K Y B E R N E T I K A — V O L U M E *59* (2 0 2 3) , N U M B E R 4 , P A G E S 5 7 5 – 5 9 1

BAYESIAN NASH EQUILIBRIUM SEEKING FOR MULTI-AGENT INCOMPLETE-INFORMATION AGGREGATIVE GAMES

Hanzheng Zhang, Huashu Qin, and Guanpu Chen

In this paper, we consider a distributed Bayesian Nash equilibrium (BNE) seeking problem in incomplete-information aggregative games, which is a generalization of either Bayesian games or deterministic aggregative games. We handle the aggregation function to adapt to incomplete-information situations. Since the feasible strategies are infinite-dimensional functions and lie in a non-compact set, the continuity of types brings barriers to seeking equilibria. To this end, we discretize the continuous types and then prove that the equilibrium of the derived discretized model is an ϵ-BNE. On this basis, we propose a distributed algorithm for an ϵ-BNE and further prove its convergence.

Keywords: aggregative games, Bayesian games, equilibrium approximation, distributed algorithms

Classification: 91A27, 91A43, 68W15

1. INTRODUCTION

In recent years, distributed design for multi-agent decision and control has become increasingly important and many distributed algorithms have been proposed for various games [4, 11, 14, 21, 22]. Aggregative games, as non-cooperative distributed games, are widely investigated. In aggregative games, each agent's cost function depends on its action and an aggregate of the decisions taken by all agents which is obtained via network communication. [11] proposed both synchronous and asynchronous distributed algorithms for aggregative games, and analyzed their convergence, while [14] considered coupled constraints in aggregative games and provided a distributed continuous-time algorithm for a generalized Nash equilibrium. In addition, [21] proposed a distributed approximation algorithm using inscribed polyhedrons to estimate local set constraints, while [10] proposed a distributed algorithm with multiple rounds of communication, and provided its linear convergence rate.

Considering uncertainties in reality, there are various incomplete-information games, and among them, Bayesian games are one of the most important and have a wide range of applications [1, 3, 5, 12]. In Bayesian games, players cannot obtain complete characteristics of the other players, which are called types subjected to a joint distribution.

DOI: 10.14736/kyb-2023-4-0575

Each player knows its own type and has access to the joint distribution [8]. In the investigation of various Bayesian games, the existence and computation of the Bayesian Nash equilibrium (BNE) are fundamental problems. Many works have studied the computation of BNE in the discrete-type Bayesian games [1, 2], by fixing the types and converting the games to deterministic ones. In addition to the centralized models, there are also many works on distributed Bayesian games [1, 13], where players make decisions based on their local information.

However, most of the aforementioned works focus on discrete-type Bayesian games. In fact, continuous-type Bayesian games are also widespread in various fields such as engineering and economics [5, 12]. Due to the continuity of types, it is hard to seek and verify BNE. Specifically, in these games, the feasible strategies are infinite-dimensional functions, and thus, their sets are not compact [6, 16]. Lack of compactness, we cannot apply the fixed point theorem for the existence of BNE, let alone seek a BNE. Fortunately, many works have tried to investigate the existence of BNE in such continuous-type situations and design the computation. For instance, [16] analyzed the existence of BNE in virtue of equicontinuous payoffs and absolutely continuous information, while [15] investigated the situation when the best responses are equicontinuous. Afterwards, [6] provided an equivalent condition of the equicontinuity and proposed an approximation algorithm. Besides, [20] regarded the BNE as a solution to a variational inequality and gave a sufficient condition of the existence of BNE, while [7] proposed two variational-inequality-based algorithms provided that the strategy forms are prior knowledge. [18] considered Bayesian games with finite actions and continuous types, and proposed a fictitious-play-based algorithm for the games with linear costs, while [9] proposed a discretization method for the finite-action and continuous-type model, and proved that the derived results converge to the BNE.

Therefore, using a Bayesian scheme to analyze an incomplete-information aggregative game is worth investigating, because it can be regarded as a generalization of either deterministic aggregative games [11, 14, 21] or Bayesian games [6, 8, 16]. Nevertheless, continuous-type aggregative Bayesian games are more challenging than both deterministic aggregative games and discrete-type Bayesian games. On the one hand, in the incomplete-information models, the strategies are functions of random variables, i. e., types. Thus, as the aggregate of strategies, the aggregation term should also be a function of types, while the existing aggregation functions for deterministic cases [11, 14, 21] cannot be directly applied to the incomplete-information cases. On the other hand, to seek a continuous-type BNE in a distributed manner, we need an effective method to convert the infinite-dimensional BNE-seeking problem into a finite-dimensional one, which also has to be friendly to distributed design.

Specifically, we consider seeking a continuous-type BNE in distributed aggregative games in this paper, where each agent has its own type following a joint distribution, and makes decisions based on its type and local information. Agents exchange their information and estimate the aggregate of all agents' decisions via time-varying graphs. The challenges lie in how to handle the aggregation term with incomplete-information and how to seek a BNE in this continuous-type model. Existing methods are insufficient to solve this problem. Although [5, 6, 7, 9] explored several ways to seek an approximate BNE, their approaches were limited in heuristic approximations and lacked the quantitative analysis of the derived results. Moreover, existing aggregative game models [11, 14, 21] cannot be directly applied to the incomplete-information cases. Considering the importance of the distributed aggregative Bayesian games, we focus on the aggregation functions which are well adapted to incomplete-information cases. To overcome bottlenecks in seeking a BNE, we provide a new discretization method that generates an ϵ-BNE with its quantitative analysis, and propose a distributed algorithm for an ϵ-BNE. The contributions are summarized as follows.

- We consider a distributed aggregative Bayesian game with continuous types, where each agent has access to its own type and the aggregate. Such generalized models can be regarded as not only Bayesian games [6, 8, 16] if each agent has access to the strategies of all agents, but also deterministic aggregative games [11, 14, 21] by letting out the uncertainties. Moreover, we focus on the incomplete-information aggregation functions when agents adopt non-constant-valued functions as strategies, which can turn to the average of strategies when types are deterministic [11, 21].

- We provide a BNE approximation method by discretizing the continuous types. By establishing a discretized model, we prove that the BNE of the derived model is an ϵ-BNE of the continuous-type model. Compared with existing methods [6, 7, 9] on continuous-type Bayesian games, our method provides an explicit error bound as well as a practical implementation different from heuristics.

- Based on the discretization, we propose a gradient-descent-based distributed algorithm for seeking a BNE of the discretized model, namely an ϵ-BNE of the original model. Furthermore, we prove that the proposed algorithm generates a sequence convergent to an ϵ-BNE of the continuous-type model using the Lyapunov theory.

The paper is arranged as follows. Section 2 summarizes the preliminaries. Section 3 formulates the problem. Section 4 provides a discretization method to generate an ϵ-BNE, while Section 5 gives a distributed algorithm for the derived ϵ-BNE and analyzes the convergence of the algorithm. Section 6 provides numerical simulations for illustrations. Finally, Section 7 concludes the paper.

2. PRELIMINARIES

2.1. Notations

Denote the n-dimensional real Euclidean space by $\mathbb{R}^n$ and its measure by μ. $B(a,\varepsilon)$ is a ball with the center a and the radius $\epsilon > 0$. Denote $col(x_1,\dots,x_n) = (x_1^T,\dots,x_n^T)^T$ and $1_n \in \mathbb{R}$ as the column vector with all entries equal to 1. For an integer $n > 0$, denote $[n] = \{1,\dots,n\}$. For column vectors $x, y \in \mathbb{R}^n$, $\langle x, y\rangle$ denotes the inner product, and $\|\cdot\|$ denotes the 2-norm. For a matrix $W \in \mathbb{R}^{n\times n}$, denote its element in the ith row and jth column by $[W]_{ij}$, $i, j \in [n]$. A function is piecewise continuous if it is continuous except at finite points in its domain. For $\boldsymbol{x} = (x_1,\dots,x_n)$, define the vector with entries of $\boldsymbol{x}$ except for i as $x_{-i} = (x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)$. For a surface $\boldsymbol{s} = \{(x_1,\dots,x_n)|S(x_1,\dots,x_n) = 0\}$, denote its differential by $\mathrm{d}\boldsymbol{s}$.

2.2. Convex analysis

A set $C \subseteq \mathbb{R}^n$ is convex if $\lambda z_1 + (1-\lambda)z_2 \in C$, $\forall\, z_1, z_2 \in C$ and $0 \leq \lambda \leq 1$. For a closed convex set $C \subseteq \mathbb{R}^n$, a projection map $\Pi_C : \mathbb{R}^n \to C$ is defined as $\Pi_C(x) = \arg\min_{y\in C}\|x - y\|$, and holds $\langle x - \Pi_C(x), \Pi_C(x) - y\rangle \geq 0$, $\forall\, y \in C$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is (strictly) convex if

$$f(\lambda x_1 + (1-\lambda)x_2)(<) \leq \lambda f(x_1) + (1-\lambda)f(x_2),\ \forall\, x_1, x_2 \in \mathbb{R}^n,\ \lambda \in (0,1).$$

For a convex differentiable function f, the gradient of f at point x is denoted by ∇f, satisfying $f(y) \geq f(x) + \langle y - x, \nabla f(x)\rangle$, $\forall\, y \in \mathbb{R}^n$. For a convex differentiable function $f(x_1, \ldots, x_n)$, denote $\nabla_i f$ as the differential of f with respect to x_i. If f is (strictly) convex, the gradient of f satisfies $\langle \nabla f(x) - \nabla f(y), x - y\rangle\, (>) \geq 0$, $\forall\, x \neq y$.

2.3. Bayesian games

Consider a Bayesian game denoted by $G = (I, \{X_i\}_{i\in I}, \boldsymbol{\Theta}, P(\cdot), \{f_i\}_{i\in I})$ with a set of agents $I = [n]$, where agent i has the feasible action set $X_i \subseteq \mathbb{R}^{m_i}$ and the cost function $f_i(x_i, x_{-i}, \theta_i)$. For $i \in I$, the incomplete information of agent i is referred to the type, denoted by $\theta_i \in \Theta_i \subseteq \mathbb{R}$, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \in \boldsymbol{\Theta}$ is a random variable mapping from the probability space $(\Omega, \mathcal{B}, P)$ to $\mathbb{R}^n$. Denote the density function of P by p with the marginal density $p_i(\theta_i) = \int_{\Theta_{-i}} p(\theta_i, \theta_{-i})\, \mathrm{d}\theta_{-i}$ and the conditional probability density $p_i(\theta_{-i}|\theta_i) = p(\theta_i, \theta_{-i})/p_i(\theta_i)$, $i \in [n]$. Throughout the paper, we use $\boldsymbol{\theta}$ to denote a random variable mapping from $(\Omega, \mathcal{B}, P)$ to $\mathbb{R}^n$, or a deterministic element in $\mathbb{R}^n$ depending on the context.

In Bayesian games, each agent $i \in I$ only knows its own type but not those of its rivals, and the joint distribution P is public information. The cost function of agent i is defined as $f_i : X_i \times X_{-i} \times \Theta_i \to \mathbb{R}$, depending on all agents' actions and the type of agent i. Each agent adopts a strategy σ_i, which is a measurable function mapping from its type set Θ_i to its action set X_i, and $\sigma_i(\theta_i)$ is the action taken by agent i when it receives the type $\theta_i \in \Theta_i$. Denote the feasible strategy set of agent i by Σ_i. For $i \in I$, define a Hilbert space $\mathcal{H}_i$ consisting of measurable functions $\beta : \mathbb{R} \to \mathbb{R}^{m_i}$ with the inner product $\langle \sigma, \sigma'\rangle_{\mathcal{H}_i} = \int_{\theta_i\in\Theta_i} \langle \sigma_i, \sigma'_i\rangle\, p_i(\theta_i)\, \mathrm{d}\theta_i,\ \sigma, \sigma' \in \mathcal{H}_i,\ i \in I$. Thus, the strategy set Σ_i is a subset of the Hilbert space $\mathcal{H}_i$.

2.4. Graph theory

An undirected graph $\mathcal{G}$ is defined by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = [n]$ is the node set and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set. Node j is a neighbor of i if $(j, i) \in \mathcal{E}$, and thus, node i is a neighbor of j. Take $(i, i) \in \mathcal{E}$. A path in $\mathcal{G}$ from i_1 to i_k is an alternating sequence $i_1 e_1 i_2 \cdots i_{k-1} e_{k-1} i_k$ such that $e_j = (i_j, i_{j+1}) \in \mathcal{E}$ for $j \in [k-1]$. $W = ([W]_{ij}) \in \mathbb{R}^{n\times n}$ is the adjacency matrix such that $[W]_{ij} > 0$ if $(j, i) \in \mathcal{E}$ and $[W]_{ij} = 0$ otherwise. $\mathcal{G}$ is connected if there exists a path in $\mathcal{G}$ from i to j for any nodes $i, j \in \mathcal{V}$.

3. PROBLEM FORMULATION

Consider an incomplete-information aggregative game, denoted by $G = (I, \{X_i\}_{i\in I}, \boldsymbol{\Theta}, P, \{f_i\}_{i\in I})$. Each agent $i \in I = [n]$ has its type $\theta_i \in \Theta_i \subseteq \mathbb{R}$, feasible action set $X_i \subseteq \mathbb{R}^m$, and cost function $f_i(x_i, \tilde{x}, \theta_i)$, where the joint type $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ follows the joint distribution $P(\boldsymbol{\theta})$ with the density $p(\boldsymbol{\theta})$, and $\tilde{x}$ is the aggregate of all agents' decisions. The type set Θ_i is compact and without loss of generalization, take $\Theta_i = [\underline{\theta}_i, \overline{\theta}_i]$. Agent i adopts a strategy σ_i, which is a measurable function from its type set Θ_i to its action set X_i. Denote the strategy set of agent i by Σ_i.

Different from deterministic games, consider the aggregation functions in incomplete-information situations where agents take non-constant-valued functions of types as strategies, which are functions mapping to different values for different types, rather than vectors in deterministic games. Here, the aggregation function is shown as follows.

$$\bar{\sigma}(\tilde{\theta}) = \int_{\sum_{i=1}^{n} \theta_i = n\tilde{\theta}} \frac{\sum_{i=1}^{n} \sigma_i(\theta_i)}{n} \bar{p}(\theta_1, \dots, \theta_n | \tilde{\theta}) \, \mathrm{d}\boldsymbol{s}, \tag{1}$$

where $\boldsymbol{s} = \{(\theta_1, \dots, \theta_n) | \sum_{i=1}^{n} \theta_i = n\tilde{\theta}\}$ and $\tilde{\theta} = (\theta_1 + \cdots + \theta_n)/n$ is the average type following the distribution

$$\bar{P}(\tilde{\theta}) = \int_{\sum_{i=1}^{n} \theta_i = n\tilde{\theta}} p(\theta_1, \dots, \theta_n) \, \mathrm{d}\boldsymbol{s},$$

with the density function $\bar{p}(\tilde{\theta})$ and the conditional probability density function $\bar{p}(\tilde{\theta}|\theta_i) = \int_{\sum_{j \neq i} \theta_j = n\tilde{\theta} - \theta_i} p(\theta_i, \theta_{-i})/p_i(\theta_i) \, \mathrm{d}s_{-i}$. Note that $\bar{\sigma}$ is a linear function with respect to $\sigma_1, \dots, \sigma_n$, denoted by $H(\sigma_1, \dots, \sigma_n)$.

Remark 3.1. The aggregation function (1) can be regarded as the average of the agents' strategies. When the model is deterministic, which means that $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_0\}$ and $P(\boldsymbol{\theta}_0) = 1$, the aggregation function (1) turns to the simple average of agents' strategies, as [11, 21].

Example 3.2. Consider a Nash-Cournot game [11, 14], where firms compete to produce a kind of commodity. For each firm i, the price of the raw materials θ_i is a continuous random variable. All firms cannot know the exact type of other firms, and instead, they know the distribution of the average price of the raw materials $\tilde{\theta}$. The cost of each firm is a function of the price of its raw materials θ_i, its production x_i, and the average production of all firms $\bar{x}$.

With the above aggregation function, the goal of agent i is to minimize the following conditional expectation of f_i

$$\begin{aligned} U_i(\sigma_i, \bar{\sigma}, \theta_i) &= \int_{\tilde{\Theta}} f_i(\sigma_i(\theta_i), \bar{\sigma}(\tilde{\theta}|\theta_i), \theta_i) \bar{p}(\tilde{\theta}|\theta_i) \, \mathrm{d}\tilde{\theta} \\ &= \int_{\Theta_{-i}} f_i(\sigma_i(\theta_i), H(\sigma_i(\theta_i), \sigma_{-i}(\theta_{-i})), \theta_i) p_i(\theta_{-i}|\theta_i) \, \mathrm{d}\theta_{-i}, \end{aligned}$$

where $\bar{\sigma}(\tilde{\theta}|\theta_i)$ is the aggregate $\bar{\sigma}$ when the average type is $\tilde{\theta} \in \tilde{\Theta}$ and agent i's type is $\theta_i \in \Theta_i$. Denote its gradient by

$$F_i(\sigma_i, \bar{\sigma}, \theta_i) = \nabla_i U_i(\sigma_i, H(\sigma_i, \sigma_{-i}), \theta_i).$$

Then we give the concept of the Bayesian Nash equilibrium.

Definition 3.3. Consider the continuous-type Bayesian game G,

(i) For agent $i \in I$, a strategy σ_{i*} is a best response with respect to the strategy profile σ_{-i} if for almost every $\theta_i \in \Theta_i$,

$$\sigma_{i*} \in \arg\min_{\sigma_i \in \Sigma_i} U_i(\sigma_i, H(\sigma_i, \sigma_{-i}), \theta_i),$$

where the best response set is denoted by $BR_i(\sigma_{-i})$.

(ii) A strategy profile $(\sigma_1^*, \dots, \sigma_n^*)$ is a Bayesian Nash equilibrium (BNE) if for any $\sigma_i \in \Sigma_i$, $i \in I$,

$$U_i(\sigma_i^*, H(\sigma_i^*, \sigma_{-i}^*), \theta_i) \leq U_i(\sigma_i, H(\sigma_i, \sigma_{-i}^*), \theta_i), \text{ for a.e. } \theta_i \in \Theta_i.$$

Agents in our model have local interactions with each other over time to estimate the aggregate, where these interactions are modeled by time-varying graphs $\mathcal{G}(t)$. At the time t, agents exchange their estimations of the aggregate with current neighbors through $\mathcal{G}(t)$.

We make the following assumptions for the aggregative game G.

Assumption 3.4. Consider the incomplete-information aggregative game G. For $i \in I$,

(i) the action set X_i is nonempty, convex, and compact;

(ii) the distribution P is atomless, i. e., $P(\boldsymbol{\theta} = \boldsymbol{\zeta}) = 0$ for any given $\boldsymbol{\zeta} \in \boldsymbol{\Theta}$. Moreover, the measure $\mu(\{\theta_i \in \Theta_i | p_i(\theta_i) > 0\}) = \mu(\Theta_i)$;

(iii) the cost function $f_i(x_i, \tilde{x}, \theta_i)$ is strictly convex in $x_i \in X_i$ and L_θ-Lipschitz continuous in $\theta_i \in \Theta_i$ for each $x_i, \tilde{x} \in \mathbb{R}^m$;

(iv) the expectation U_i is well defined for every $\sigma_j \in \Sigma_j$ and $\theta_i \in \Theta_i$, $j \in I$, and for any $\theta_i \in \Theta_i$, its gradient F_i is D-Lipschitz continuous in $\sigma_i(\theta_i) \in X_i$ for any $\sigma_{-i} \in \Sigma_{-i}$, and is L_u-Lipschitz continuous in $\bar{\sigma}$ for any $\sigma_i(\theta_i) \in X_i$;

(v) the graph sequence $\mathcal{G}(t)$ is uniformly jointly strongly connected, i. e., there exists an integer $\mathcal{B} > 0$ such that $\cup_{k=t}^{t+\mathcal{B}} \mathcal{G}(k)$ is strongly connected, and its adjacency matrix $W(t)$ satisfies $[W(t)]_{ij} > \eta$ $(\eta > 0)$ when $(j, i) \in \mathcal{E}$, and $\sum_{i=1}^n [W(t)]_{ij} = \sum_{j=1}^n [W(t)]_{ij} = 1$.

Assumption 3.4 was widely used in the study of aggregative games and Bayesian games [6, 7, 11, 16, 22]. The atomless property in Assumption 3.4(ii) is a common assumption in Bayesian games [6, 15, 16], and the measure condition can be guaranteed by removing types in $\{\theta_i \in \Theta_i, \exists \varepsilon > 0, p_i(\theta_i') > 0, \forall \theta_i' \in B(\theta_i, \varepsilon)\}$. Assumption 3.4(v) ensures the connectivity, and the property of the adjacency matrix $W(t)$ ensures that each agent can substantially and equally influence the aggregation term in the long run, which holds for a variety of graphs[11, 17].

The existence of the BNE can be guaranteed by the variational inequalities [7, 20], summarized as follows.

Lemma 3.5. Under Assumption 3.4, there exists a unique BNE of game G.

In Bayesian games, it is hard to seek a BNE due to the continuity of types. As the Riesz's Lemma shows [19], any infinite-dimensional normed space contains a sequence of unit vectors $\{x_n\}$ with $\|x_n - x_m\| > \alpha$ for any $0 < \alpha < 1$ and $n \neq m$. Then the strategy set $\mathbf{\Sigma}$ lying in the infinite-dimensional space $\mathcal{H}_1 \times \cdots \times \mathcal{H}_n$ is not compact, which poses obstacles to computation. There are a few attempts to seek a continuous-type BNE. For example, [7] considered the situation that the strategy forms are prior knowledge, in which the forms are usually unavailable, while [6] utilized polynomial approximations to estimate a BNE without the estimation error. Moreover, [9] adopted heuristic approximations in discrete-action Bayesian games, but their method was NP-hard and not practical to be implemented in continuous-action games. Thus, since directly seeking a BNE is difficult, we introduce the following concept.

Definition 3.6. Denote $EU_i(\sigma_i, \bar{\sigma}) = \int_{\Theta_i} U_i(\sigma_i, \bar{\sigma}, \theta_i) p_i(\theta_i) \, \mathrm{d}\theta_i$. For any $\epsilon > 0$, a strategy profile $\hat{\boldsymbol{\sigma}}^* = (\hat{\sigma}_1^*, \ldots, \hat{\sigma}_n^*)$ is an ϵ-Bayesian Nash equilibrium (ϵ-BNE) of G if for any $\sigma_i \in \Sigma_i$, $i \in I$

$$EU_i(\sigma_i, H(\sigma_i, \hat{\sigma}_{-i}^*)) \geq EU_i(\hat{\sigma}_i^*, H(\hat{\boldsymbol{\sigma}}^*)) - \epsilon.$$

The goal of this paper is to design a distributed algorithm to seek an approximate BNE of the proposed model, summarized as follows.

Problem 3.7. Seek an approximate BNE of the incomplete-information aggregative game $G = (I, \{X_i\}_{i \in I}, \mathbf{\Theta}, P, \{f_i\}_{i \in I})$ in a distributed manner.

To overcome the above bottlenecks in seeking BNE, in the following sections, we first provide a discretization method to convert the infinite-dimensional problem to a finite-dimensional one, and then propose a distributed algorithm for an ϵ-BNE.

4. DISCRETIZATION

In this section, we give a discretization method and show its effectiveness in approximating the best responses and the BNE of the continuous-type model G.

For each agent i, we select N points θ_i^k from Θ_i satisfying

$$P_i(\theta_i^k) = \frac{k}{N}, \ k \in [N].$$

Denote the corresponding discrete type set by $\hat{\Theta}_i$. Define $\theta_i^0 = \underline{\theta}_i$ and $\hat{\mathbf{\Theta}} = \hat{\Theta}_1 \times \cdots \times \hat{\Theta}_n$. In the discretized model, we regard all types in the interval $(\theta_i^{k-1}, \theta_i^k]$ as θ_i^k, then the discrete type $\boldsymbol{\theta}$ follow the below joint distribution

$$\hat{P}(\theta_1^{k_1}, \ldots, \theta_n^{k_n}) = \int_{\theta_1^{k_1 - 1}}^{\theta_1^{k_1}} \cdots \int_{\theta_n^{k_n - 1}}^{\theta_n^{k_n}} p(\theta_1, \ldots, \theta_n) \, \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_n, \ (\theta_1^{k_1}, \ldots, \theta_n^{k_n}) \in \hat{\mathbf{\Theta}}. \quad (2)$$

Correspondingly, the marginal distribution $\hat{P}_i(\theta_i) = \sum_{\theta_i \in \hat{\Theta}_i} \hat{P}(\theta_i, \theta_{-i})$ and the conditional distribution $\hat{P}_i(\theta_{-i} | \theta_i) = \hat{P}(\theta_i, \theta_{-i}) / \hat{P}_i(\theta_i)$.

Due to Assumption 3.4(ii) that $\mu(\{\theta_i \in \Theta_i | p_i(\theta_i) > 0\}) = \mu(\Theta_i)$, the gap between the adjacent discrete points $\theta_i^k - \theta_i^{k-1}$ tends to 0 as N tends to infinity. Since we use θ_i^k to represent the interval $(\theta_i^k, \theta_i^{k-1}]$, we choose the length of such intervals as small as possible, which can effectively reduce the error. Additionally, our design is friendly to distributed algorithms, while other choices of discrete points will bring extra computation when agents update the strategies.

Based on the above discretization, we formulate a discretized model as $\hat{G} = (I, \{X_i\}_{i\in I}, \hat{\Theta}, \hat{P}, \{f_i\}_{i=1}^n)$. In this model, strategies are restricted to N-dimensional vectors. Denote the strategy set of agent i in $\hat{G}$ by $\hat{\Sigma}_i$. Thus, the aggregate of the discretized strategies is

$$\hat{\bar{\sigma}}(\tilde{\theta}) = \sum_{\sum_{i=1}^n \theta_i = n\tilde{\theta}, \theta_i \in \hat{\Theta}_i} \frac{\sum_{i=1}^n \hat{\sigma}_i(\theta_i)}{n} \hat{\bar{P}}(\theta_1, \ldots, \theta_n|\tilde{\theta}),$$

where $\tilde{\theta} = \sum_{i=1}^n \theta_i/n$ follows the below discrete distribution

$$\hat{\bar{P}}(\tilde{\theta}) = \sum_{\sum_{i=1}^n \theta_i = n\tilde{\theta}, \theta_i \in \hat{\Theta}_i} \hat{P}(\theta_1, \ldots, \theta_n),$$

with the conditional probability $\hat{\bar{P}}(\cdot|\tilde{\theta})$. Since the $\hat{\bar{P}}$ is a discrete distribution, the aggregation function can be written as $\hat{\bar{\sigma}} = \sum_{i=1}^n h_i(\hat{\sigma}_i)$, where h_i is a linear function mapping from $\hat{\Sigma}_i$ to $\mathbb{R}^{nNm}$. Denote $h_{-i}(\hat{\sigma}_{-i}) = \sum_{j\neq i} h_j(\hat{\sigma}_j)$ for $\hat{\sigma}_{-i} \in \hat{\Sigma}_{-i}$, then the expectation of the cost of i in $\hat{G}$ is

$$\hat{U}_i(\hat{\sigma}_i, \hat{\bar{\sigma}}, \theta_i) = \sum_{\theta_{-i}\in\hat{\Theta}_{-i}} f_i(\hat{\sigma}_i(\theta_i), h_i(\hat{\sigma}_i(\theta_i)) + h_{-i}(\hat{\sigma}_{-i}(\theta_{-i})), \theta_i)\hat{P}_i(\theta_{-i}|\theta_i),\ \theta_i \in \hat{\Theta}_i.$$

Then we define the following best response and equilibrium of $\hat{G}$.

Definition 4.1. Consider the discretized model $\hat{G}$,

(i) For agent $i \in I$, given a strategy profile $\hat{\sigma}_{-i} \in \hat{\Sigma}_{-i}$, a strategy $\hat{\sigma}_{i*}^N$ is a best response to σ_{-i} if for any $\theta_i \in \hat{\Theta}_i$,

$$\hat{\sigma}_{i*}^N = \arg \min_{\hat{\sigma}(\theta_i)\in X_i} \hat{U}_i(\sigma_i, h_i(\hat{\sigma}_i) + h_{-i}(\hat{\sigma}_{-i}), \theta_i).$$

Denote the set of the best responses to $\hat{\sigma}_{-i}$ by $BR_i^N(\hat{\sigma}_{-i})$ in $\hat{G}$.

(ii) A strategy profile $(\hat{\sigma}_1^*, \ldots, \hat{\sigma}_n^*)$ is a BNE of $\hat{G}$, or a DBNE(N) of $\hat{G}$ if for $i \in I$,

$$\hat{\sigma}_i^* \in BR_i^N(\hat{\sigma}_{-i}^*).$$

The existence of DBNE can be guaranteed by variational inequalities [7, 20] or Browner fixed point theorem [6, 16], summarized as follows.

Lemma 4.2. Under Assumption 3.4, there exists a unique DBNE(N) of the discretized model $\hat{G}$.

To approximate the strategies in the continuous-type model G with the strategies in the discretized model $\hat{G}$, we extend the domains of strategies from $\hat{\Theta}_i$ to Θ_i, and define the strategies of $\hat{G}$ at type $\theta_i \in (\theta_i^{k-1}, \theta_i^k]$ as

$$\hat{\sigma}_i(\theta_i) = \hat{\sigma}_i(\theta_i^k).$$

Denote $\hat{\tilde{\Theta}} = \{\tilde{\theta} | n\tilde{\theta} = \sum_{i=1}^n \theta_i, \theta_i \in \Theta_i\} = \{\tilde{\theta}^1, \dots, \tilde{\theta}^S\}$, $S \le nN$, and $\tilde{\theta}^0 = \sum_{i=1}^n \underline{\theta}_i / n$. The aggregates in the discretized model $\hat{G}$ and the continuous-type model G satisfy

$$\bar{\sigma}(\tilde{\theta}) = \hat{\bar{\sigma}}(\tilde{\theta}^k),\ \tilde{\theta} \in (\tilde{\theta}^{k-1}, \tilde{\theta}^k],\ k \in [S].$$

With this extension, we denote $\hat{\bar{\sigma}}$ of the discretized strategies as $\bar{\sigma}$ for simplification.

Then we estimate the best response in G. Actually, a best response $\hat{\sigma}_{i*}$ needs to respond to any strategies in Σ_{-i}, rather than strategies in $\hat{\Sigma}_{-i} \subseteq \Sigma_{-i}$. To this end, we modify the best responses $\hat{\sigma}_{i*}$ with respect to $\sigma_{-i} \in \Sigma_{-i}$ as follows. For $\theta_i^k \in \hat{\Theta}_i$,

$$\hat{\sigma}_{i*}(\theta_i^k) = \arg \min_{\hat{\sigma}_i(\theta_i) \in X_i} \int_{\Theta_{-i}} f_i(\hat{\sigma}(\theta_i^k), H(\hat{\sigma}_i(\theta_i^k), \sigma_{-i}(\theta_{-i})), \theta_i^k) \frac{\int_{\theta_i^{k-1}}^{\theta_i^k} p(\theta_i, \theta_{-i})\, \mathrm{d}\theta_i}{\hat{P}_i(\theta_i^k)}\, \mathrm{d}\theta_{-i}.$$

The following lemma shows the relation between the best response in the discretized model $\hat{G}$ and in the continuous-type model G.

Lemma 4.3. Let Assumption 3.4(ii) hold. For a given $\sigma_{-i} \in \Sigma_{-i}$, if all the best responses in $BR_i^N(\sigma_{-i})$ of G are piecewise continuous, then the best responses in $BR_i^N(\sigma_{-i})$ of $\hat{G}$ are almost surely the best responses of G, as N tends to infinity. Specifically, for any $\hat{\sigma}_{i*}^N \in BR_i^N(\sigma_{-i})$, there exists $\sigma_{i*} \in BR_i(\sigma_{-i})$ such that

$$\lim_{N\to\infty} \hat{\sigma}_{i*}^N(\theta_i) = \sigma_{i*}(\theta_i),\ \text{for a.e. } \theta_i \in \Theta_i.$$

Proof. Due to the L'Hospital's rule,

$$\lim_{N\to\infty} \frac{\int_{\theta_i^{r-1}}^{\theta_i^k} p(\theta_i, \theta_{-i})\, \mathrm{d}\theta_i}{\int_{\theta_i^{r-1}}^{\theta_i^k} p_i(\theta_i)\, \mathrm{d}\theta_i} = \frac{p(\theta_i^k, \theta_{-i})}{p_i(\theta_i^k)} = p_i(\theta_{-i} | \theta_i^k).$$

Thus, for any $\hat{\sigma}_{i*} \in BR_i^N(\sigma_{-i})$ and $\theta_i \in \hat{\Theta}_i$, there exists a strategy $\sigma_{i*} \in BR_i(\sigma_{-i})$ such that, as N tends to infinity, $\hat{\sigma}_{i*}^N(\theta_i) \to \sigma_{i*}$. Since σ_{i*} is piecewise continuous, for any $\varepsilon > 0$, there exists $\delta > 0$ such that, for any $\theta_i \in \Theta_i$ except for finite points and $\theta_i' \in B(\theta_i, \delta) \cap \Theta_i$, $|\sigma_{i*}(\theta_i) - \sigma_{i*}(\theta_i')| < \varepsilon$. Take $\varepsilon = \max_k \{\theta_i^k - \theta_i^{k-1}\}$, thus

$$\int_{\Theta_i} |\hat{\sigma}_{i*}^N(\theta_i) - \sigma_{i*}(\theta_i)|^2 p_i(\theta_i)\, \mathrm{d}\theta_i \le \sum_{i=1}^N \varepsilon^2 (\theta_i^k - \theta_i^{k-1}) = \varepsilon^2 (\overline{\theta}_i - \underline{\theta}_i).$$

As N tends to infinity, ε tends to 0, and thus, $\hat{\sigma}_{i*}^N$ tends to σ_{i*} for almost every $\theta_i \in \Theta_i$. □

Lemma 4.3 implies that agents can utilize the best responses derived from $\hat{G}$ to estimate the best responses in G. That is to say, agents are willing to adopt the best responses in $\hat{G}$, and thus, these best responses form a DBNE.

In the next theorem, we show that the DBNE(N) of $\hat{G}$ is an ϵ-BNE of G. In addition, we provide an explicit error bound of our approximation, compared with heuristic approximations [6, 7, 9].

Theorem 4.4. Under Assumption 3.4(i), (ii), and (iv), the DBNE $(\hat{\sigma}_1^*, \ldots, \hat{\sigma}_n^*)$ of the discretized model $\hat{G}$ is an ϵ-BNE, where $\epsilon = O(\max_{i\in I} \max_{k\in[N]}(\theta_i^k - \theta_i^{k-1}))$.

Proof. Due to the Lipschitz continuity of f_i, $|f_i(x_i, x_{-i}, \theta_i) - f_i(x_i, x_{-i}, \theta_i')| \le L_\theta \|\theta_i - \theta_i'\|$ for $x_i \in X_i,\ x_{-i} \in X_{-i},\ \theta_i, \theta_i' \in \Theta_i$. Denote $E\hat{U}_i(\hat{\sigma}_i, \bar{\sigma}) = \sum_{\hat{\Theta}_i} \hat{U}_i(\hat{\sigma}_i, \bar{\sigma}, \theta_i)\hat{P}_i(\theta_i)$. Define $\epsilon_0 = \max_{i\in I} \max_{k\in[N]}(\theta_i^k - \theta_i^{k-1})$. Then for any $\hat{\sigma}_i \in \hat{\Sigma}_i$ and $\hat{\sigma}_{-i} \in \hat{\sigma}_{-i}$,

$$\begin{aligned} EU_i(\hat{\sigma}_i, \bar{\sigma}) =& \sum_{k=1}^{N} \int_{\theta_i^{k-1}}^{\theta_i^k} U_i(\hat{\sigma}_i, \bar{\sigma}, \theta_i) p_i(\theta_i) \, \mathrm{d}\theta_i \\ =& \sum_{k=1}^{N} \int_{\theta_i^{k-1}}^{\theta_i^k} (U_i(\hat{\sigma}_i, \bar{\sigma}, \theta_i) - U_i(\hat{\sigma}_i, \bar{\sigma}, \theta_i^k) + U_i(\hat{\sigma}_i, \bar{\sigma}, \theta_i^k)) p_i(\theta_i) \, \mathrm{d}\theta_i \\ \le& E\hat{U}_i(\hat{\sigma}_i, \bar{\sigma}) + L_\theta \epsilon_0. \end{aligned} \tag{3}$$

Denote the aggregate of the DBNE by $\hat{\bar{\sigma}}^*$. Recalling the definition of DBNE, for any $\theta_i \in \hat{\Theta}_i$ and $\hat{\sigma}_i \in \hat{\sigma}_i$,

$$\hat{U}_i(\hat{\sigma}_i^*, \hat{\bar{\sigma}}^*, \theta_i) \le \hat{U}_i(\hat{\sigma}_i, h_i(\hat{\sigma}_i) + h_{-i}(\hat{\sigma}_{-i}^*), \theta_i). \tag{4}$$

Then we convert the DBNE from the discretized model to the continuous-type model. Due to Assumption 3.4(ii), the distribution P is continuous, and thus, p is L_p-Lipschitz continuous over $\mathbf{\Theta}$. Therefore, for any $\theta_i \in (\theta_i^{k-1}, \theta_i^k]$, $k \in [N]$,

$$\begin{aligned} &\left| p(\theta_i, \theta_{-i}) - \int_{\theta_i^{k-1}}^{\theta_i^k} \frac{p(\theta_i', \theta_{-i}) \, \mathrm{d}\theta_i'}{\theta_i^k - \theta_i^{k-1}} \right| \\ \le& \frac{1}{\theta_i^k - \theta_i^{k-1}} \int_{\theta_i^{k-1}}^{\theta_i^k} |p(\theta_i, \theta_{-i}) - p(\theta_i', \theta_{-i})| \, \mathrm{d}\theta_i' \le L_p \epsilon_0. \end{aligned}$$

Thus, for any $\sigma_i \in \Sigma_i$ and $k \in [N]$,

$$\begin{aligned} &\int_{\theta_i^{k-1}}^{\theta_i^k} \int_{\Theta_{-i}} f_i(\sigma_i(\theta_i), H(\sigma_i(\theta_i), \hat{\sigma}_{-i}^*(\theta_{-i})), \theta_i^k) p(\theta_i, \theta_{-i}) \, \mathrm{d}\theta_i \mathrm{d}\theta_{-i} \\ \ge& \int_{\theta_i^{k-1}}^{\theta_i^k} \int_{\Theta_{-i}} f_i(\sigma_i(\theta_i), H(\sigma_i(\theta_i), \hat{\sigma}_{-i}^*(\theta_{-i})), \theta_i^k) \frac{\int_{\theta_i^{k-1}}^{\theta_i^k} p(\theta_i', \theta_{-i}) \, \mathrm{d}\theta_i'}{\theta_i^k - \theta_i^{k-1}} \, \mathrm{d}\theta_i \mathrm{d}\theta_{-i} \\ &- M L_p (\theta_i^k - \theta_i^{k-1}) \mu(\Theta_{-i}) \epsilon_0. \end{aligned} \tag{5}$$

With (4) and (5), for any $\sigma_i \in \Sigma_i$,

$$\begin{aligned}
&E\hat{U}_i(\hat{\sigma}_i^*, \hat{\bar{\sigma}}^*)\\
\leq&\sum_{k=1}^{N}\bigg(\int_{\theta_i^{k-1}}^{\theta_i^k}\int_{\Theta_{-i}} f_i(\sigma_i(\theta_i), H(\sigma_i(\theta_i), \hat{\sigma}_{-i}^*(\theta_{-i})), \theta_i^k)p(\theta_i, \theta_{-i})\,\mathrm{d}\theta_i\mathrm{d}\theta_{-i}\\
&+ML_p(\theta_i^k-\theta_i^{k-1})\mu(\Theta_{-i})\epsilon_0)\bigg)\\
\leq&\sum_{k=1}^{N}\bigg(\int_{\theta_i^{k-1}}^{\theta_i^k}\int_{\Theta_{-i}} (f_i(\sigma_i(\theta_i), H(\sigma_i(\theta_i), \hat{\sigma}_{-i}^*(\theta_{-i})), \theta_i)+L_\theta\epsilon_0)p(\theta_i, \theta_{-i})\,\mathrm{d}\theta_i\mathrm{d}\theta_{-i}\\
&+ML_p(\theta_i^k-\theta_i^{k-1})\mu(\Theta_{-i})\epsilon_0\bigg)\\
=&EU(\sigma_i, \hat{\bar{\sigma}}^*)+(ML_p\mu(\mathbf{\Theta})+L_\theta)\epsilon_0.
\end{aligned} \tag{6}$$

Define $C = ML_p\mu(\mathbf{\Theta}) + 2L_\theta$. Combining (3) and (6), for any $\sigma_i \in \Sigma_i$,

$$EU(\hat{\sigma}_i^*, \hat{\bar{\sigma}}^*) \leq E\hat{U}(\hat{\sigma}_i^*, \hat{\bar{\sigma}}^*) + L_\theta\epsilon_0 \leq EU_i(\sigma_i, H(\sigma_i, \hat{\sigma}_{-i}^*)) + C\epsilon_0,$$

which means that the DBNE is an ϵ-BNE with $\epsilon = C\epsilon_0$. □

Remark 4.5. Taking advantage of the probability $\frac{1}{\theta_i^k-\theta_i^{k-1}}\int_{\theta_i^{k-1}}^{\theta_i^k} p(\theta_i', \theta_{-i})\,\mathrm{d}\theta_i'$, we convert the discrete distribution $\hat{P}$ of $\hat{G}$ into a continuous distribution of G, thereby overcoming the challenges in quantitative analysis. Compared with existing methods which were limited in heuristic approximations [6, 7, 9], we quantitatively analyze the derived DBNE of $\hat{G}$ and provide the explicit error bound of the DBNE. Moreover, our discretization is a practical method rather than heuristics.

Remark 4.6. Since we discretize the continuous types, the dimension of the aggregation term in the discretized model $\hat{G}$ depends on the number of discrete points. For higher accuracy, we need to take more discrete points, which brings barriers to distributed computation when agents exchange such high-dimensional aggregation terms. In fact, these bottlenecks are common in discretization methods [7, 9], and we are still trying to handle this challenge.

5. DISTRIBUTED ALGORITHM

In this section, we propose a distributed algorithm for the BNE of the discretized model $\hat{G}$, namely an ϵ-BNE of the continuous-type model G.

At the time t, agent i estimates the aggregate according to neighbors' approximations as

$$u_i^t = \sum_{j=1}^{n}[W(t)]_{ij}v_j^t, \tag{7}$$

where v_j^t is the approximation of the aggregate made by agent j at the time t. Due to the uncertainties, we can also regard $u_i^t(\tilde{\theta})$ as a function mapping from $\hat{\tilde{\Theta}}$ to $\mathbb{R}^m$, where

$\tilde{\theta}$ follows the discrete distribution $\hat{\hat{P}}$ as defined in Section 4. Then agent i evaluates its subgradient as

$$g_i^t = (F_i(\sigma_i^t, u_i^t, \theta_i^1), \ldots, F_i(\sigma_i^t, u_i^t, \theta_i^N))/N, \tag{8}$$

where $F_i(\sigma_i^t, u_i^t, \theta_i^k)$ $(k \in [N])$ was defined in Section 3. We summarize the above procedures as follows.

Algorithm 1 Algorithm for an ϵ-BNE of continuous-type Bayesian games.

Initialization: For $i \in I$: take $\sigma_i(0) \in \widetilde{\Sigma}_i$, and $v_i(0) = h_i(\sigma_i(0))$.
Discretization: For $i \in I$, take N discrete points from the type set Θ_i as (2).
Iterate until $t \geq T$:
Communicate and Update: Agent i evaluates the aggregate of neighbors $u_i(t)$ based on (7) and the gradient g_i^t based on (8), then updates $\sigma_i(t)$ and its observation $v_i(t)$ by

$$\sigma_i^{t+1} = \Pi_{\hat{\Sigma}_i}(\sigma_i^t - \alpha(t) g_i^t),$$
$$v_i^{t+1} = u_i^t - n(h_i(\sigma_i^t) + h_i(\sigma_i^{t+1})).$$

The stepsize $\alpha(t)$ taken in Algorithm 1 satisfies

(i) $\alpha(t)$ is a positive non-increasing sequence.

(ii) $\sum_{t=0}^{\infty} \alpha(t) = \infty$, $\sum_{t=0}^{\infty} \alpha^2(t) < \infty$.

Therefore, we give the following main result of this paper to show the convergence of Algorithm 1 to an ϵ-BNE, or the DBNE(N) with an explicit error bound ϵ.

Theorem 5.1. Under Assumption 3.4, Algorithm 1 generates a sequence convergent to the DBNE of $\hat{G}$, which is an ϵ-BNE of G, with $\epsilon = O(\max_{i \in I} \max_{k \in [N]} (\theta_i^k - \theta_i^{k-1}))$.

Proof. Define the average of the estimations v_i^t as $\bar{v}^t = \sum_{i=1}^n v_i^t / n$. Firstly, we prove that $\bar{v}^t = \sum_{i=1}^n h_i(\sigma_i^t)$ by induction on t.

For $t = 0$, the above relation holds trivially. Since the adjacency matrices have columns sum up to 1, assume the above relation holds for $t-1$, as the induction step,

$$\begin{aligned}
\sum_{i=1}^n \frac{v_i^t}{n} &= \sum_{i=1}^n \left(\frac{u_i^{t-1}}{n} + h_i(\sigma_i^t) - h_i(\sigma_i^{t-1}) \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n [W(t-1)]_{ij} \left(\frac{v_j^{t-1}}{n} + h_j(\sigma_j^t) - h_j(\sigma_j^{t-1}) \right) \\
&= \sum_{i=1}^n \left(\frac{v_i^{t-1}}{n} + h_i(\sigma_i^t) - h_i(\sigma_i^{t-1}) \right) = \sum_{i=1}^n h_i(\sigma_i^t).
\end{aligned}$$

Thus, $\bar{v}^t = \sum_{i=1}^n h_i(\sigma_i^t)/n$ holds for all $t \geq 0$.

Secondly, we establish a relation between the estimation u_i^t and the average $\bar{v}^t$. From the update rule of v_i^t, for $t \geq 1$,

$$u_i^t = \sum_{j=1}^{n} [\Phi(t,0)]_{ij} v_j^0 + \sum_{r=1}^{t} \sum_{j=1}^{n} [\Phi(t,r)]_{ij} n(h_j(\sigma_j^r) - h_j(\sigma_j^{r-1})). \tag{9}$$

Then, for $t \geq 1$, $\bar{v}^t$ can be reconstructed as

$$\bar{v}^t = \frac{1}{n} \sum_{j=1}^{n} v_j^0 + \sum_{r=1}^{t} \sum_{i=1}^{n} (h_i(\sigma_i^r) - h_i(\sigma_i^{r-1})). \tag{10}$$

Since h_i is linear, there exists a constant $C > 0$ such that h_i is $\frac{C}{n}$-Lipschitz continuous for $i \in I$. Due to the property of projection, $\|\Pi_{\hat{\Sigma}_i}(\sigma_i^t - \alpha(t) g_i^t) - \sigma_i^t\| \leq \alpha(t)\|g_i^t\| \leq \alpha(t) D$ for $\sigma_i^t \in \hat{\Sigma}_i$.

Claim 5.2. (Koshal et al. [11], Nedic et al. [17]) Denote the transition matrices $\Phi(k,s)$ from time s to $k > s$ as $\Phi(k,s) = W(k)W(k-1)\cdots W(s)$ for $0 \leq s < k$. Under Assumption 3.4(v),

(i) $\lim_{k\to\infty} \Phi(k,s) = \frac{1}{n} 1 1^T$ for all $s \geq 0$.

(ii) $|[\Phi(k,s)]_{ij} - 1/n| \leq \Gamma \beta^{k-s}$ for all $k \geq s \geq 0$ and $i, j \in I$, where $\Gamma = (1-\eta/(4n^2))^{1/\mathcal{B}}$ and $\beta = (1-\eta/(4n^2))^{1/\mathcal{B}}$.

Combining (9) and (10), with Claim 5.2, for $t \geq 1$,

$$\|u_i^t - \bar{v}^t\| \leq \Gamma \beta^t R_0 + C\Gamma D \sum_{r=1}^{t} \beta^{t-r} \alpha(r),$$

where $R_0 = \sum_{j=0}^{n} v_j^0$. Since $\alpha(t)$ is non-increasing,

$$\sum_{t=0}^{\infty} \alpha(t) \|v_i^t - \bar{v}^t\| \leq \Gamma \frac{1}{1-\beta} R_0 \alpha(0) + C\Gamma D \frac{1}{1-\beta} \sum_{t=0}^{\infty} \alpha^2(t). \tag{11}$$

Because the stepsize $\alpha(t)$ satisfies $\sum_{t=0}^{\infty} \alpha(t) = \infty$ and $\sum_{t=0}^{\infty} \alpha^2(t) < \infty$, $\sum_{t=0}^{\infty} \alpha(t)\|v_i^t - \bar{v}^t\| < \infty$.

Thirdly, we give the convergence result. With the update rule of σ_i^t,

$$\|\sigma_i^{t+1} - \hat{\sigma}_i^*\| = \|\Pi_{\hat{\Sigma}_i}(\sigma_i^t - \alpha(t) g_i^t) - \hat{\sigma}_i^*\| \leq \|\sigma_i^t - \hat{\sigma}_i^* - \alpha(t)(g_i^t + g_i^*)\|,$$

where $g_i^* = col(F_i(\hat{\sigma}_i^*, \hat{\bar{\sigma}}^*, \theta_i^1), \ldots, F_i(\hat{\sigma}_i^*, \hat{\bar{\sigma}}^*, \theta_i^N))$. Then

$$\|\sigma_i^{t+1} - \hat{\sigma}_i^*\|^2 \leq \|\sigma_i^t - \hat{\sigma}_i^*\|^2 + \alpha^2(t)\|g_i^t - g_i^*\|^2 - 2\alpha(t) \left\langle g_i^t - g_i^*, \sigma_i^t - \hat{\sigma}_i^* \right\rangle. \tag{12}$$

To prove that σ_i^t converges to $\hat{\sigma}_i^*$, we need to show that $\sum_{t=0}^{\infty} \alpha(t)\|\sigma_i^t - \sigma_i^*\| < \infty$. Since $\|g_i^t\| \leq \sqrt{N} D$ and $\|g_i^*\| \leq \sqrt{N} D$,

$$\sum_{t=0}^{\infty} \alpha^2(t) \|g_i^t - g_i^*\| < \infty.$$

Based on the property of subgradients, for $\theta_i^k \in \hat{\Theta}_i$,

$$N\left\langle g_i^t, \sigma_i^t - \hat{\sigma}_i^*\right\rangle = \sum_{k=1}^{N}\left\langle F_i(\sigma_i^t, u_i^t, \theta_i^k) - F_i(\hat{\sigma}_i^*, \hat{\bar{\sigma}}^*, \theta_i^k), \sigma_i^t(\theta_i^k) - \hat{\sigma}_i^*(\theta_i^k)\right\rangle.$$

Since F is Lipschitz continuous,

$$\|F_i(\sigma_i^t, u_i^t, \theta_i^k) - F_i(\sigma_i^t, \bar{v}^t, \theta_i^k)\| \leq L_u\|u_i^t(\theta_i^k) - \bar{v}^t(\theta_i^k)\|,$$

where $u_i^t(\theta_i^k)$ and $\bar{v}^t(\theta_i^k)$ denoted the aggregates $u_i^t(\tilde{\theta}|\theta_i^k)$ and $\bar{v}^t(\tilde{\theta}|\theta_i^k)$ with $\tilde{\theta} \in \{\sum_{i=1}^n \theta_i/n, \theta_i = \theta_i^k, \theta_{-i} \in \Theta_{-i}\}$, satisfying $\|u\|^2 = \sum_{k=1}^N \|u(\theta_i^k)\|^2$ for $i \in I$ and $u = u_i^t, \bar{v}^t$. Then

$$\begin{aligned} & N\left\langle g_i^t - g_i^*, \sigma_i^t - \hat{\sigma}_i^*\right\rangle \\ \leq & \sum_{k=1}^{N}\left\langle F_i(\sigma_i^t, \bar{v}^t, \theta_i^k) - F_i(\hat{\sigma}_i^*, \hat{\bar{\sigma}}^*, \theta_i^k), \sigma_i^t(\theta_i^k) - \hat{\sigma}_i^*(\theta_i^k)\right\rangle + \sqrt{N}L_u\|u_i^t - \bar{v}^t\|. \end{aligned}$$

Denote $EF_i(\sigma_i, \bar{\sigma}) = \sum_{k=1}^N F(\sigma_i, \bar{\sigma}, \theta_i^k)$ for $\sigma \in \hat{\boldsymbol{\Sigma}}$, where $\bar{\sigma}$ is the aggregate of $\boldsymbol{\sigma}$, and $E\boldsymbol{F}(\boldsymbol{\sigma}) = col(EF_1(\sigma_1, \bar{\sigma}), \ldots, EF_n(\sigma_n, \bar{\sigma}))$. From the strict convexity of f_i, EU_i is strict convex, and thus,

$$\left\langle E\boldsymbol{F}(\boldsymbol{\sigma}^t, H(\boldsymbol{\sigma}^t)) - E\boldsymbol{F}(\hat{\boldsymbol{\sigma}}^*, H(\hat{\boldsymbol{\sigma}}^*)), \boldsymbol{\sigma}^t - \hat{\boldsymbol{\sigma}}^*\right\rangle < 0.$$

Combining with (11), $\sum_{i=1}^n \|\sigma_i^t - \hat{\sigma}_i^*\|^2 = \|\boldsymbol{\sigma}^t - \hat{\boldsymbol{\sigma}}^*\|^2$ is convergent. Furthermore,

$$\sum_{t=0}^{\infty}\alpha(t)\left\langle E\boldsymbol{F}(\boldsymbol{\sigma}^t, H(\boldsymbol{\sigma}^t)) - E\boldsymbol{F}(\hat{\boldsymbol{\sigma}}^*, H(\hat{\boldsymbol{\sigma}}^*)), \boldsymbol{\sigma}^t - \hat{\boldsymbol{\sigma}}^*\right\rangle < \infty.$$

Since $\sum_{t=0}^\infty \alpha(t) = \infty$, there exists a subsequence $\{t_l\}$ such that

$$\lim_{l\to\infty}\left\langle E\boldsymbol{F}(\boldsymbol{\sigma}^{t_l}, H(\boldsymbol{\sigma}^{t_l})) - E\boldsymbol{F}(\hat{\boldsymbol{\sigma}}^*, H(\hat{\boldsymbol{\sigma}}^*)), \boldsymbol{\sigma}^{t_l} - \hat{\boldsymbol{\sigma}}^*\right\rangle = 0.$$

Due to the strict convexity of f_i, $E\boldsymbol{F}$ is strictly monotone, and thus, $\lim_{l\to\infty}\|\boldsymbol{\sigma}^{t_l} - \hat{\boldsymbol{\sigma}}^*\| = 0$. Since $\|\boldsymbol{\sigma}^t - \hat{\boldsymbol{\sigma}}^*\|$ is convergent,

$$\lim_{t\to\infty}\|\boldsymbol{\sigma}^t - \hat{\boldsymbol{\sigma}}^*\| = 0.$$

Therefore, we complete the proof. □

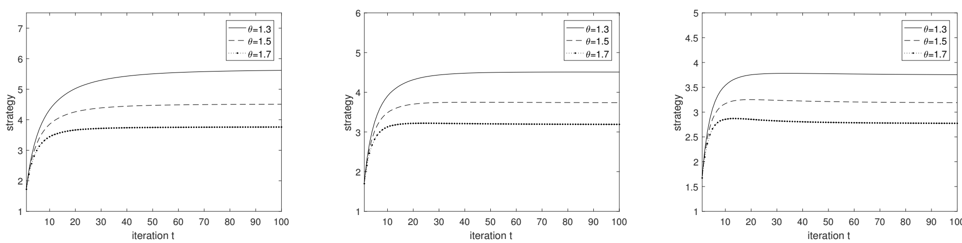
Remark 5.3. Compared with existing works which were restricted in deterministic aggregative games [11, 14, 21], Algorithm 1 handles the equilibrium seeking problem on aggregative Bayesian games, and Theorem 5.1 provides its convergence to an ϵ-BNE of G with an explicit bound ϵ.

6. NUMERICAL SIMULATIONS

In this section, we provide numerical simulations to illustrate the effectiveness of Algorithm 1 on aggregative Bayesian games.

Consider a Nash-Cournot game played by 5 competitive firms to produce a kind of commodity. The costs of productivity, for example, the price of raw materials, are uncertain in the game, which is referred to the type $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_5) \in [1,2]^5$ and $\theta_1, \ldots, \theta_5$ are independent and uniformly distributed over $[1,2]$, respectively. For firm i, $i \in \{1, \ldots, 5\}$, it has a feasible action set $X_i = [0, 10]$ and cost function $f_i(x_i, \tilde{x}, \theta_i) = -(40 - 5\bar{x})x_i + (3\theta_i + 0.5(i+1))x_i^2$, where $\tilde{x}$ is the aggregate of actions. Each firm i can only obtain the probability distribution of its own type θ_i and the average type $\tilde{\theta}$. Firm i adopts a strategy σ_i, which means that when it receives a type $\theta_i \in [1,2]$, it takes $\sigma_i(\theta_i) \in X_i = [0, 10]$ as the quantity of the commodity to produce. The expectation of the cost and the aggregation function were defined in Section 3. Firms exchange their aggregates via a connected time-varying graph $\mathcal{G}(t)$, which is randomly generated.

Firstly, we show the convergence of Algorithm 1. Fig. 1 presents the trajectories of strategies for different agents at some specific types under $N = 100$. We can see that the agents' strategies at these specific types converge, which indicates the convergence of Algorithm 1 in Theorem 5.1.



(a) agent 1 (b) agent 2 (c) agent 3

Fig. 1. Strategies of agents at specific types with $N = 100$.

Next, we verify the effectiveness of the discretization. Since Algorithm 1 generates an ϵ-BNE of the continuous-type model G, we calculate the value of ϵ for different numbers of discrete points N. Denote ϵ_i as the difference of the expectations EU_i at the BNE $(\sigma_i^*, \sigma_{-i}^*)$ and at the DBNE $(\hat{\sigma}_i^*, \hat{\sigma}_{-i}^*)$, i. e., $\epsilon_i = EU_i(\hat{\sigma}_i^*, \hat{\sigma}_{-i}^*) - EU_i(\sigma_i^*, \sigma_i^*)$. Tab. 1 shows ϵ_i for different agents and N's. Since $\epsilon = \max_{i \in I} \epsilon_i$, we find that ϵ tends to 0 as N tends to infinity and is consistent with Theorem 4.4 that $\epsilon = O(\max_{i \in I} \max_{k \in [N]}(\theta_i^k - \theta_i^{k-1}))$.

7. CONCLUSION

In this paper, we considered an aggregative Bayesian game, which is a generalization of both Bayesian games and deterministic aggregative games. We handled the aggregation function with the incomplete-information situations. To break through barriers in seeking BNE, we provided a discretization method, and proved that the DBNE of the gen-

N	ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5
50	3.60×10^{-2}	1.20×10^{-2}	1.16×10^{-3}	4.33×10^{-3}	7.28×10^{-3}
100	1.82×10^{-2}	6.23×10^{-3}	8.02×10^{-4}	1.96×10^{-3}	3.45×10^{-3}
200	1.40×10^{-2}	7.09×10^{-3}	3.75×10^{-3}	1.91×10^{-3}	8.24×10^{-4}
500	4.80×10^{-3}	2.20×10^{-3}	9.64×10^{-4}	3.05×10^{-4}	7.44×10^{-5}

Tab. 1. The value ϵ_i for different agents and N's.

erated discretized model is an ϵ-BNE of the original model with an explicit error bound. On this basis, we proposed a distributed algorithm for the DBNE of the discretized model, namely an ϵ-BNE of the continuous-type model, and proved its convergence. Because the dimension of the aggregation term can be very huge after discretization, which may cause problems in communications as discussed in Remark 4.5, we are trying to explore more efficient distributed designs that can reduce the resulting communication burdens.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (No. 62173250). Correspondent to Guanpu Chen.

(Received November 8, 2022)

REFERENCES

[1] K. Akkarajitsakul, E. Hossain, and D. Niyato: Distributed resource allocation in wireless networks under uncertainty and application of Bayesian game. IEEE Commun. Magazine *49* (2011), 8, 120–127. DOI:10.1109/MCOM.2011.5978425

[2] U. Bhaskar, Y. Cheng, Y. K. Ko, and C. Swamy: Hardness results for signaling in Bayesian zero-sum and network routing games. In: Proc. 2016 ACM Conference on Economics and Computation, pp. 479–496.

[3] G. Chen, K. Cao, and Y. Hong: Learning implicit information in Bayesian games with knowledge transfer. Control Theory Technol. *18* (2020), 3, 315–323. DOI:10.1007/s11768-020-0086-2

[4] G. Chen, Y. Ming, Y. Hong, and P. Yi: Distributed algorithm for ϵ-generalized Nash equilibria with uncertain coupled constraints. Automatica *123*, (2021), 109313. DOI:10.1016/j.automatica.2020.109313

[5] M. Großhans, C. Sawade, M. Bruckner, and T. Scheffer: Bayesian games for adversarial regression problems. In: International Conference on Machine Learning 2013, pp. 55–63.

[6] S. Guo, H. Xu, and L. Zhang: Existence and approximation of continuous Bayesian Nash equilibria in games with continuous type and action spaces. SIAM J. Optim. *31* (2021), 4, 2481–2507. DOI:10.1137/19M1298032

[7] W. Guo, M. I. Jordan, and T. Lin: A variational inequality approach to Bayesian regression games In: 2021 60th IEEE Conference on Decision and Control (CDC), pp. 795–802.

[8] J. C. Harsanyi: Games with incomplete information played by "Bayesian" players, I–III; Part I. The basic model. Management Sci. *14* (1967), 3, 159–182. DOI:10.1287/mnsc.14.3.159

[9] L. Huang and Q. Zhu: Convergence of Bayesian Nash equilibrium in infinite Bayesian games under discretization. arXiv:2102.12059 (2021).

[10] S. Huang, J. Lei, and Y. Hong: A linearly convergent distributed Nash equilibrium seeking algorithm for aggregative games. IEEE Trans. Automat. Control *68* (2023), 3, 1753–1759. DOI:10.1109/TAC.2022.3154356

[11] J. Koshal, A. Nedi, and U. V. Shanbhag: Distributed algorithms for aggregative games on graphs. Oper. Res. *64* (2016), 3, 680–704. DOI:10.1287/opre.2016.1501

[12] V. Krishna: Auction Theory. Academic Press, London 2009.

[13] V. Krishnamurthy and H. V. Poor: Social learning and Bayesian games in multiagent signal processing: How do local and global decision makers interact? IEEE Signal Process. Magazine, *30* (2013) 3, 43–57. DOI:10.1109/MSP.2012.2232356

[14] S. Liang, P. Yi, and Y. Hong: Distributed Nash equilibrium seeking for aggregative games with coupled constraints. Automatica *85* (2017), 179–185. DOI:10.1016/j.automatica.2017.07.064

[15] A. Meirowitz: On the existence of equilibria to Bayesian games with non-finite type and action spaces. Econom. Lett. *78* (2003), 2, 213–218. DOI:10.1016/S0165-1765(02)00216-1

[16] P. R. Milgrom and R. Journal Weber: Distributional strategies for games with incomplete information. Math. Oper. Res. *10* (1985), 4, 619–632. DOI:10.1287/moor.10.4.619

[17] A. Nedic, A. Ozdaglar, and P. A. Parrilo: Constrained consensus and optimization in multi-agent networks. IEEE Trans. Automat. Control *55* (2010), 4, 922–938. DOI:10.1109/TAC.2010.2041686

[18] Z. Rabinovich, V. Naroditskiy, E. H. Gerding, and R. J. Nicholas: Computing pure Bayesian-Nash equilibria in games with finite actions and continuous types. Artif. Intell. *195* (2013), 106–139.

[19] B. Rynne and M. A. Youngson: Linear Functional Analysis. Springer Science and Business Media, 2007.

[20] T. Ui: Bayesian Nash equilibrium and variational inequalities. J. Math. Econom. *63* (2016), 139–146. DOI:10.1016/j.jmateco.2016.02.004

[21] G. Xu, G. Chen, H. Qi, and Y. Hong: Efficient algorithm for approximating Nash equilibrium of distributed aggregative games. IEEE Trans. Cybernet. *53* (2023), 7, 4375–4387. DOI:10.1109/TCYB.2022.3175831

[22] H. Zhang, G. Chen, and Y. Hong: Distributed algorithm for continuous-type Bayesian Nash equilibrium in subnetwork zero-sum games. arXiv:2209.06391 (2022).

Hanzheng Zhang, Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai. P. R. China.
e-mail: zhanghanzheng@amss.ac.cn

Huashu Qin, Academy of Mathematics and Systems Science, CAS, Beijing. P. R. China.
e-mail: qin@iss.ac.cn

Guanpu Chen, Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm. Sweden.
e-mail: guanpu@kth.se