Certifying -equilibria in stochastic games with limited data availability

European Journal of Control xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Control

journal homepage: www.sciencedirect.com/journal/european-journal-of-control

Certifying ε-equilibria in stochastic games with limited data availability

Filippo Fabiani [a,*], Barbara Franci [b,1]

[a] IMT School for Advanced Studies Lucca, Piazza San Francesco 19, Lucca, 55100, Italy
[b] Department of Mathematical Sciences, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, 10129, Italy

ARTICLE INFO

Recommended by Prof. T Parisini

Keywords:
Data-driven methods
Robust decision-making
Stochastic Nash games
Statistical learning

ABSTRACT

We focus on SGNEP to establish probabilistic guarantees on the quality of the solution produced by a data-driven version of the EG algorithm for SGNE seeking. Specifically, we consider a setting where the distribution of the random variable influencing the multi-agent process is unknown, and only a finite number of realizations is available for its characterization. While convergence to a SGNE should not be expected with finite data, we leverage a sample average-based approximation of the game mapping to derive distribution-free certificates bounding the distance between the true SGNE and the solution produced by the data-driven EG method. Consistently, the proposed bound shrinks as the number of samples grows to infinite.

1. Introduction

Several works have recently focused on SGNE seeking in SGNEP (Franci & Grammatico, 2020, 2021; Koshal et al., 2013; Ravat & Shanbhag, 2011). Besides its tight connection with (SVIs) (Ravat & Shanbhag, 2011), this research area has gained popularity due to its broad applicability in real-world scenarios that involve the coordination and control of self-interested agents such as the electricity or gas markets (Abada et al., 2013; Henrion & Römisch, 2007) and ride-haling services (Fabiani & Franci, 2022).

In words, a SGNEP amounts to a collection of mutually coupled optimization problems in which the cost each agent aims at minimizing does not only depend on the decisions of its opponents, but also on a random variable whose distribution is typically *unknown*. Therefore, evaluating the expected cost of each agent is not feasible. To overcome this issue, available approximation schemes provide an estimation of the pseudogradient mapping that can then be implemented within iterative procedures eventually leading to an SGNE.

Among the most popular techniques, *stochastic approximation* relies on samples of the random variable to be drawn at every iteration. Specifically, one can either select a single sample and control the approximation with a vanishing step-size sequence (Fabiani & Franci, 2022; Franci & Grammatico, 2021; Koshal et al., 2013), or increase the batch size and average the resulting operators (Franci & Grammatico, 2020, 2021; Iusem et al., 2017). The underlying approaches can demonstrate asymptotic convergence to an SGNE, however, this inevitably requires an infinite number of uncertainty realizations, which is often unrealistic in practical scenarios where only limited samples are available. Prominent examples can be found in smart grids and electricity markets based on weather forecast data. Some progress have been recently made in cases where the cost functions consist of finite sums, i.e., with the batch size being finite by definition, although too big to be easily computed (Alacaoglu & Malitsky, 2022; Franci, 2025).

Motivated by these practical considerations, we examine the scenario where only a finite number of samples are given, and use them to approximate the pseudogradient mapping of the SGNEP at hand, which is then incorporated into a popular EG method (Iusem et al., 2017; Korpelevich, 1976). Given such a limited data availability, exact convergence to an SGNE should not be expected. Thus, we leverage algorithmic stability arguments (Bousquet & Elisseeff, 2002) to derive a distribution-free certificate bounding the distance between the true SGNE and the solution obtained through the data-driven EG scheme. Consistently, it turns out that our bound shrinks with increasing samples, and hence allows one to recover an SGNE with infinite data.

Besides (Fabiani & Franci, 2026), this is the first work addressing SGNE seeking in a limited data framework while providing rigorous guarantees on the quality of the solution produced that i) are easy to compute, ii) hold with high confidence, and iii) are valid regardless the probability distribution underlying data. For direct comparison, Table 1 reports the most common algorithms for S(G)NE seeking, together with the amount of samples needed at each iteration for convergence.

* Corresponding author.
E-mail addresses: filippo.fabiani@imtlucca.it (F. Fabiani), barbara.franci@polito.it (B. Franci).

[1] Barbara Franci was supported by the "Programma Rita Levi Montalcini" of Italian Ministry of University and Research (53_RID26FRABAR).

https://doi.org/10.1016/j.ejcon.2026.101558
Received 27 May 2026; Accepted 18 July 2026
Available online 21 July 2026

Table 1
The top row reports available S(G)NE algorithms, where we tacitly omit an "S" for "stochastic" in front of each acronym, and use "F" for "forward", "B" for "backward", "L" for "loopless"; "Tik" stands for Tikhonov regularization, RFB denotes a relaxed FB, while VRG is a variance reduced gradient iteration The samples needed at every iteration are indicated in the second row: 1 means stochastic approximation, N_k indicates a time-varying increasing batch size, $0 \leq s < \infty$, while $\mathcal{B}(p)$ is a Bernoulli distribution of probability p. The latter to indicate that, at every iteration, the underlying algorithm relies on $1+N$ samples with probability p, or 1 with probability $1-p$. The finite-time or asymptotic convergence is reported in the third row ($0 \leq K < \infty$), while the bottom one indicates whether such convergence returns an SNE, SGNE, or one of these by considering cost functions with finite sum (marked with an asterisk).

	FBF (Boţ et al., 2021)	Tik (Fabiani & Franci, 2022)	RFB (Franci & Grammatico, 2021)	RFB (Franci & Grammatico, 2021)	VRG (Gower et al., 2020)	FB (Franci & Grammatico, 2020)	FB (Fabiani & Franci, 2026)	LEG (Franci, 2025)	EG (Iusem et al., 2017)	EG (This paper)
Sample(s)	N_k	1	N_k	1	$1+N\mathcal{B}(p)$	N_k	s	$1+N\mathcal{B}(p)$	N_k	s
Convergence	$k \to \infty$	$k \to \infty$	$k \to \infty$	$k \to \infty$	$k \to \infty$	$k \to \infty$	$k < K$	$k \to \infty$	$k \to \infty$	$k < K$
Equilibrium	VI/SNE	SGNE	SGNE	SNE	VI/SNE*	SGNE	ε-SGNE	SGNE*	SGNE/VI	ε-SGNE

1.1. Algorithmic stability notions and results

Let $\mathcal{D}_s := \{\xi^{(1)}, \dots, \xi^{(s)}\}$ be a dataset consisting of s iid samples drawn from an *unknown* probability measure $\mathbb{P}$ attached to Ξ^s, where $\xi^{(i)} \in \Xi$ for all $i = 1, \dots, s$. Then, an *algorithm* amounts to an indexed family of mappings $\{A_s\}_{s\geq 0}$, with $A_s : \Xi^s \to \mathbb{R}^n$ taking a dataset $\mathcal{D}_s$ and returning a hypothesis $H_s := A_s(\{\xi^{(i)}\}_{i=1}^s) = A_s(\mathcal{D}_s)$ (Vidyasagar, 2013). Given $\mathcal{D}_s$, we denote with $\mathcal{D}_s^i = \{\xi^{(1)}, \dots, \xi^{(i-1)}, \xi', \xi^{(i+1)}, \dots, \xi^{(s)}\}$ the set obtained by replacing the ith element of $\mathcal{D}_s$ with $\xi' \in \Xi$ drawn according to $\mathbb{P}$, iid wrt $\mathcal{D}_s \setminus \{\xi^{(i)}\}$. Accordingly, the associated hypothesis is $H_{s^i} := A_s(\mathcal{D}_s^i)$.

The performance of some algorithm, and particularly of the produced hypothesis H wrt an example ξ, are usually measured through a *loss function* $\ell : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$. Attached to the latter, one can define the so-called *generalization error*, a random variable (since the hypothesis generated depends on $\mathcal{D}_s$) defined as follows:

$$r(A, s) = \mathbb{E}_{\mathbb{P}}\big[\ell(H_s, \xi)\big], \tag{1}$$

where we use A instead of A_s as first argument. Computing (1) then requires $\mathbb{P}$, which might be unavailable in several cases. An estimate of (1) refers to the *empirical error*:

$$\hat{r}(A, s) = \frac{1}{s}\sum_{i=1}^{s} \ell(H_s, \xi^{(i)}). \tag{2}$$

Definition 1 (Kutin and Niyogi (2002)). An algorithm $\{A_s\}_{s\geq 0}$ has uniform stability β wrt the loss function ℓ if for all $s \geq 0$, dataset with s samples $\mathcal{D}_s$, and $i \in \{1, \dots, s\}$, $|\ell(H_s, \xi) - \ell(H_{s^i}, \xi)| \leq \beta$ for all $\xi \in \Xi$.

In words, an algorithm is uniformly stable if changing one sample $\xi^{(i)}$ in its input dataset does not significantly alter the output wrt to some predefined loss ℓ. In general, the smaller the β the better the stability features possessed by the algorithm at hand, measured through ℓ. The following result, based on uniform stability, will be key in the remainder:

Lemma 1 (Kutin & Niyogi, 2002, Th. 3.2). *Let $\{A_s\}_{s\geq 0}$ be an algorithm with uniform stability β wrt a loss function $0 \leq \ell(H_s, \xi) \leq M$, $M \geq 0$, for all $\xi \in \Xi$ and $\mathcal{D}_s$. Then, for any $s \geq 1$ and $\delta \in (0, 1)$, it holds that: $\mathbb{P}^s\{\mathcal{D}_s \in \Xi^s : r(A, s) \leq \hat{r}(A, s) + \beta + (s\beta + M)\sqrt{2\ln(1/\delta)/s}\} \geq 1 - \delta$.*

Specifically, Lemma 1 provides with an upper bound on the risk associated to hypothesis $H_s = A_s(\mathcal{D}_s)$, which holds with arbitrary confidence $1 - \delta$, such that i) if β scales proportional to $1/s$, vanishes as $s \to \infty$, and ii) it is exponential in the confidence parameter δ. The former condition is called *consistency* property, and provides a "tight" certificate in the sense that the upper bound in Lemma 1 vanishes as $s \to \infty$.

1.2. Notation and operator-theoretic definitions

$\mathbb{R}$ denotes the set of real numbers and $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. The standard inner product is represented by $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, while $\|\cdot\|$ denotes the associated Euclidean norm. $\mathbf{0}_m$ indicates the vector with m entries all equal to 0. Given N vectors $x_1, \dots, x_N$, with each $x_i \in \mathbb{R}^{n_i}$, $\boldsymbol{x} := \operatorname{col}(x_1, \dots, x_N) = [x_1^\top, \dots, x_N^\top]^\top$, while $\boldsymbol{x}_{-i} := \operatorname{col}((x_j)_{j\neq i}) = [x_1^\top, \dots, x_{i-1}^\top, x_{i+1}^\top, \dots, x_N^\top]^\top$. With some abuse of notation, we also use $\boldsymbol{x} = (x_i, \boldsymbol{x}_{-i})$. The uniform distribution on $[a, b]$ is denoted by $\mathcal{U}(a, b)$, and the normal distribution with mean μ and variance σ^2 by $\mathcal{N}(\mu, \sigma^2)$.

For all $x, y \in \operatorname{dom}(F)$, a mapping $F : \operatorname{dom} F \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is κ-Lipschitz continuous if, for some $\kappa > 0$, $\|F(x) - F(y)\| \leq \kappa\|x - y\|$; pseudo-monotone if $\langle F(y), (x - y)\rangle \geq 0 \Rightarrow \langle F(x), (x - y)\rangle \geq 0$; monotone if $\langle F(x) - F(y), x - y\rangle \geq 0$; μ-strongly monotone if, for some $\mu > 0$, $\langle F(x) - F(y), x - y\rangle \geq \mu\|x - y\|^2$. The following result will be useful to apply our stability-based arguments:

Lemma 2 (Lei and Shanbhag (2022)). *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be an μ-strongly monotone and κ-Lipschitz continuous mapping, with κ, $\mu > 0$. If $\gamma \in (0, 2\mu/\kappa^2)$, then it holds that $\|x - \gamma F(x) - (y - \gamma F(y))\| \leq \rho\|x - y\|$, with $\rho := \sqrt{1 - 2\gamma\mu + \gamma^2\kappa^2} < 1$.*

2. Problem formulation

We start by introducing SGNEP and discussing a popular reformulation based on barrier functions for imposing (soft) constraint satisfaction. Then, we formalize the considered SGNE seeking problem with limited data availability.

2.1. Stochastic generalized Nash equilibrium problems

We consider a set $\mathcal{I} = \{1, \dots, N\}$ indexing N noncooperative agents taking part to a Nash game. In particular, each of them chooses its strategy $x_i \in \mathbb{R}^{n_i}$, aiming at minimizing its local cost function $\mathbb{J}_i : \mathbb{R}^{n_i} \times \mathbb{R}^{n-n_i} \to \mathbb{R}$, $n := \sum_{i\in\mathcal{I}} n_i$, which also depends on the decisions of the ith agent's opponents $\boldsymbol{x}_{-i} \in \mathbb{R}^{n-n_i}$. Instead, $\boldsymbol{x} \in \mathbb{R}^n$ represents the collective decision strategy of the population of agents $\mathcal{I}$, taking values in the set $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n : g(\boldsymbol{x}) \leq \mathbf{0}_m\}$, where $g : \mathbb{R}^n \to \mathbb{R}^m$. We indicate the feasible set of agent i with $\mathcal{X}_i(\boldsymbol{x}_{-i}) \subseteq \mathbb{R}^{n_i}$, which explicitly accounts for the possible dependency on the other participants' decision variables. Local constraints are, instead, implicitly included in $\mathcal{X}$.

Standing Assumption 1. *Let $g(\boldsymbol{x}) = \operatorname{col}((g_j(\boldsymbol{x}))_{j=1}^m)$, with each $g_j : \mathbb{R}^n \to \mathbb{R}$. The following conditions hold:*

(i) Each $\boldsymbol{x} \mapsto g_j(\boldsymbol{x})$ is convex and of class C^2;
(ii) The set $\mathcal{X}$ is nonempty, convex, compact, and satisfies Slater's constraint qualification.

These assumptions are standard for SGNEP (Franci & Grammatico, 2021; Romano & Pavel, 2022, 2023). Moreover, we assume each agents' cost function in the form:

$$\mathbb{J}_i(x_i, \boldsymbol{x}_{-i}) := \mathbb{E}_{\mathbb{P}}[f_i(x_i, \boldsymbol{x}_{-i}, \xi(\omega))], \tag{3}$$

where $f_i : \mathbb{R}^{n_i} \times \mathbb{R}^{n-n_i} \times \mathbb{R}^d \to \mathbb{R}$. The variable $\xi : \Xi \to \mathbb{R}^d$ represents a random quantity affecting each local cost, fully characterized through the probability space $(\Xi, \mathcal{F}, \mathbb{P})$, while $\mathbb{E}_{\mathbb{P}}$ denotes the expectation wrt the distribution $\mathbb{P}$ of ξ, which is assumed here to be *unknown*. We make the

blanket assumption that $\mathbb{E}_{\mathbb{P}}[f_i(\boldsymbol{x},\xi)]$ is well defined for all $\boldsymbol{x}\in\mathcal{X}$. With some abuse of notation, from now on we simply use ξ to indicate the sample $\xi(\omega)$, as common in the literature (Iusem et al., 2017).

Thus, each agent wishes to solve its local optimization problem, thereby originating a so-called SGNEP:

$$\forall i\in\mathcal{I}:\quad \begin{cases}\min\limits_{x_i\in\mathbb{R}^{n_i}} & \mathbb{J}_i\left(x_i,\boldsymbol{x}_{-i}\right)\\ \text{s.t.} & g(x_i,\boldsymbol{x}_{-i})\leq \mathbf{0}_{\mathbf{m}}.\end{cases} \tag{4}$$

Next, we formalize a popular solution for SGNEP (Franci & Grammatico, 2020, 2021):

Definition 2. An SGNE is a collective decision strategy $\boldsymbol{x}^\star\in\mathcal{X}$ such that, for all $i\in\mathcal{I}$,

$$\mathbb{J}_i(x_i^\star,\boldsymbol{x}_{-i}^\star)\leq\mathbb{J}_i(y_i,\boldsymbol{x}_{-i}^\star)\text{ for all }y_i\in\mathcal{X}_i(\boldsymbol{x}_{-i}^\star).$$

According to (Ravat & Shanbhag, 2011, Prop. 3.5), we make now few assumptions ensuring that at least an SGNE exists for the SGNEP in (4):

Standing Assumption 2. *The following conditions hold:*

(i) Each $x_i\mapsto f_i\left(x_i,\boldsymbol{x}_{-i},\xi\right)$ is convex, of class $\mathcal{C}^1$, and Lipschitz continuous with integrable Lipschitz constant w.r.t. ξ, while $\xi\mapsto f_i\left(x_i,\boldsymbol{x}_{-i},\xi\right)$ is measurable;

(ii) Each $x_i\mapsto\mathbb{J}_i\left(x_i,\boldsymbol{x}_{-i}\right)$ is convex and of class $\mathcal{C}^1$.

Among all possible SGNE, we focus on those corresponding to the solution set of the SVI associated to the SGNEP. To characterize them, let us first define the pseudogradient mapping $\mathbb{F}:\mathbb{R}^n\rightrightarrows\mathbb{R}^n$ related to each cost function:

$$\mathbb{F}(\boldsymbol{x})=\operatorname{col}\left(\left(\mathbb{E}_{\mathbb{P}}[\nabla_{x_i}f_i(x_i,\boldsymbol{x}_{-i},\xi)]\right)_{i\in\mathcal{I}}\right). \tag{5}$$

Thus, the associated SVI problem is formalized as:

$$\text{Find }\boldsymbol{x}^\star\in\mathcal{X}\text{ such that }\langle\mathbb{F}(\boldsymbol{x}^\star),\boldsymbol{x}-\boldsymbol{x}^\star\rangle\geq 0\,\forall\boldsymbol{x}\in\mathcal{X}. \tag{6}$$

While being a solution to the SVI above is necessary and sufficient to be a SNE of a game with no coupling constraints, the set of (v-SGNE) of (6) is a subset of all the equilibria of the SGNEP in (4) (Facchinei et al., 2007). We, however, use v-SGNE and SGNE interchangeably in the following. To impose constraint satisfaction while turning the SGNEP in (4) into an unconstrained SNEP, a popular approach relies on barrier functions (Fabiani & Caiti, 2019; Romano & Pavel, 2022, 2023). We can then adjust each cost function in (4) by adding a barrier term as follows:

$$\forall i\in\mathcal{I}:\ \min_{x_i\in\mathbb{R}^{n_i}}\ \mathbb{J}_i\left(x_i,\boldsymbol{x}_{-i}\right)+\phi(\boldsymbol{x}), \tag{7}$$

where the function ϕ is a log-barrier function, with $\varrho>0$:

$$\phi(\boldsymbol{x})=\begin{cases}-\varrho\sum_{j=1}^m\log\left(-g_j(\boldsymbol{x})\right), & g(\boldsymbol{x})<\mathbf{0}_m\\ +\infty, & \text{otherwise.}\end{cases} \tag{8}$$

From Standing Assumption 1, $\boldsymbol{x}\mapsto\phi(\boldsymbol{x})$ is convex and of class $\mathcal{C}^2$ on $\operatorname{int}(\mathcal{X})$, and hence its gradient $\nabla\phi(\boldsymbol{x})$ is monotone and locally Lipschitz with constant ℓ_ϕ (Romano & Pavel, 2022, 2023).

Remark 1. Solving (7) typically provides an approximate solution of the SGNEP in (4) (Romano & Pavel, 2023, Lemma 4) while for ϱ shrinking to zero, the SNE of (7) converges to an SGNE of (4) (Romano & Pavel, 2022, Lemma 2). In practice, however, setting a small enough $\varrho>0$ is usually sufficient to recover an SGNE of the SGNEP, and it is actually what we will implicitly assume throughout the paper. This is why, when referring to a solution to the unconstrained SNEP in (7), we will equivalently refer to an SGNE of (4).

The resulting pseudogradient mapping for the unconstrained SNEP in (7), $\mathbb{B}:\mathbb{R}^n\rightrightarrows\mathbb{R}^n$, formally reads as:

$$\mathbb{B}(\boldsymbol{x})=\mathbb{F}(\boldsymbol{x})+\nabla\phi(\boldsymbol{x}). \tag{9}$$

Standing Assumption 3. *The following conditions hold:*

(i) $\mathbb{B}$ is μ-strongly monotone;

(ii) $\mathbb{B}$ is κ-Lipschitz continuous;

(iii) $\exists M>0$ such that $\|\mathbb{B}(\boldsymbol{x})\|\leq M$, for all $\boldsymbol{x}\in\operatorname{int}(\mathcal{X})$.

Remark 2. The first requirement holds, for instance, if the pseudogradient mapping $\mathbb{F}$ in (5) is strongly monotone, since $\nabla\phi(\boldsymbol{x})$ is monotone. Under this assumption on $\mathbb{F}$, the solution to (6) is also unique (Facchinei & Pang, 2007). Similarly, Standing Assumption 3.(ii) is a consequence of Standing Assumption 1, and possibly the Lipschitz continuity of $\mathbb{F}$ (Franci, 2025; Romano & Pavel, 2022, 2023).

Standing Assumption 3.(iii), instead, holds for instance in case of polyhedral constraints, i.e., $\mathcal{X}=\{\boldsymbol{x}\in\mathbb{R}^n:A\boldsymbol{x}\leq b\}$, for $A\in\mathbb{R}^{m\times n}$ and $b\in\mathbb{R}^m$. Collectively, these assumptions are common in the (S)GNE seeking literature (Franci & Grammatico, 2020, 2021).

2.2. A framework with limited data availability

According to (Facchinei & Pang, 2007), if $\boldsymbol{x}^\star$ is an SGNE of (7), then $\mathbb{B}(\boldsymbol{x}^\star)=0$. Thus, we define the set SGNE_ε, which consist of all those points ε-close to an SGNE of (7) (and hence of (4)), as:

$$\text{SGNE}_\varepsilon:=\{\boldsymbol{x}\in\mathcal{X}:\|\mathbb{B}(\boldsymbol{x})\|\leq\varepsilon\}.$$

As common in SGNEPs, the mapping $\mathbb{F}(\cdot)$ is either unavailable, or quite challenging to calculate, because it involves the expected value $\mathbb{E}_{\mathbb{P}}[\cdot]$, i.e., a multidimensional integration that depends on the distribution $\mathbb{P}$ of ξ which might be unknown. Therefore, by assuming to have available only a few realizations of the uncertain variable we approximate $\mathbb{B}(\cdot)$ in a data-driven fashion by means of an oracle $O:\Omega\times\Xi\to\mathbb{R}^n$, assumed as a Borel function satisfying:

$$\text{For all }\boldsymbol{x}\in\Omega,\ \mathbb{E}_{\mathbb{P}}[O(\boldsymbol{x},\xi)]=\mathbb{B}(\boldsymbol{x}).$$

In the remainder, we tacitly assume that the oracle O possesses the same properties postulated for the pseudogradient operator $\mathbb{B}$ in (9) (Farnia & Ozdaglar, 2021; Hardt et al., 2016). Relying on an oracle is a rather common assumption for stochastic algorithms and essentially means that it is possible to compute the pseudogradient given one sample (Alacaoglu & Malitsky, 2022; Fabiani & Franci, 2026; Iusem et al., 2017). In our analysis, we hence leverage the following data-based approximation for $\mathbb{B}(\cdot)$:

$$\hat{B}_s(\boldsymbol{x})=\hat{B}(\boldsymbol{x},\mathcal{D}_s):=\frac{1}{s}\sum_{r=1}^{s}O(\boldsymbol{x},\xi^{(r)}), \tag{10}$$

which depends on s iid samples made available from the set Ξ, and collected into $\mathcal{D}_s:=\{\xi^{(r)}\}_{r=1}^s\in\Xi^s$.

The question we wish to answer is then the following: given a finite dataset, are we able to quantify through an explicit calculation of some $\varepsilon>0$, how far we can get from an SGNE by employing $\hat{B}_s$ in place of the actual $\mathbb{B}$?

3. SGNE seeking under the lens of algorithmic stability

First, we introduce the EG scheme applied to SGNE seeking, and the data-driven version adapted for our analysis. Next, we show its stability properties wrt a tailored loss function, which is crucial for proving our main result.

3.1. The extragradient method

Our analysis focuses on the procedure reported in Algorithm 1, which consists in the distributed and data-driven version of the following EG in compact form (Ryu & Boyd, 2016, §7.2):

$$\begin{cases}\boldsymbol{y}^k=\boldsymbol{x}^k-\gamma\mathbb{B}(\boldsymbol{x}^k)\\ \boldsymbol{x}^{k+1}=\boldsymbol{x}^k-\gamma\mathbb{B}(\boldsymbol{y}^k).\end{cases} \tag{11}$$

The expected value required to evaluate $\mathbb{B}$ can then be replaced through a data-driven approximation of the pseudogradient mapping

Algorithm 1: Approximated EG method.

Initialization: Each x_i^0 so that $\boldsymbol{x}^0 \in \text{int}(\mathcal{X})$

Iteration $k \in \mathbb{N}_0$**:** Agent i

1. Receives $\boldsymbol{x}_{-i}^k$, then updates

$$y_i^k = x_i^k - \frac{\gamma}{s}\sum_{r=1}^{s} O(x_i^k, \boldsymbol{x}_{-i}^k, \xi^{(r)})$$

2. Receives $\boldsymbol{y}_{-i}^k$, then updates

$$x_i^{k+1} = x_i^k - \frac{\gamma}{s}\sum_{r=1}^{s} O(y_i^k, \boldsymbol{y}_{-i}^k, \xi^{(r)})$$

as in (10). To simplify the analysis we have adopted the same, constant step-size $\gamma > 0$ and the same dataset $\mathcal{D}_s$ for all the agents, although a generalization to local, possibly time-varying $\{\gamma_i^k\}_{i\in\mathcal{I}}$ and $\mathcal{D}_{s^i}$ is doable.

Remark 3. We note that the scheme in (11) converges almost surely to an SGNE when the true operator $\mathbb{B}(\cdot)$ is at least pseudomonotone and Lipschitz continuous, and an increasing batch $s = s^k \to \infty$ is available at every iteration (Iusem et al., 2017, Th. 3.18), (Franci, 2025, Theoram. 1). Thus, if we knew $\mathbb{B}(\cdot)$, the conditions in Standing Assumption 3 would suffice for the convergence of (11).

In a data-limited context our goal is hence to investigate how far we can get from the (unique) v-SGNE by running Algorithm 1 for an arbitrary number of iterations K. Specifically, we want to characterize the resulting $\boldsymbol{x}^{K+1}$ as an ε-SNE of the SNEP in (7). To this end, we will make use of the algorithmic stability arguments discussed in Section 1.1.

3.2. A computable bound for ε-SGNE

By replacing $\mathbb{B}$ with the approximation $\hat{B}_s$ in (10), the resulting data-driven variant then proposes the distributed updates in Algorithm 1. According to Section 1.1, $\omega := \text{col}(\boldsymbol{y}, \boldsymbol{x}) \in \mathbb{R}^{2n}$ then amounts to the hypothesis of our algorithm. Under the postulated assumptions we will prove next that Algorithm 1 is uniformly stable wrt the following loss function, which measures the performance of a generic hypothesis H:

$$\begin{aligned}\ell(H,\xi) &= \left\|[0_n \quad I_n]H - \gamma O([\mathbf{0}_n \quad I_n]H,\xi)\right\| \\ &= \|x - \gamma O(x,\xi)\|.\end{aligned} \tag{12}$$

In particular, we will obtain a stability parameter $\beta \propto (1/s)$, thereby making the probabilistic condition in Lemma 1 tight.

Lemma 3. *Let $\gamma \in (0, 2\mu/\kappa^2)$ be chosen so that $\tau = \tau(\gamma) := (1+\gamma\kappa)\rho + \gamma\kappa < 1$. Then, the data-driven EG in Algorithm 1 possesses uniform stability proportional to $1/s$ wrt the loss function $\ell(H,\xi)$ in (12).*

Proof. Fix some $s \geq 1$, and consider two datasets, $\mathcal{D}_s$, $\mathcal{D}_s^t \in \Xi^s$, both consisting of s-iid random samples and differing on the tth instance only, i.e., some ξ' replaces $\xi^{(t)}$ in $\mathcal{D}_s^t$. After iterating (a compact version of) Algorithm 1 for $K \geq 1$ times starting from some $\boldsymbol{x}^0 \in \text{int}(\mathcal{X})$, we obtain $\omega^{K+1} = \text{col}(\boldsymbol{y}^K, \boldsymbol{x}^{K+1})$ and $\tilde{\omega}^{K+1} = \text{col}(\tilde{\boldsymbol{y}}^K, \tilde{\boldsymbol{x}}^{K+1})$ as the two hypotheses leveraging samples in $\mathcal{D}_s \in \Xi^s$ and $\mathcal{D}_s^t \in \Xi^s$, respectively. Once picked an arbitrary $\xi \in \Xi$, we have:

$$\begin{aligned}&|\ell(H_s,\xi) - \ell(H_{s^t},\xi)| \\ &\quad\overset{(a)}{\leq} \left\|\boldsymbol{x}^{K+1} - \gamma O(\boldsymbol{x}^{K+1},\xi) - \tilde{\boldsymbol{x}}^{K+1} + \gamma O(\tilde{\boldsymbol{x}}^{K+1},\xi)\right\| \\ &\quad\overset{(b)}{\leq} \rho\left\|\boldsymbol{x}^{K+1} - \tilde{\boldsymbol{x}}^{K+1}\right\|,\end{aligned} \tag{13}$$

where (a) follows by the reverse triangle inequality, while (b) from Lemma 2, $\rho \in (0,1)$. It remains to bound $\|\boldsymbol{x}^{K+1} - \tilde{\boldsymbol{x}}^{K+1}\|$. To this end, consider Algorithm 1 at the generic kth iteration, which produces $\boldsymbol{x}^{k+1}$ and $\tilde{\boldsymbol{x}}^{k+1}$ such that:

$$\begin{aligned}&\|\boldsymbol{x}^{k+1} - \tilde{\boldsymbol{x}}^{k+1}\| \\ &= \|\boldsymbol{x}^k - \tfrac{\gamma}{s}\textstyle\sum_{r=1}^s O(\boldsymbol{x}^k - \tfrac{\gamma}{s}\sum_{r=1}^s O(\boldsymbol{x}^k,\xi^{(r)}),\xi^{(r)}) \\ &\qquad - (\tilde{\boldsymbol{x}}^k - \tfrac{\gamma}{s}\textstyle\sum_{r=1}^s O(\tilde{\boldsymbol{x}}^k - \tfrac{\gamma}{s}\sum_{r=1}^s O(\tilde{\boldsymbol{x}}^k,\xi^{(r)}),\xi^{(r)})\| \\ &\leq \|\boldsymbol{x}^k - \tfrac{\gamma}{s}\textstyle\sum_{r=1,r\neq t}^s O(\boldsymbol{x}^k - \tfrac{\gamma}{s}\sum_{r=1}^s O(\boldsymbol{x}^k,\xi^{(r)}),\xi^{(r)}) \\ &\qquad - (\tilde{\boldsymbol{x}}^k - \tfrac{\gamma}{s}\textstyle\sum_{r=1,r\neq t}^s O(\tilde{\boldsymbol{x}}^k - \tfrac{\gamma}{s}\sum_{r=1}^s O(\tilde{\boldsymbol{x}}^k,\xi^{(r)}),\xi^{(r)})\| \\ &\qquad + \tfrac{\gamma}{s}\|O(\boldsymbol{x}^k - \tfrac{\gamma}{s}\textstyle\sum_{r=1}^s O(\boldsymbol{x}^k,\xi^{(r)}),\xi^{(t)})\| \\ &\qquad + \tfrac{\gamma}{s}\|O(\tilde{\boldsymbol{x}}^k - \tfrac{\gamma}{s}\textstyle\sum_{r=1}^s O(\tilde{\boldsymbol{x}}^k,\xi^{(r)}),\xi')\| \\ &\leq \|\boldsymbol{x}^k - \tfrac{\gamma}{s}\textstyle\sum_{r=1,r\neq t}^s O(\boldsymbol{x}^k - \tfrac{\gamma}{s}\sum_{r=1}^s O(\boldsymbol{x}^k,\xi^{(r)}),\xi^{(r)}) \\ &\qquad - (\tilde{\boldsymbol{x}}^k - \tfrac{\gamma}{s}\textstyle\sum_{r=1,r\neq t}^s O(\tilde{\boldsymbol{x}}^k - \tfrac{\gamma}{s}\sum_{r=1}^s O(\tilde{\boldsymbol{x}}^k,\xi^{(r)}),\xi^{(r)})\| \\ &\qquad + \tfrac{2\gamma M}{s},\end{aligned}$$

where the last inequality follows by virtue of Standing Assumption 3.(iii). Note that the arguments of both outer operators $O(\cdot)$ coincide with the gradient descent update evaluated from $\boldsymbol{x}^k$ and $\tilde{\boldsymbol{x}}^k$ with the same pool of samples, denoted as $\boldsymbol{z}^{k+1} = \boldsymbol{x}^k - \frac{\gamma}{s}\sum_{r=1,r\neq t}^s O(\boldsymbol{x}^k,\xi^{(r)})$ and $\tilde{\boldsymbol{z}}^{k+1} = \tilde{\boldsymbol{x}}^k - \frac{\gamma}{s}\sum_{r=1,r\neq t}^s O(\tilde{\boldsymbol{x}}^k,\xi^{(r)})$, respectively. Thus, after summing and subtracting both $\boldsymbol{z}^{k+1}$ and $\tilde{\boldsymbol{z}}^{k+1}$ inside the norm above, standard manipulations yield:

$$\begin{aligned}\|\boldsymbol{x}^{k+1} - \tilde{\boldsymbol{x}}^{k+1}\| &\leq \|\boldsymbol{z}^{k+1} - \tilde{\boldsymbol{z}}^{k+1}\| + \tfrac{\gamma}{s}\textstyle\sum_{\substack{r=1\\r\neq t}}^s \|O(\boldsymbol{z}^{k+1},\xi^{(r)}) \\ &\qquad - O(\tilde{\boldsymbol{z}}^{k+1},\xi^{(r)}) + O(\boldsymbol{x}^k,\xi^{(r)}) - O(\tilde{\boldsymbol{x}}^k,\xi^{(r)})\| + \tfrac{2\gamma M}{s}, \\ &\leq (1+\gamma\kappa)\|\boldsymbol{z}^{k+1} - \tilde{\boldsymbol{z}}^{k+1}\| + \gamma\kappa\|\boldsymbol{x}^k - \tilde{\boldsymbol{x}}^k\| + \tfrac{2\gamma M}{s}.\end{aligned}$$

By relying on the growth lemma (Farnia & Ozdaglar, 2021, Lemma 2), we can then upper bound the last relation, particularly the term $\|\boldsymbol{z}^{k+1} - \tilde{\boldsymbol{z}}^{k+1}\|$, as follows: $\|\boldsymbol{x}^{k+1} - \tilde{\boldsymbol{x}}^{k+1}\| \leq ((1+\gamma\kappa)\rho + \gamma\kappa)\|\boldsymbol{x}^k - \tilde{\boldsymbol{x}}^k\| + \frac{2\gamma M}{s}$, where $\rho < 1$ is the contraction factor characterizing the gradient descent step in Lemma 2 for $\gamma \in (0, 2\mu/\kappa^2)$. Then, recursing over the K-steps, and considering that both trajectories leading to $\boldsymbol{x}^{K+1}$ and $\tilde{\boldsymbol{x}}^{K+1}$ originate from the same initial conditions leaves us with:

$$\begin{aligned}\|\boldsymbol{x}^{K+1} - \tilde{\boldsymbol{x}}^{K+1}\| &\leq \frac{2\gamma M}{s}\sum_{k=1}^{K}((1+\gamma\kappa)\rho + \gamma\kappa)^{K-k} \\ &\leq \frac{2\gamma M}{s(1-\gamma\kappa-(1+\gamma\kappa)\rho)},\end{aligned}$$

where we have exploited the relation $\sum_{k=1}^K \alpha^{K-k} = \frac{1-\alpha^K}{1-\alpha} < \frac{1}{1-\alpha}$ which holds true for any $\alpha \in [0,1)$, i.e., $(1+\gamma\kappa)\rho + \gamma\kappa = \tau < 1$. Finally, combining the latter inequality with (13) allows us to conclude that:

$$|\ell(H_s,\xi) - \ell(H_{s^t},\xi)| \leq \frac{2\gamma\rho M}{s(1-\tau)},$$

thereby proving uniform stability proportional to $1/s$. □

Remark 4. Lemma 3, and later Theorem 1, assume the existence of some γ so that $(1+\gamma\kappa)\rho + \gamma\kappa < 1$. The latter condition ensures a sort of global contraction for the data-driven EG scheme. While establishing explicit conditions on γ guaranteeing $\tau < 1$ is hard, one can numerically observe that meeting such a condition is possible, making the resulting assumption reasonable–see Appendix A.

Then, equipped with this uniform stability result, we are now ready to state and prove what follows:

Theorem 1. *Fix $s \geq 1$, $\delta \in (0,1)$, and let $\gamma \in (0, 2\mu/\kappa^2)$ be chosen so that $\tau(\gamma) < 1$. Given any dataset $\mathcal{D}_s \in \Xi^s$, there exists a computable $\varepsilon = \varepsilon(s,\delta) > 0$ such that $\boldsymbol{x}^{K+1} \in \text{SGNE}_\varepsilon$ with probability at least $1-\delta$, where $\boldsymbol{x}^{K+1}$ is a subvector of ω^{K+1} obtained by iterating the data-driven EG in Algorithm 1 K-times starting from $\boldsymbol{x}^0 \in \text{int}(\mathcal{X})$.*

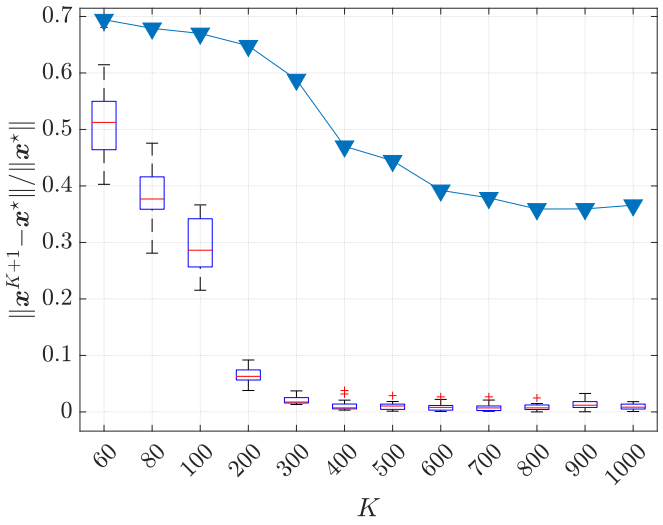

Fig. 1. Box plots capturing the relative approximation error $\|\boldsymbol{x}^{K+1}-\boldsymbol{x}^\star\|$ produced by different run K of Algorithm 1 over 20 trials with fixed number of samples $s = 1000$. The blue down-pointing triangles denote the resulting averaged bounds ε offered by Theorem 1 with $\delta = 0.05$.

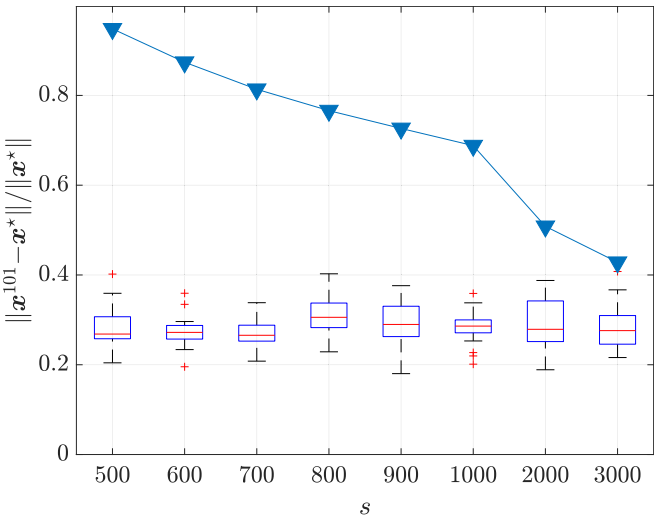

Fig. 2. Box plots capturing the relative approximation error $\|\boldsymbol{x}^{101}-\boldsymbol{x}^\star\|$ produced by $K = 100$ iterations of Algorithm 1 over 20 trials considering different dataset size. The blue down-pointing triangles denote the resulting averaged bounds ε offered by Theorem 1 with $\delta = 0.05$.

Proof. After iterating the instructions reported in Algorithm 1 K-times, consider the following relations:

$$\begin{aligned}\|\boldsymbol{x}^\star-\gamma\mathbb{B}(\boldsymbol{x}^\star)\| = 0 \le \left\|\boldsymbol{x}^{K+1}-\gamma\mathbb{B}(\boldsymbol{x}^{K+1})\right\| \\ \overset{(a)}{=} \left\|\boldsymbol{x}^{K+1}-\gamma\mathbb{E}_{\mathbb{P}}\left[O(\boldsymbol{x}^{K+1},\xi)\right]\right\| \\ \overset{(b)}{=} \left\|\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{x}^{K+1}-\gamma O(\boldsymbol{x}^{K+1},\xi)\right]\right\| \\ \overset{(c)}{\le} \mathbb{E}_{\mathbb{P}}\left[\left\|\boldsymbol{x}^{K+1}-\gamma O(\boldsymbol{x}^{K+1},\xi)\right\|\right] \\ = \mathbb{E}_{\mathbb{P}}\left[\ell(H_s,\xi)\right] = r(\mathcal{A},s),\end{aligned}$$

where we have used in (a) the fact that the oracle operator is unbiased, in (b) the fact that the data-driven EG in Algorithm 1 yields a deterministic output ω^{K+1} once fixed the dataset $\mathcal{D}_s$, as well as the linearity of the expected value, and in (c) the standard Jensen's inequality. Thus, upper bounding the risk $\mathbb{E}_{\mathbb{P}}\left[\ell(H_s,\xi)\right]$ provides an equivalent information on the distance of $\boldsymbol{x}^{K+1}$ from a fixed point of the dynamics produced by iterating the operator $\mathrm{Id}-\gamma\mathbb{B}\circ(\mathrm{Id}-\gamma\mathbb{B})$ with perfect knowledge on $\mathbb{B}$. Therefore, with probability at least $1-\delta$, the direct application of the bound in Lemma 3 leads to $\left\|\boldsymbol{x}^{K+1}-\gamma\mathbb{B}(\boldsymbol{x}^{K+1})\right\| \le \varepsilon$ with $\varepsilon := \sum_{r=1}^{s}\left\|\boldsymbol{x}^{K+1}-\gamma O(\boldsymbol{x}^{K+1},\xi^{(r)})\right\|/s + \beta + (s\beta+M)\sqrt{2\ln(1/\delta)/s}$ and $\beta = 2\gamma\rho M/s(1-\tau)$. □

Thus, Theorem 1 provides a computable bound, which holds with arbitrarily high confidence $1-\delta$ wrt any possible dataset drawn from Ξ^s and probability distribution $\mathbb{P}$ underlying these data, on the distance between the output of Algorithm 1 after K steps, and an SGNE of the SGNEP in (4). Note that the specific choice of the loss function in (12) makes the calculation of ε possible a-posteriori (i.e., after K iterations). In fact, all of the terms $\left\|\boldsymbol{x}^{K+1}-\gamma O(\boldsymbol{x}^{K+1},\xi^{(r)})\right\|$ do not explicitly depend on unknown quantities, e.g., the true SGNE $\boldsymbol{x}^\star$, while the other terms depend on design parameters (δ,γ), problem data (s, M, μ, κ), or both (ρ,τ).

4. Illustrative example

We now test our bounds on an academic example involving $N = 20$ agents controlling an unconstrained scalar decision variable (i.e., $n_i = 1$) to minimize a quadratic cost function $f_i(x_i,\boldsymbol{x}_{-i},\xi) = a_i x_i^2 + (b_i+\xi)x_i + (\sum_{j\in\mathcal{I}\setminus\{i\}} c_{i,j}x_j)x_i$. Specifically, each parameter in the cost is chosen randomly, i.e., $a_i \sim \mathcal{U}(0.2, 0.8)$, $b_i \sim \mathcal{U}(0.02, 0.075)$, $c_{i,j} \sim \mathcal{N}(0, 0.01)$, while the stochastic variable ξ is distributed according to $\mathcal{N}(-1, 0.2)$. Note that the results from the previous section do not directly involve quantities related to the collective strategy profile $\boldsymbol{x}$, e.g., $n = \sum_{i\in\mathcal{I}} n_i$, nor $\nabla\phi(\cdot)$, and hence the constraints – see the discussion in Remark 2. This explains the unconstrained nature of the considered game with scalar strategies, which has the sole purpose of computing the proposed bounds numerically. In all the instances, the true SNE $\boldsymbol{x}^\star$ is computed via (Franci & Grammatico, 2020, Algorithm. 1).

We start by investigating the behaviour of Algorithm 1 by fixing the number of samples s and varying K. Specifically, we perform 20 trials for each value of $K \in \{60, 80, \dots, 1000\}$, and populate the dataset $\mathcal{D}_s$ with $s = 1000$ samples. In Fig. 1 we illustrate the numerical results obtained with confidence level $\delta = 0.05$ and step-size γ chosen to meet the condition in Theorem 1. While, in practice, Algorithm 1 well approximates the unique SNE when run for $K \ge 300$, the theoretical bound established in Theorem 1 reveals an intrinsic dependency on K itself. In fact, the averaged bound obtained appears dominated by the empirical error term, which progressively reduces its impact with larger K, settling around 0.35. By running Algorithm 1 for $K = 1000$ iterations with $s = 1000$ samples, for instance, one can then claim that the data-driven EG produces an error in the approximation of $\boldsymbol{x}^\star$ of at most 35% with probability of at least 95%. Note that this holds regardless the distribution of ξ and the dataset $\mathcal{D}_{1001}$ at hand.

Successively, we compare the theoretical bound ε provided by Theorem 1 with the actual distance $\|\boldsymbol{x}^{1001}-\boldsymbol{x}^\star\|$ (both scaled by $\|\boldsymbol{x}^\star\|$) by considering a small number of iterations $K = 101$ and several values for s. Also in this case, we perform 20 trials for each $s \in \{500, 600, \dots, 3000\}$. In Fig. 2, we note that the relative error produced by the data-driven EG in Algorithm 1, measured by the box plots, is quite significant (around 30%), yet almost equivalent for the different s considered. Consistently, the theoretical bound ε decreases as the number of samples adopted increases.

As a concluding remark, we note that, in general, the proposed certificates are not expected to be tight. The resulting numerical gap between the theoretical bound and the empirical evaluation–approximately one order of magnitude in the worst case, as seen in Fig. 1–is quite common for these type of probabilistic guarantees holding with high-confidence. Similar results have been observed in other domains, e.g., robust control design (Schildbach et al., 2014), or robustness evaluation (Fabiani et al., 2022).

5. Conclusion

We have derived rigorous probabilistic guarantees on the quality of the solution produced by a data-driven version of an EG scheme applied to SGNE seeking with limited data availability. Specifically, we have leveraged a finite number of uncertainty realizations to construct a sample average-based approximation for the game mapping, and exploited algorithmic stability arguments to bound the distance between the true SGNE and the solution obtained through the EG method. A distinct feature of our bounds is that they i) are easy to compute, ii) hold with high confidence, and iii) remain valid regardless of the underlying data distribution.

Future work will explore alternative techniques for solving SGNEP under the lens of algorithmic stability, potentially incorporating nonlinear operators to better handle both local and coupling constraints. Besides extending results for the EG considered here, prominent examples include algorithms based on standard operator splitting methods, such as forward-backward-forward or Douglas-Rachford.

CRediT authorship contribution statement

Filippo Fabiani: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing; **Barbara Franci:** Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. On the condition ensuring contraction

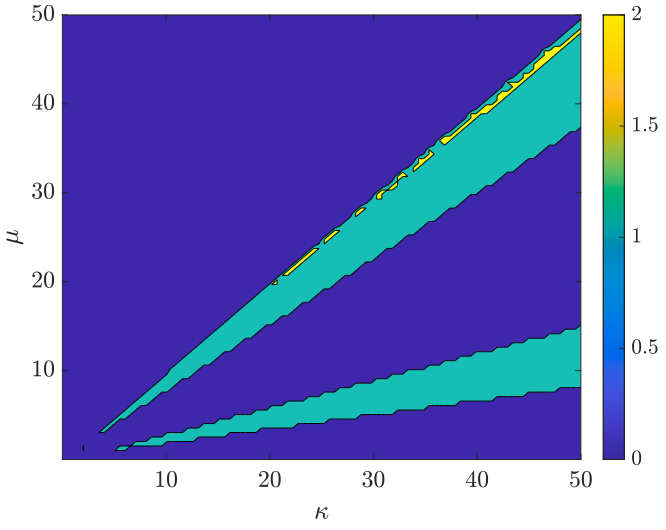

Fig. A.3. Map showing the feasibility (blue area), infeasibility (green area), or reached solver limits (yellow area) of (A.1) for 10^4 random combinations of $(\kappa, \mu) \sim \mathcal{U}(10^{-6}, 50)^2$.

In Fig. A.3 we report the binary values associated to the feasibility of the nonlinear optimization problem for the design of an optimal γ satisfying the condition for Theorem 1, i.e.,

$$\begin{aligned} \min_{\gamma \in (0, 2\mu/\kappa^2)} \quad & -\gamma \\ \text{s.t.} \quad & (1+\gamma\kappa)\sqrt{1-2\gamma\mu+\gamma^2\kappa^2}+\gamma\kappa < 1, \end{aligned} \tag{A.1}$$

for 100 randomly sampled values of $\kappa \sim \mathcal{U}(10^{-6}, 50)$ and $\mu \sim \mathcal{U}(10^{-6}, 50)$. From Fig. A.3, the feasible area covers around the 80.6% of the considered cases. Therefore, assuming the existence of some γ ensuring $\tau(\gamma) < 1$ appears a reasonable condition at the basis of our analysis.

References

Abada, I., Gabriel, S., Briat, V., & Massol, O. (2013). A generalized Nash–Cournot model for the northwestern European natural gas markets with a fuel substitution demand function: The GaMMES model. *Networks and Spatial Economics*, *13*, 1–42.

Alacaoglu, A., & Malitsky, Y. (2022). Stochastic variance reduction for variational inequality methods. In *Conference on learning theory* (pp. 778–816). PMLR.

Boţ, R. I., Mertikopoulos, P., Staudigl, M., & Vuong, P. T. (2021). Minibatch forward-backward-forward methods for solving stochastic variational inequalities. *Stochastic Systems*, *11*(2), 112–139.

Bousquet, O., & Elisseeff, A. (2002). Stability and generalization. *The Journal of Machine Learning Research*, *2*, 499–526.

Fabiani, F., & Caiti, A. (2019). Nash equilibrium seeking in potential games with double-integrator agents. In *2019 18th european control conference (ECC)* (pp. 548–553). IEEE.

Fabiani, F., & Franci, B. (2022). A stochastic generalized Nash equilibrium model for platforms competition in the ride-hail market. In *2022 IEEE 61st conference on decision and control (CDC)* (pp. 4455–4460). IEEE.

Fabiani, F., & Franci, B. (2026). Finite-sample guarantees for data-driven forward-backward operator methods. *IEEE Transactions on Automatic Control*, *71*(7), 4665–4677.

Fabiani, F., Margellos, K., & Goulart, P. J. (2022). Probabilistic feasibility guarantees for solution sets to uncertain variational inequalities. *Automatica*, *137*, 110120.

Facchinei, F., Fischer, A., & Piccialli, V. (2007). On generalized Nash games and variational inequalities. *Operations Research Letters*, *35*(2), 159–164.

Facchinei, F., & Pang, J.-S. (2007). Finite-dimensional variational inequalities and complementarity problems. Springer Science & Business Media.

Farnia, F., & Ozdaglar, A. (2021). Train simultaneously, generalize better: Stability of gradient-based minimax learners. In *International conference on machine learning* (pp. 3174–3185). PMLR.

Franci, B. (2025). On variance-reduced extragradient methods for stochastic generalized Nash equilibrium problems. *IEEE Control Systems Letters*, *9*, 2333–2338. https://doi.org/10.1109/LCSYS.2025.3615593

Franci, B., & Grammatico, S. (2020). A distributed forward–backward algorithm for stochastic generalized Nash equilibrium seeking. *IEEE Transactions on Automatic Control*, *66*(11), 5467–5473.

Franci, B., & Grammatico, S. (2021). Stochastic generalized Nash equilibrium-seeking in merely monotone games. *IEEE Transactions on Automatic Control*, *67*(8), 3905–3919.

Gower, R. M., Schmidt, M., Bach, F., & Richtárik, P. (2020). Variance-reduced methods for machine learning. *Proceedings of the IEEE*, *108*(11), 1968–1983.

Hardt, M., Recht, B., & Singer, Y. (2016). Train faster, generalize better: Stability of stochastic gradient descent. In *International conference on machine learning* (pp. 1225–1234). PMLR.

Henrion, R., & Römisch, W. (2007). On M-stationary points for a stochastic equilibrium problem under equilibrium constraints in electricity spot market modeling. *Applications of Mathematics*, *52*(6), 473–494.

Iusem, A. N., Jofré, A., Oliveira, R. I., & Thompson, P. (2017). Extragradient method with variance reduction for stochastic variational inequalities. *SIAM Journal on Optimization*, *27*(2), 686–724.

Korpelevich, G. M. (1976). The extragradient method for finding saddle points and other problems. *Matecon*, *12*, 747–756.

Koshal, J., Nedić, A., & Shanbhag, U. V. (2013). Regularized iterative stochastic approximation methods for stochastic variational inequality problems. *IEEE Transactions on Automatic Control*, *58*(3), 594–609.

Kutin, S., & Niyogi, P. (2002). Almost-everywhere algorithmic stability and generalization error. In *Proceedings of the 18th conference on uncertainty in artificial intelligence* (pp. 275–282).

Lei, J., & Shanbhag, U. V. (2022). Distributed variable sample-size gradient-response and best-response schemes for stochastic Nash equilibrium problems. *SIAM Journal on Optimization*, *32*(2), 573–603.

Ravat, U., & Shanbhag, U. V. (2011). On the characterization of solution sets of smooth and nonsmooth convex stochastic Nash games. *SIAM Journal on Optimization*, *21*(3), 1168–1199.

Romano, A. R., & Pavel, L. (2022). Exact convergence to GNE using penalty methods. In *2022 European control conference (ECC)* (pp. 2285–2290). IEEE.

Romano, A. R., & Pavel, L. (2023). An inexact-penalty method for GNE seeking in games with dynamic agents. *IEEE Transactions on Control of Network Systems*, *10*(4), 2072–2084.

Ryu, E. K., & Boyd, S. (2016). Primer on monotone operator methods. *Applied and Computational Mathematics*, *15*(1), 3–43.

Schildbach, G., Fagiano, L., Frei, C., & Morari, M. (2014). The scenario approach for stochastic model predictive control with bounds on closed-loop constraint violations. *Automatica*, *50*(12), 3009–3018.

Vidyasagar, M. (2013). Learning and generalisation: With applications to neural networks. Springer Science & Business Media.