

# Learning to Acquire Resources in Competition

Safwan Hossain\*  
Harvard University

Mirah Shi\*  
University of Pennsylvania

Andrew Bennett  
Morgan Stanley

Neil Chriss  
EdgePeak Capital

Michael Kearns  
University of Pennsylvania

Anderson Schneider  
Morgan Stanley

Yuriy Nevmyvaka  
Morgan Stanley

## Abstract

We consider multiple agents competing to acquire stakes in some costly divisible resource (*e.g.* shares of a financial asset, compute resources, or commodities) over time. We propose a novel game-theoretic model for this problem that generalizes settings studied in diverse literatures, and analyze it under different assumptions on agent information. Given complete-information, we establish the existence and uniqueness of a pure Nash equilibrium (NE) in this generalized setting. This is shown to be efficiently computable but has worst-case unbounded price of anarchy. Alternatively, under partial-information with a common prior, we establish the existence and uniqueness of a Bayesian Nash equilibrium (BNE), which is also efficiently computable. Finally, we propose a more realistic learning setting for the game, where agents have partial information but no common prior. Instead, they must learn how to act given online contextual feedback from interactions in stochastically sampled game instances. We provide sufficient conditions on agents doing simultaneous no-regret learning for convergence to Bayesian coarse-correlated equilibrium (BCCE) or last-iterate convergence to the BNE. In each setting, we provide detailed simulations, which empirically validates our theory and provides new insights into strategic behavior of resource acquisition.

---

\*Safwan Hossain and Mirah Shi contributed equally and completed this as part of an internship in the Morgan Stanley Machine Learning Research Team during Summer 2025. Correspondence to shossain@g.harvard.edu or mirahshi@seas.upenn.edu

# 1 Introduction

Consider multiple traders attempting to acquire a position in a stock ahead of an earnings release, under the belief that its price will rise afterward. If each trader were acting in isolation, they might follow a classical optimal execution strategy – such as that of Almgren and Chriss [2] – to minimize their trading costs. However, if multiple traders are pursuing their strategies simultaneously, their aggregate activity influences prices and liquidity. This interaction transforms the problem from one of individual optimization into one of understanding the intra-agent strategic behavior, where each agent’s decisions affect the market environment faced by others.

This challenge of acquiring costly resources in competitive, dynamically priced environments extends well beyond financial markets. For instance, a firm training a large machine learning model may need to secure substantial cloud computing resources within a given time frame. Here too, spot prices are shaped by aggregate demand across many users, requiring firms to account not only for their own scheduling and budget constraints but also for how their actions interact with others [26]. Further, agents in many such environments may only have partial or incomplete knowledge of other market participants or the market itself.

While recent works have attempted to capture these strategic perspectives, they do so with several limitations. Chriss [8, 9, 7] all consider a complete-information setting, which is unrealistic in all but very limited scenarios. Chriss [10], Kearns and Shi [15] more recently consider some extensions to deal with this, but these also have major limitations, requiring shared common priors over uncertain information or repeated play of fixed game instances respectively. Furthermore, all of these works: (1) require agents to acquire a fixed target position, thereby ruling out more general action constraints; (2) do not allow for custom objectives that agents may wish to optimize alongside acquisition costs; and most importantly (3) do not address the computational and learning challenges that arise when agents act under incomplete information without common priors. In addition, these works are all finance-specific. The broad scope and practical relevance of this problem thus necessitates a general game-theoretic framework that can gracefully accommodate diverse applications and practical limitations of real-world markets, which is the foundational premise of our work.

## 1.1 Our Contribution

- In Section 2 we propose a novel model for this problem, which generalizes and improves on past settings, by allowing for convex constraints, concave idiosyncratic utility functions, and unknown or incomplete information available to the agents.
- In Section 3 we characterize the complete-information equilibria properties of this game. Even within our very general model, the game still has a unique, pure NE that is efficiently computable. We also show that in the worst-case, the price of anarchy of this game is unbounded.
- In Section 4 we consider the partial-information Bayesian setting: each agent only observes their own private information (their “type”), but all agents have common knowledge of the prior distribution over agent types and game parameters. We extend the complete-information results here, establishing the uniqueness and efficient computability of the Bayesian NE (BNE).
- In Section 5 we further extend our model to a more realistic learning-based setting, where agents only observe their own type, but do not have any knowledge of the prior distribution over types and game parameters. Instead, they learn from repeated interaction, where they iteratively decide their strategy conditioned on their realized type. This naturally models learning to

acquire resources in competition, given contextual information. We establish sufficient conditions under which agents engaging in simultaneous no-regret learning either convergence to a Bayesian coarse-correlated equilibrium (BCCE) on average over rounds, or to the BNE in the final round.

- For each setting, we provide simulations showcasing the respective algorithms.

## 1.2 Related Work

The most relevant related work to our setting is the recent line of work on optimal position building under competition in Chriss [8, 9, 7, 10], Kearns and Shi [15] that we discussed above. Our work can be seen as a generalization of these, for both financial and non-financial applications. We provide a detailed discussion of how our setting relates to and subsumes the settings in these works in Appendix A.2. To the best of our knowledge, there is no existing work that have considered strategic aspects of resource application in on-financial settings, although researchers have noted the relevance of such considerations, *e.g.* in compute markets Shastri and Irwin [26].

More broadly, our model captures standard notions of *market impact* in finance, of which there is a large literature (see *e.g.* Webster [29], Li et al. [18] and citations therein for a recent detailed overview). These works broadly consider how prices change in response to trading (both theoretically and empirically from real markets). In this literature, market impact is often decomposed into permanent and temporary impact [2, 4, 20], which is the same approach that we take. There are also some more flexible models such as the *propagator model* [6, 12, 22] that allow for transient impacts in between these extremes; we do not consider these, but such extensions would be an interesting direction for future work.

Our work also relates to the literature on learning in games more broadly. Of particular note, our setting in Section 5 is very similar in spirit to the setting of Hartline et al. [13], who provide a general framework for no-regret learning in repeated Bayesian games. Although our model is slightly outside their framework (as it allows continuous market types), and has some specific structure that we can leverage (strongly monotone) our Theorem 4 is very motivated by their theory.

Our strategic setup is conceptually related to resource allocation — or more generally, congestion — games, where agents share resources, and the cost of any resource depends on its demand [24]. In contrast to this setting, our work studies behavior under time-varying prices, which endogenously adjusts based on supply and demand. Lastly, while we study competitive resource acquisition in a market setting, a non-market variant has long been studied in the context of *fair division* [21] for both divisible and indivisible goods. Recent literature here has extended this problem to an online setting (see Aleksandrov and Walsh [1] for a survey), and often incorporates learning and predictions [5], spiritually motivating our model in Section 5.

## 2 Model

**Preliminaries:** We consider a market consisting of  $n$  strategic agents looking to trade (buy/sell) some costly, divisible resource (stock, bond, compute time, *etc.*) over a period of  $T$  rounds. In the simplest setting, each strategic agent  $i$ 's action is a  $T$ -dimensional vector  $\mathbf{h}_i$ , where  $h_{i,t}$  denotes how much they purchased at time  $t \in [T]$ . Note that  $\mathbf{h}_i$  is a signed vector, and we conventionally denote positive values as buying and negative values as selling throughout. We assume that each agent has some set of convex constraints on their allowable actions, which we represent by a feasible set of trajectories  $G_i \subseteq \mathbb{R}^T$ . For example, if agent  $i$  wants to procure at most  $V_i$  equity

shares without short selling or over-buying, they could represent their constraints via  $G_i = \{\mathbf{h}_i : 0 \leq \sum_{l=1}^t h_{i,l} \leq V_i \forall t \in [T]\}$ . We also assume that each agent has some idiosyncratic utility function on their strategy, which can capture (1) the utility of their final position and/or any preferences on their acquisition schedule. For an agent  $i$ , we represent this via a concave function  $f_i : \mathbb{R}^T \rightarrow \mathbb{R}$  (with the same units as price)<sup>1</sup>. For example, if agent  $i$  wishes to impose a concave utility function  $\phi_i$  on their final position and penalize selling, their idiosyncratic utility could be  $f_i(\mathbf{h}_i) = \phi_i(\mathbf{1}^\top \mathbf{h}_i) - \zeta \sum_{t=1}^T |h_{i,t}| \mathbb{I}\{h_{i,t} < 0\}$  with  $\zeta \geq 0$ , which is clearly concave. If an agent has a private valuation  $r_i$  for the asset, they could use  $\phi_i(x) = r_i x$ . In settings like compute markets or optimal trade execution, where the agents' goal is to acquire a fixed target position as cheaply as possible, one can set  $f_i = 0$  and include a hard constraint on  $\mathbf{1}^\top \mathbf{h}_i$  in  $G_i$ . Lastly, and inspired by the seminal work of Kyle [17], we allow the market to contain a non-strategic (possibly random) *exogenous* agent, which captures all non-strategic trade flow. Following convention, the exogenous agent's action is given by a signed vector  $\mathbf{s} \in \mathbb{R}^T$  with positive values indicating buying.

**Price Model:** Core to understanding how agents strategically interact in acquiring costly resources is how their demand/supply levels influence resource prices. We assume the following dynamic model for determining resource prices from agents' trading schedules:

**Assumption 1** (Price Dynamics). *All agents pay the same price  $p_t$  for each share of the resource at time  $t$ , where  $p_t$  is determined from the total trading schedule of all agents up to and including time  $t$  according to the following equations:*

$$p_t = p_t^w + \beta \left( \sum_{i=1}^n h_{i,t} + s_t \right); \quad p_t^w = p_{t-1}^w + \alpha \left( \sum_{i=1}^n h_{i,t} + s_t \right), \quad (1)$$

where  $p_0 = p_0^w$  is the initial price, and  $\alpha, \beta \geq 0$  are some problem parameters.

The dynamic process for  $p_t^w$  can be seen as a discretization of the Walrasian price dynamics from general equilibrium theory, which posits that prices evolve from an imbalance of supply and demand:  $dp_t = \alpha(\text{demand}_t - \text{supply}_t)dt$ , where  $\alpha$  is a sensitivity factor [28]. The additional  $\beta(\sum_{i=1}^n h_{i,t} + s_t)$  term in  $p_t$  accounts for additional costs imposed by market makers who provide liquidity to balance supply and demand, causing temporary deviations from the Walrasian price process ( $\beta$  controls the strengths of this impact). This also maps to how prices are modeled within the theory of optimal trade execution, with  $\alpha$  and  $\beta$  corresponding to permanent and temporary impact coefficients, respectively [2]. We discuss in detail in Appendix A.1.

**Game Payoff Structure:** We model the total utility for each agent according to their personal utility  $f_i$ , minus the total cost they incur buying and selling. Formally:

**Definition 1** (Game Payoffs). *Let the price parameters  $p_0$ ,  $\alpha$ , and  $\beta$ , and exogenous action  $\mathbf{s}$ , be given. In addition, let  $\mathbf{h}_{-i}$  denote the trading schedules of all strategic agents other than  $i$ , and  $\mathbf{p}(\mathbf{h}_i, \mathbf{h}_{-i}, p_0, \boldsymbol{\lambda}) \in \mathbb{R}^T$  denote the sequence of prices under Assumption 1 for  $\mathbf{h}_i$ ,  $\mathbf{h}_{-i}$ , and  $\boldsymbol{\lambda} = (f_1, \dots, f_n, p_0, \alpha, \beta, \mathbf{s})$ . Then, the overall utility for agent  $i$  is:*

$$u_i(\mathbf{h}_i; \mathbf{h}_{-i}, \boldsymbol{\lambda}) = f_i(\mathbf{h}_i) - \mathbf{p}(\mathbf{h}_i, \mathbf{h}_{-i}, \boldsymbol{\lambda})^\top \mathbf{h}_i. \quad (2)$$

<sup>1</sup>This allows general concave utility on final position, which corresponds to diminishing marginal utility and is a natural restriction in economics and game theory. See [19, 11].

We note that these payoff (utility) functions, along with the constraint sets  $G_i$  for each agent  $i$ , fully define the strategic game for fixed game parameters. In addition, we note that these payoff functions can easily be shown to be equivalent to the linear/quadratic cost function considered in Chriss [8, 9], in the case where  $f_i = 0$  for all  $i$ ; we provide details of this in Appendix A.2.

**Bayesian Game Extension:** In addition to considering fixed instances of the strategic resource acquisition game, as defined above (which we analyze in detail in Section 3,) we also consider a Bayesian game extension, where there is uncertainty in the game parameters, and each agent only observes some private information that may be correlated with these. This is needed for the partial-information settings we consider in Section 4 and Section 5.

**Definition 2** (Bayesian Game). *The Bayesian extension of our game is formalized by the following:*

1. **Market Type:** We define the market type as  $\lambda = (f_1, \dots, f_n, p_0, \alpha, \beta, \mathbf{s})$ , as in Definition 1.
2. **Agent Type:** We let  $\theta_i$  denote the type of agent  $i$ , which is the set of all information known to them prior to acting.  $\Theta_i$  denotes the possible type space for agent  $i$ , where  $|\Theta_i| = k_i < \infty$ .
3. **Agent Constraints:** Agent types  $\theta_i$  fully determines their constraints, which we denote  $G_i(\theta_i)$ .
4. **Distribution over game instances:** Letting  $\mathcal{I} = (\theta_1, \dots, \theta_n, \lambda)$  denote a full game instance, there exists a joint probability distribution  $P(\theta_1, \dots, \theta_n, \lambda)$  over the components of  $\mathcal{I}$ .
5. **Agent Strategy:** In this extended model, the strategy for agent  $i$  is a function  $\mathbf{h}_i : \Theta_i \rightarrow \mathbb{R}^T$ , which determines how they would behave under each possible type.
6. **Feasible Strategy Sets:** We let  $\mathcal{H}_i$  denote the set of feasible strategies for each agent  $i$ , which we formally define as  $\mathcal{H}_i = \{\mathbf{h} : \mathbf{h}(\theta) \in G_i(\theta) \ \forall \theta \in \Theta_i\}$ .
7. **Agent Utility:** Each agent  $i$  defines their utility given strategy  $\mathbf{h}_i$  and opponent strategies  $\mathbf{h}_{-i}$  as the expected utility over  $\mathcal{I} \sim P$ , which is given by  $\mathbb{E}_{\theta_i, \theta_{-i}, \lambda \sim P}[u_i(\mathbf{h}_i(\theta_i); \mathbf{h}_{-i}(\theta_{-i}), \lambda)]$ .

It is trivial to verify that the goal of all agents maximizing their overall expected utility is equivalent to each agent maximizing their conditional expected utility given their private information, given by  $\mathbb{E}_{\theta_{-i}, \lambda \sim P|\theta_i}[u_i(\mathbf{h}_i(\theta_i); \mathbf{h}_{-i}(\theta_{-i}), \lambda)]$ . In this model, the only requirement of the agent types  $\theta_i$  is that they specify the agent's constraints  $G_i(\theta_i)$ . Other than this, the information specified by the type may be generic; it may either completely determine, be partially correlated with, or completely uninformative of any given component in  $\lambda$ . In particular, we do *not* generally assume that  $\theta_i$  specifies  $f_i$ , since we allow for idiosyncratic utilities to depend on uncertain market valuations. In addition, the structure of agent types may be generic, but in practice we may often think of them as defined by some structured set of features that the respective agent uses to decide their trading schedule. Finally, we note that the types for different agents may have very different strengths of correlation with other agent types and components of  $\lambda$ , which allows for information asymmetry.

### 3 Complete Information Setting

In the complete information setting, the market type  $\lambda$  and agent constraints  $G_1, \dots, G_n$  are observed by all  $n$  strategic agents. Therefore, it is unnecessary to consider agent types or the distribution  $P$ , so we instead just analyze an arbitrary fixed game instance  $\mathcal{I}$  defined by  $G_1, \dots, G_n, \lambda$ . Recall from Section 2 that for fixed game instances, we let  $\mathbf{h}_i$  denote a fixed trading schedule (*i.e.*  $\mathbf{h}_i \in \mathbb{R}^T$ ) rather than a function of agent types.

Complete information games are routinely studied in game theoretic models, since: (1) they provide clearer intuitions on the strategic dynamics of the problem; and (2) they are the basis for common solution concepts, namely Nash Equilibria (NE) and Price of Anarchy (PoA) [25]:

**Definition 3.** For a complete information instance  $\mathcal{I}$ , the strategies  $(\mathbf{h}_1^{eq}, \dots, \mathbf{h}_n^{eq})$  are a Pure Nash Equilibrium if and only if: for all buyers  $i$  and any strategy  $\mathbf{h}'_i$ :  $u_i(\mathbf{h}_i^{eq}; \mathbf{h}_{-i}^{eq}, \boldsymbol{\lambda}) \geq u_i(\mathbf{h}'_i; \mathbf{h}_{-i}^{eq}, \boldsymbol{\lambda})$ .

**Definition 4.** For a complete information instance  $\mathcal{I}$ , let  $NE(\mathcal{I})$  denote the set of all NE strategies, and let  $welf(\mathbf{h}_1, \dots, \mathbf{h}_n, \boldsymbol{\lambda}) = \sum_{i=1}^n u_i(\mathbf{h}_i; \mathbf{h}_{-i}, \boldsymbol{\lambda})$  denote the welfare function. The Price of Anarchy ratio is then defined as:  $\sup_{\mathbf{h}_1 \in G_1, \dots, \mathbf{h}_n \in G_n, \mathbf{h}^{eq} \in NE(\mathcal{I})} \frac{welf(\mathbf{h}_1, \dots, \mathbf{h}_n, \boldsymbol{\lambda})}{welf(\mathbf{h}_1^{eq}, \dots, \mathbf{h}_n^{eq}, \boldsymbol{\lambda})}$ .

Informally, the NE are the set of strategies such that no agent has any incentive to unilaterally deviate, and the PoA characterizes the ratio of the best obtainable welfare if agents were to cooperate to the worst obtainable welfare obtainable from NE. We begin with an explicit expression of agent utility, which we show to be strictly concave, allowing us to characterize the above notions. We provide details and derivation of the lemma in Appendix B.

**Lemma 1.** By unrolling the auto-regressive price definition in Equation (1), the utility of agent  $i$  in instance  $\mathcal{I}$  for joint strategy  $(\mathbf{h}_1, \dots, \mathbf{h}_n)$  is strictly concave in their strategy, and is given by:

$$u_i(\mathbf{h}_i; \mathbf{h}_{-i}, \boldsymbol{\lambda}) = f_i(\mathbf{h}_i) - \frac{1}{2} \mathbf{h}_i^T Q \mathbf{h}_i - \sum_{j \neq i} (A \mathbf{h}_j)^T \mathbf{h}_i - \mathbf{s} A \mathbf{h}_i - p_0(\mathbf{1}^T \mathbf{h}_i),$$

where  $Q$  and  $A$  are  $n \times n$  matrices defined in terms of  $\alpha$  and  $\beta$ , and  $Q$  is symmetric and strictly positive definite (PD).

Our definitions and statements so far have been framed with respect to pure (deterministic) strategies. In general, strategic agents may use mixed (randomized) strategies, which begs the following question: could mixed strategies appear in NE? In the following lemma (proof in Appendix B), we answer this question in the negative by characterizing an agent's best response – their optimal strategy for a fixed set of others' strategies – as being pure.

**Lemma 2.** For any fixed game instance  $\mathcal{I}$ , the best response of any agent  $i$  is always unique and deterministic, even when others are playing some mixed (possibly correlated) strategies.

Next, we turn to the first of our two central results in this section, regarding characterization and computation of the NE, which we address via the following theorem. The proof is technical and stems from casting the equilibrium conditions as a variational inequality, and then proving that the operator for this variational inequality is strongly monotone. This property immediately implies uniqueness of the NE, and gives us an efficient gradient-based algorithm for computing it. We formalize these results below, with the proof and full algorithm details given in Appendix B.

**Theorem 1.** For every fixed game instance  $\mathcal{I}$ , there is a unique pure NE. In addition, the extra-gradient algorithm [16] converges linearly to this equilibrium.

Finally, we turn to the question of characterizing the PoA, which we show in general is unbounded. The proof is based on an explicit counterexample, whose intuition is as follows: consider two agents with differing valuation on their final position, where one agent wants to buy and other wants to sell. If they coordinate, they can provide liquidity to each other, which eliminates trading frictions and allows them to trade a large quantity and obtain high welfare. However, if

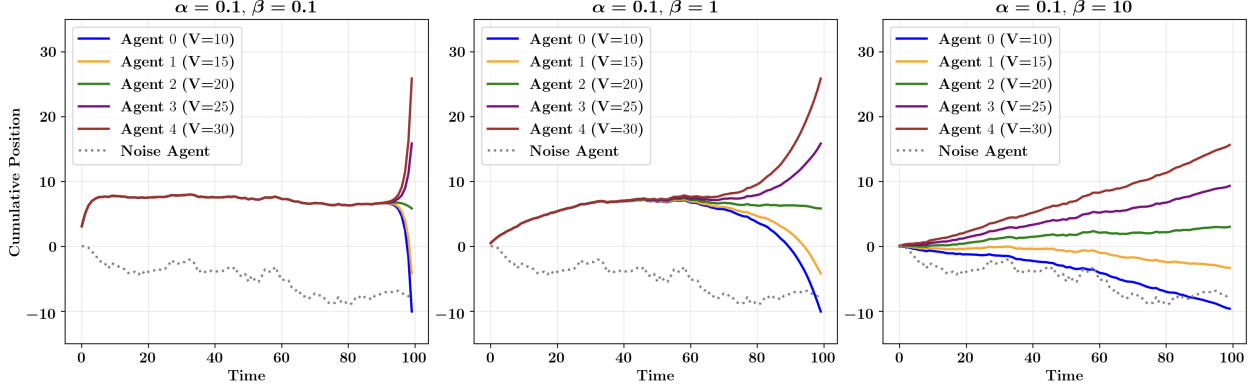


Figure 1: Cumulative position over time for agents in NE, for fixed  $\alpha$  and varying  $\beta$ . The initial price is  $p_0 = 2$ , the reserve prices are (4, 5, 6, 7, 8) and the constraint values are  $V = (10, 15, 20, 25, 30)$ .

they agreed to do this, each would have an incentive to cheat by providing less liquidity to the other; by doing so, the trade imbalance would move the price favorably for them, which they could profit from. Because of this, the NE involves both agents trading almost nothing, and achieving very low welfare. We provide full details in Appendix B.

**Theorem 2.** *For any constants  $\alpha, \beta, T$ , and any  $\xi > 0$ , there exists an instance  $\mathcal{I}$  of the complete information game with PoA ratio at least  $\xi$ . Therefore, the PoA is unbounded.*

### 3.1 Empirical Simulations of Equilibrium

To further understand the agents' behavior in equilibrium, we empirically compute the NE for a simple, yet practically motivated setting. Consider 5 strategic agents with a linear final position utility  $f_i(\mathbf{h}_i) = r_i \sum_t h_{i,t}$  for some reserve price  $r_i$ , and constraints  $-V_i \leq \mathbf{1}^T \mathbf{h}_i \leq V_i$  (i.e. final position must be in range  $[-V, V]$ ). We randomly sample the actions  $s_t$  of the exogenous agent using i.i.d. zero-mean random variables. In Figure 1 we plot the cumulative positions ( $\sum_{l=1}^t h_{i,l}$ ) of each agent  $i$ , for three different sets of problem parameters, where we fix  $\alpha = 0.1$  and vary  $\beta \in \{0.1, 1, 10\}$ . We observe that, as  $\beta$  increases, the total volume traded decreases, which is unsurprising since  $\beta$  corresponds to trading frictions. More interestingly, we note a phase transition. For small  $\beta$ , the NE approaches a pair of block trades – a first at time 0 where all agents purchase an identical quantity of the resource, and a second at time  $T$  where all agents buy or sell to reach some final position. For large  $\beta$ , the NE approaches all agents trading at a constant rate, with some interpolation between these for intermediate  $\beta$ .

## 4 Partial Information Setting with Common Knowledge of Prior

We now consider the partial-information setting: all agents have common knowledge of the joint distribution  $P$  of agent and market types, but each agent only observes their own private information via their type  $\theta_i$ . For this setting, we only consider characterization and computation of equilibria, since the unbounded PoA for complete information settings automatically carries over here. As discussed in Section 2, we think of the strategy of each agent as an *ex-ante* mapping from each possible type  $\theta_i$  to a feasible trading schedule in  $G_i(\theta_i)$ , which we denote by the function  $\mathbf{h}_i \in \mathcal{H}_i$ . Game play in this Bayesian setting operates via the following sequence of events: (1) each agent

decides their *ex-ante* strategy  $\mathbf{h}_i$ ; (2) a game instance  $\mathcal{I} \sim P$  is sampled and the corresponding type information  $\theta_i$  is privately revealed to each agent; and (3) each agent executes their strategy  $\mathbf{h}_i(\theta_i)$ . This setting can be formally studied within the Bayesian game theory framework, using the standard equilibrium notion as follows:

**Definition 5** (Bayesian Nash Equilibrium). *For a Bayesian instance, the strategies  $(\mathbf{h}_1^{eq}, \dots, \mathbf{h}_n^{eq})$  are in a Bayesian Nash Equilibrium (BNE) if for all agents  $i$ , all  $\mathbf{h}'_i \in \mathcal{H}_i$ , and all  $\theta_i \in \Theta_i$ , we have:  $\mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P|\theta_i}[u_i(\mathbf{h}_i^{eq}(\theta_i); \mathbf{h}_{-i}^{eq}(\theta_{-i}), \boldsymbol{\lambda})] \geq \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P|\theta_i}[u_i(\mathbf{h}'_i(\theta_i); \mathbf{h}_{-i}^{eq}(\theta_{-i}), \boldsymbol{\lambda})]$ .*

As in Section 3, this equilibrium is defined in terms of pure strategies (deterministic trading schedule for each type). The following lemma, which generalizes Lemma 2 to the Bayesian game setting, ensures that this restriction does not restrict the BNE (proof details in Appendix C).

**Lemma 3.** *For any Bayesian instance, the best response of any agent  $i$  is always unique and deterministic (meaning trading schedule  $\mathbf{h}_i(\theta_i)$  for every type  $\theta_i$  is deterministic), even if others are playing some mixed (possibly correlated) set of strategies for each type.*

We now present the central result in this setting. Theorem 3 generalizes the result of Theorem 1 to the Bayesian game setting, namely that there is a unique and efficiently computable equilibrium. Similar to the complete information setting, this follows by casting the equilibrium problem as a variational inequality, which we show is strongly monotone; this implies uniqueness of the BNE, and gives an efficient gradient based algorithm for computing it. We provide full details in Appendix C.

**Theorem 3.** *For every Bayesian game instance (given by distribution  $P$ ), there is a unique pure BNE, and the extra-gradient algorithm [16] converges linearly to this BNE.*

#### 4.1 Empirical Simulations of Equilibrium

We simulate the BNE for a similar scenario as in Section 3.1, where agents' have linear utility and inequality constraints on their final positions. We use 2 agents here, each with 3 possible types, and the type  $\theta_i$  determines the final position bounds  $V_i$  and expected reserve price  $r_i$  for each agent. We set  $\mathbf{s} = 0$  in this simulation. As before, we fix  $\alpha = 0.1$ , and vary  $\beta$ ; in each case,  $\beta$  is continuously distributed within some bounded range. We provide full details in Appendix C.

We see similar phase transition dynamics as we vary  $\beta$  from small to large as we observed in the complete information setting. However, in this case, since the cumulative positions in each agent's strategy are type-dependent, the strategies do not consistently overlap during early time steps. That is, the partial information induces a richer, type-dependent dynamic. However, we do observe overlap in trade execution between some pairs of types for the two agents, which is interesting and warrants exploration in future work.

### 5 Partial Information Setting with no Prior via Online Learning

We now move to a more realistic setting, where agents neither have complete information, nor any *a priori* knowledge of the prior  $P$ . Instead, in this setting agents must learn via interaction. Specifically, we consider a mode of repeated game play that occurs over  $R \in \mathbb{N}^+$  rounds, where in each round  $r$  the game play follows the sequence of events for Bayesian game play described in Section 4, and the *ex-ante* strategy  $\mathbf{h}_i^r$  that each agent  $i$  selects in round  $r$  is chosen adaptive to their feedback following rounds 1 through  $r - 1$ . The feedback that each agent observes after each round is formalized by the following assumption:



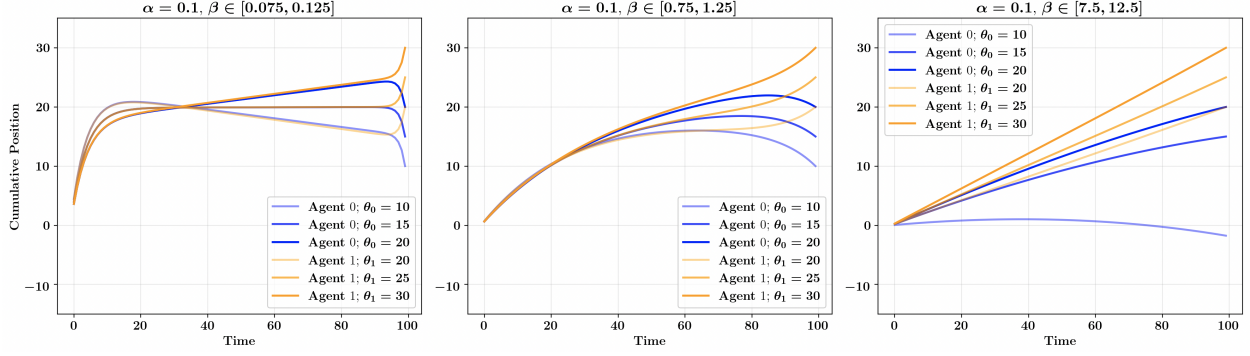


Figure 2: Cumulative position over time for agents under the BNE. Types are distributed uniformly and correspond to the constraint  $V$ . The type conditioned expected reserves are  $(3, 5, 7)$  for agent 1, and  $(6, 8, 10)$  for agent 2.  $p_0 = 0, \alpha = 0.1$  and the conditional  $\beta$  lies in the given range.

**Assumption 2** (End of Round Feedback). *Let  $\mathbf{h}_i^r$  and  $\theta_i^r$  denote the ex-ante strategy and sampled type for each agent  $i$  in round  $r$ , and let  $\boldsymbol{\lambda}^r$  denote the sampled market type. Then, at the end of round  $r$ , each agent observes the cost function  $c_i^r : \mathcal{H}_i \rightarrow \mathbb{R}$  as feedback, given by*

$$c_i^r(\mathbf{h}_i) = -u_i(\mathbf{h}_i(\theta_i^r); \mathbf{h}_{-i}^r(\theta_{-i}^r), \boldsymbol{\lambda}^r).$$

This feedback is the cost (negative utility) that the agent would have received if they had committed to a different strategy in that round, letting the strategies of all other agents and game instance be fixed. Although technically  $c_i^r$  could be written in a slightly simpler way with domain  $G_i(\theta_i^r)$  rather than  $\mathcal{H}_i$ , the latter is more convenient as it ensures that  $c_i^r$  is defined on the same domain for all  $r \in [R]$ . In many applications, the cost function  $c_i^r$  could be inferred by performing some regressions on observed market data to compute the sequence of aggregate outside demands  $\sum_{j \neq i} h_{j,t} + s_t$  along with  $\alpha$  and  $\beta$ . Finally, although we assume that agents observe the full cost function  $c_i^r$ , in practice many algorithms only require the gradient  $\nabla_{\mathbf{h}_i^r(\theta_i^r)} c_i^r(\mathbf{h}_i^r)$ , so in practice this assumption could be loosened. In particular, this is the case for the algorithm of Jordan et al. [14] discussed below.

To obtain concrete guarantees, we impose some boundedness regularity conditions, which are restrictions on the constraint sets, distribution over exogenous actions, and idiosyncratic utilities.

**Assumption 3** (Boundedness). *We assume that there exists some finite, fixed values  $B, S, U$ , and  $U'$ , such that for all agents  $i \in [n]$  and strategies  $\mathbf{h}_i \in \mathcal{H}_i$ , the following bounds hold almost surely: (1)  $\|\mathbf{h}_i(\theta_i)\|_2 \leq B$ ; (2)  $\|s\|_2 \leq S$ ; (3)  $|f_i(\mathbf{h}_i(\theta_i))| \leq U$ ; and (4)  $\|\nabla f_i(\mathbf{h}_i(\theta_i))\|_2 \leq U'$ .*

Next, we define an (unobserved) *population loss* for each agent  $i$  in any given round  $r$ , as follows:

**Definition 6** (Population Loss). *Let  $\mathbf{h}_i^r$  denote the ex-ante strategies of agent  $i$  in round  $r$  for all  $i$ . Then, the expected loss for each agent  $i$  in round  $r$  is a function  $\ell_i^r : \mathcal{H}_i \rightarrow \mathbb{R}$  given by:*

$$\ell_i^r(\mathbf{h}_i) = \mathbb{E}_{\theta_i, \theta_{-i}, \boldsymbol{\lambda} \sim P} [-u_i(\mathbf{h}_i(\theta_i); \mathbf{h}_{-i}^r(\theta_{-i}), \boldsymbol{\lambda})].$$

In other words,  $\ell_i^r(\mathbf{h}_i) = \mathbb{E}_{\mathcal{I} \sim P} [c_i^r(\mathbf{h}_i)]$  for all  $\mathbf{h}_i \in \mathcal{H}_i$ ; likewise for its derivatives. Therefore, the observed losses  $c_i^r$  or their gradients can be interpreted as unbiased stochastic estimates of the population losses  $\ell_i^r$  or their gradients respectively. It is easy to see that if  $\mathbf{h}_i^r$  minimizes  $\ell_i^r$  for every agent  $i$  simultaneously in some round  $r$ , then the agents are in BNE. Therefore,

this intuitively suggests that we could apply existing theory and algorithms for online convex optimization with stochastic feedback to establish convergence to equilibrium. We proceed by formalizing this intuition. First, we define some additional central concepts:

**Definition 7** (Regret). *Fix a agent  $i \in [n]$  and a sequence of loss functions  $\chi_i^r : \mathcal{H}_i \rightarrow \mathbb{R}$  for  $r \in [R]$ . For any sequence of strategies  $\mathbf{h}_i^r$  for  $r \in [R]$ , we define their (average) regret as*

$$\text{regret}(\mathbf{h}_i^1, \dots, \mathbf{h}_i^R; \chi_i^1, \dots, \chi_i^R) = \frac{1}{R} \sum_{r=1}^R \chi_i^r(\mathbf{h}_i^r) - \min_{\mathbf{h}_i \in \mathcal{H}_i} \frac{1}{R} \sum_{r=1}^R \chi_i^r(\mathbf{h}_i)$$

**Definition 8** ( $\epsilon$ -BCCE). *Let  $\sigma \in \Delta(\mathcal{H}_1 \times \dots \times \mathcal{H}_n)$  be a joint distribution over strategy profiles. Then,  $\sigma$  is an  $\epsilon$ -approximate Bayesian coarse correlated equilibrium ( $\epsilon$ -BCCE) if and only if for all agents  $i$ ,  $\theta_i \in \Theta_i$ , and  $\mathbf{h}_i' \in \mathcal{H}_i$ :*

$$\mathbb{E}_{\mathbf{h}_i, \mathbf{h}_{-i} \sim \sigma, \theta_{-i}, \boldsymbol{\lambda} \sim P|\theta_i} [u_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \geq \mathbb{E}_{\mathbf{h}_{-i} \sim \sigma, \theta_{-i}, \boldsymbol{\lambda} \sim P|\theta_i} [u_i(\mathbf{h}_i'(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] - \epsilon$$

Regret measures how much we could decrease our average loss by if, retrospectively, we swapped from the actual sequence of chosen strategies to some fixed alternative strategy. *No-regret algorithms* are well studied in the online learning literature; these are algorithms that ensure that the average regret converges to zero under arbitrarily (possibly adversarially) chosen loss functions. On the other hand, approximate BCCE is a weaker equilibrium notion than BNE, in two respects: (1) it allows for correlation between the strategies in equilibrium; and (2) it allows for  $\epsilon$ -sub-optimality of the chosen strategies (in conditional expectation given  $\theta_i$ ).

Although no-regret algorithms and BCCE may seem like unrelated ideas at first, they are deeply connected since multiple agents simultaneously following no-regret dynamics with a fixed game objective will induce an approximate coarse correlated equilibrium (CCE) in the game. Although we are considering Bayesian games and BCCE rather than CCE, similar reasoning gives us the following theorem (proof in Appendix D).

**Theorem 4.** *Suppose every agent  $i \in [n]$  selects their strategy  $\mathbf{h}_i^r$  at each round  $r$  via some on-line algorithm  $\text{Alg}_i$ , with the following properties: (1)  $\text{Alg}_i$  selects strategy  $\mathbf{h}_i^r$  at round  $i$  only using unbiased stochastic cost-function observations  $\tilde{\chi}_i^r$  for some true sequence of cost functions  $\chi_i^r \in \mathcal{C}_i$ , where  $\mathcal{C}_i$  is a set containing all population losses  $\ell_i^r$  almost surely; (2) it ensures that  $\text{regret}(\mathbf{h}_i^1, \dots, \mathbf{h}_i^R; \chi_i^1, \dots, \chi_i^R) \leq \epsilon_i(R)$  for some  $\epsilon_i(R)$  that is independent of the cost functions  $\chi_i^r \in \mathcal{C}_i$ , which could be adversarially chosen. In addition, let  $\sigma^R \in \Delta(\mathcal{H}_1 \times \dots \times \mathcal{H}_n)$  denote the uniform distribution over  $(\mathbf{h}_1^r, \dots, \mathbf{h}_n^r)$  across rounds. Then,  $\sigma^R$  is an  $\epsilon$ -BCCE, for some  $\epsilon$  that is bounded by  $\frac{\epsilon_i(R)}{\Pr(\theta_i)}$  for all  $i \in [n]$ ,  $\theta_i \in \Theta_i$ .*

We make a few comments on this theorem. First, we note that it is very general, and establishes convergence to equilibria for agents who simultaneously engage in no-regret learning using *any* stochastic-feedback no regret-learning algorithm with their observed costs  $c_i^r$ , and the algorithms could be different for each agent. Second, the restriction of losses in the theorem statements to some sets  $\mathcal{C}_i$  is necessary since adversarial no-regret will generally be impossible without some bounds on the allowed stochastic/true losses. In general, the constants involved in the regrets  $\epsilon_i(R)$  obtainable by a given algorithm may depend on  $\mathcal{C}_i$ , the choice of which in practice will depend on the bounds we can place on  $c_i^r$  and  $\ell_i^r$ . Finally, the dependence of the result on the worst-case  $1/\Pr(\theta_i)$  arises from bounding the conditional sub-optimality in the definition of BCCE uniformly over *all*  $i$  and  $\theta_i$ . For any given  $i, \theta_i$ , we can bound this sub-optimality by  $\epsilon_i(R)/\Pr(\theta_i)$ ,

so the presence of rarely-occurring types don't cause sub-optimality conditional on common types to suffer, and the *average* sub-optimality over all types is bounded by  $\epsilon_i(R)$  (see proof for details).

Although the guarantees from Theorem 4 are slightly weak in that they only ensure approximate convergence to a BCCE for the average-iterate strategy, not for the final iterate  $\mathbf{h}_1^R, \dots, \mathbf{h}_n^R$ , we can do better if all agents apply a *doubly optimal* algorithm [14], which is an algorithm that ensures no-regret, as well as last-iterate convergence to a NE if applied by all agents in a strongly monotone game with stochastic gradient feedback. The specific algorithm proposed by Jordan et al. [14] that obtains this property is online gradient descent (OGD) [30] with a specific stochastic scheme for reducing learning rates. We provide details of this algorithm (Alg. 2) and its theoretical guarantees in Appendix D.2. The following theorem establishes that, if all agents independently follow this algorithm, their joint strategy profile converges to the BNE.

**Theorem 5.** *Suppose all agents use Alg. 2 to decide their strategy in each round  $r$ . Then, letting  $k = \max_{i \in [n]} |\Theta_i|$ , the final iterate strategies  $\mathbf{h}_1^R, \dots, \mathbf{h}_n^R$  are an  $\epsilon$ -approximate BNE, for some  $\epsilon$  that satisfies the following in expectation over the algorithm's randomness:*

$$\mathbb{E}[\epsilon] = O \left( \frac{\text{poly}(n, T, k, \alpha, \beta, p_0, B, S, U')}{\min_{i \in [n], \theta_i \in \Theta_i} \Pr(\theta_i)} \cdot \frac{\log^{3/2}(R)}{\sqrt{R}} \right)$$

Even though this result only establishes convergence to BNE if all agents follow it, by the *doubly optimal* property discussed above it is no-regret. Thus, the algorithm is also robust to possibly adversarial environments. Compared with the extra-gradient algorithm (which we previously showed can efficiently compute the BNE), the benefits of Alg. 2 are two-fold: (1) each agent's learning procedure is *prior-independent*, since to update their strategies they only need gradient information about their realized cost; and (2) the procedure is fully decentralized, since agents can run their learning algorithms independently using only the information privately revealed to them. As with Theorem 4, our result depends on the worst-case  $1/\Pr(\theta_i)$ , but the same comments we made there apply about how this is not a major theoretical limitation. We finally note that the above concept of double optimality is very new, and it is possible that other decentralized algorithms could obtain similar guarantees, but this is a question for future research.

## 5.1 Simulations

We conclude with an empirical investigation on convergence to equilibrium in actual implementation, simulating repeated play as described in Section 5, where all agents follow Alg. 2. We use the same Bayesian game instance as in Section 4.1, with 2 agents, each having 3 possible types. We provide full details of this simulation setting in Appendix D.4.

We show the results of this simulation in Figure 3. On the left we directly compare the final iterate strategies from online learning with the BNE. We see that these almost exactly overlap, with only very minor discrepancies, which can be explained by noise in the observed market parameters. On the right we plot the convergence of the agent strategies during online learning to the BNE strategies in terms of mean-squared error (MSE); we see that the MSE very rapidly approaches 0, even faster than guaranteed by our theory. Overall, these results are a strong empirical validation of Theorem 5.

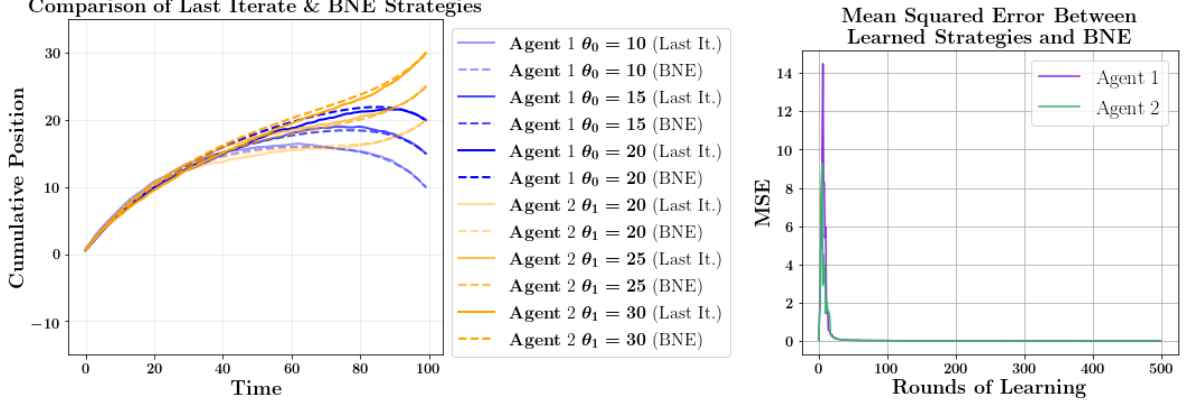


Figure 3: Comparison of Alg. 2 (over 500 rounds) to exact BNE strategies: (left) we plot the last-iterate strategies returned by Alg. 2 (solid lines) along with the true BNE (dashed lines) for all agents and types; and (right) we show the convergence in mean-squared error between the strategies from Alg. 2 and the BNE over the 500 rounds.

## References

- [1] Martin Aleksandrov and Toby Walsh. Online fair division: A survey. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13557–13562, 2020.
- [2] Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40, 2000.
- [3] Robert Almgren, Chee Thum, Emmanuel Hauptmann, and Hong Li. Direct estimation of equity market impact. *Risk*, 18(7):58–62, 2005.
- [4] Emmanuel Bacry, Adrian Iuga, Matthieu Lasnier, and Charles-Albert Lehalle. Market impacts and the life cycle of investors orders. *Market Microstructure and Liquidity*, 01(02):1550009, 2015. doi: 10.1142/S2382626615500094. URL <https://doi.org/10.1142/S2382626615500094>.
- [5] Siddhartha Banerjee, Vasilis Gkatzelis, Safwan Hossain, Billy Jin, Evi Micha, and Nisarg Shah. Proportionally fair online allocation of public goods with predictions. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 20–28, 2023.
- [6] Jean-Philippe Bouchaud, Yuval Gefen, Marc Potters, and Matthieu Wyart. Fluctuations and response in financial markets: the subtle nature of random price changes. *Quantitative finance*, 4(2):176, 2003.
- [7] Neil A Chriss. Competitive equilibria in trading. *arXiv preprint arXiv:2410.13583*, 2024.
- [8] Neil A Chriss. Optimal position-building strategies in competition. *arXiv preprint arXiv:2409.03586*, 2024.
- [9] Neil A Chriss. Position-building in competition with real-world constraints. *arXiv preprint arXiv:2409.15459*, 2024.
- [10] Neil A Chriss. Position building in competition is a game with incomplete information. *arXiv preprint arXiv:2501.01241*, 2025.

- [11] Gerard Debreu. *Theory of value: An axiomatic analysis of economic equilibrium*, volume 17. Yale University Press, 1959.
- [12] Jim Gatheral and Alexander Schied. Dynamical models of market impact and algorithms for order execution. *Handbook on Systemic Risk*, Jean-Pierre Fouque, Joseph A. Langsam, eds, pages 579–599, 2013.
- [13] Jason Hartline, Vasilis Syrgkanis, and Éva Tardos. No-regret learning in bayesian games. In *Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 2*, NIPS’15, page 3061–3069, Cambridge, MA, USA, 2015. MIT Press.
- [14] Michael Jordan, Tianyi Lin, and Zhengyuan Zhou. Adaptive, doubly optimal no-regret learning in strongly monotone and exp-concave games with gradient feedback. *Oper. Res.*, 73(3):1675–1702, May 2024. ISSN 0030-364X. doi: 10.1287/opre.2022.0446. URL <https://doi.org/10.1287/opre.2022.0446>.
- [15] Michael Kearns and Mirah Shi. Algorithmic aspects of strategic trading. *arXiv preprint arXiv:2502.07606*, 2025.
- [16] Galina M Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.
- [17] Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985.
- [18] Fengpei Li, Vitalii Ihnatiuk, Yu Chen, Jiahe Lin, Ryan J Kinnear, Anderson Schneider, Yuriy Nevmyvaka, and Henry Lam. Do price trajectory data increase the efficiency of market impact estimation? *Quantitative Finance*, 24(5):545–568, 2024.
- [19] Andreu Mas-Colell, Michael Dennis Whinston, Jerry R Green, et al. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.
- [20] Esteban Moro, Javier Vicente, Luis G. Moyano, Austin Gerig, J. Doyne Farmer, Gabriella Vaglica, Fabrizio Lillo, and Rosario N. Mantegna. Market impact and trading profile of hidden orders in stock markets. *Phys. Rev. E*, 80:066102, Dec 2009. doi: 10.1103/PhysRevE.80.066102. URL <https://link.aps.org/doi/10.1103/PhysRevE.80.066102>.
- [21] Hervé Moulin. *Fair division and collective welfare*. MIT press, 2004.
- [22] Anna A Obizhaeva and Jiang Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial markets*, 16(1):1–32, 2013.
- [23] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.
- [24] Robert W. Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973. doi: 10.1007/BF01737559.
- [25] Tim Roughgarden. Algorithmic game theory. *Communications of the ACM*, 53(7):78–86, 2010.
- [26] Supreeth Shastri and David Irwin. Cloud index tracking: Enabling predictable costs in cloud spot markets. In *Proceedings of the ACM Symposium on Cloud Computing*, pages 451–463, 2018.

- [27] Neha S Wadia, Yatin Dandi, and Michael I Jordan. A gentle introduction to gradient-based optimization and variational inequalities for machine learning. *Journal of Statistical Mechanics: Theory and Experiment*, 2024(10):104009, 2024.
- [28] Donald A Walker. Walras’s theories of tatonnement. *Journal of Political Economy*, 95(4): 758–774, 1987.
- [29] K. Webster. *Handbook of price impact modeling*. Chapman and Hall/CRC, 2023.
- [30] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, ICML’03, page 928–935. AAAI Press, 2003. ISBN 1577351894.

## Statement on Usage of LLMs

We used LLM tools for the purpose of checking language use in writing our paper, as well as a tool for suggesting relevant existing results when conducting our research (in particular, for exploring existing results related to uniqueness of Nash Equilibria, for which it suggested reading literature / existing results on variational inequalities.) However, all proofs were derived and written completely by the authors, and the paper was written completely by the authors (outside of the usage of LLMs for language checking as mentioned above.)

## A Additional Model Discussion

### A.1 Derivation of Price Model

First, we discuss in more detail how our price dynamics in Assumption 1 relate to Walrasian price dynamics. As mentioned in Section 2, this model positions that (mean) prices evolve according to the continuous time differential equation

$$dp_t = \alpha(\text{demand}_t - \text{supply}_t)dt,$$

for some price-sensitivity factor  $\alpha$ . Given this, the dynamic model for  $p_t^w$  can be viewed as a discretization of this process. In addition, allowing for noise in  $s_t$ , this turns it into a discretization of the corresponding stochastic differential equation

$$dp_t = \alpha(\text{demand}_t - \text{supply}_t)dt + \sigma dW_t,$$

for some noise process  $dW_t$  (*e.g.* Brownian motion). The actual price that traders must pay differs from this Walrasian process by amount  $\beta(\sum_{i=1}^n h_{i,t} + s_t)$ . We can interpret this difference as a temporary (instantaneous) price adjustment from  $p_t^w$  driven by the imbalance of supply and demand; when supply and demand are not balanced, the difference must be met by *market makers*, who provide liquidity. These market makers require some spread from  $p_t^w$  in order to account for the risk they take by providing liquidity. For example, if demand outstrips supply at time  $t$ , the market makers will balance this by selling an equal amount at a slight premium; hence, the instantaneous market price  $p_t$  will be slightly higher than  $p_t^w$ . We implicitly assume as part of Assumption 1 that this difference is linear in the imbalance  $\sum_{i=1}^n h_{i,t} + s_t$ , with coefficient  $\beta$ .

Alternatively, this model can also be justified from the literature on market impact. For example, the seminal model of Almgren and Chriss [2], which is the basis for much of the classical theory on (non-strategic) optimal trade execution, posits almost exactly the same model for price impact from trade execution over time, except that they consider a slightly more general offset based on supply and demand imbalance of the kind  $\psi(\sum_{i=1}^n h_{i,t} + s_t)$ , for some concave function  $\psi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ . Therefore, our price model is equivalent to theirs in the case of  $\psi(x) = \beta x$ . We note that empirical research (see *e.g.* Almgren et al. [3]) suggests power-law models of the kind  $\psi(x) = \beta x^\gamma$  with  $\gamma \approx 3/5$  to be well supported by real data. Such a model would be more challenging to study under strategic interaction, as it could break strong monotonicity without some additional assumptions on  $f_i$  and/or  $G_i$ ; we leave the investigation of alternative price models like this to future work.

## A.2 Relation of Model to Existing Models

Here, we make some concrete comparisons of our model with the models used in the recent lines of work on position building under competition Chriss [8, 9, 7, 10], Kearns and Shi [15].

**Relation to Existing Discrete Time Model** First, consider the special case of our model with no idiosyncratic utilities ( $f_i = 0$  for all  $i$ ), and no exogenous actions ( $\mathbf{s} = 0$ ). In this case, we can re-formulate the objective for each agent as minimizing a cost function  $c_i(\mathbf{h}_i, \mathbf{h}_{-i})$  given by the negative of the utility  $u_i$ , which if we unroll the autoregressive price definitions like in the proof of Lemma 1, we can easily verify is given by

$$c_i(\mathbf{h}_i; \mathbf{h}_{-i}) = \alpha \sum_{t=1}^T h_{i,t} \sum_{l=1}^t \sum_{j=1}^n h_{j,l} + \beta \sum_{t=1}^T h_{i,t} \sum_{j=1}^n h_{j,t} + \sum_{t=1}^T h_{i,t} p_0$$

Now, assume further that the constraints  $G_i$  contains a constraint of the kind  $\sum_{t=1}^T h_{i,t} = V_i$  for some fixed  $V_i$ . Then, the third term above can be ignored, as it is always equal to  $p_0 V_i$  for any feasible  $\mathbf{h}_i \in G_i$ . Given this, and with some slight re-arranging of terms, we have that the cost structure is given by

$$c_i(\mathbf{h}_i; \mathbf{h}_{-i}) = \alpha \sum_{t=1}^T h_{i,t} \sum_{j=1}^n x_{j,t-1} + (\alpha + \beta) \sum_{t=1}^T h_{i,t} \sum_{j=1}^n h_{j,t},$$

where we define

$$x_{i,t} = \sum_{l=1}^t h_{i,l}$$

as the cumulative position acquired by agent  $i$  over the first  $t$  time steps. This corresponds exactly to the kind of cost structure assumed in Kearns and Shi [15], who considered a discrete time version of optimal position building, with cost function

$$c_i^{\text{KS}}(\mathbf{h}_i; \mathbf{h}_{-i}) = \kappa \sum_{t=1}^T h_{i,t} \sum_{j=1}^n x_{j,t-1} + \sum_{t=1}^T h_{i,t} \sum_{j=1}^n h_{j,t}.$$

Following terminology for literature on optimal position building, they denote first term as the *permanent-impact cost*, and the second term as the *temporary-impact cost*, with permanent-impact coefficient  $\kappa$ , and unit temporary-impact coefficient (which is completely general up to normalization of cost). Therefore, if we normalize our cost by  $\alpha + \beta$ , we see that under the above model restrictions it recovers theirs with  $\kappa = \alpha / (\alpha + \beta)$ .

Although our model may seem less general given the above reduction, as they allow for any  $\kappa \geq 0$  but ours only allows  $\kappa \in [0, 1]$ , we argue that this restriction does not have much or any material impact in practice. First, as discussed in Kearns and Shi [15], if they decompose their cost into zero-sum and potential (*i.e.* congestion game-style cost) components, the coefficient in front of potential cost becomes negative when  $\kappa > 2$ . This implies that agents are rewarded rather than punished from congestion of their trading schedule, which therefore encourages agents to behave as aggressively as their constraints will allow (this is reflected *e.g.* in the unstable dynamics they observe when agents play no-regret with  $\kappa > 2$ ). Given this, we would probably wish to restrict to  $\kappa < 2$  in such a discrete model in practice. Second, and perhaps more importantly, to the extent



that their model is justified as a discretization of the continuous time model discussed below, the convergence of this discretization as we make it more and more fine-grained only works if we let the ratio of temporary-impact-coefficient to permanent-impact-coefficient (*i.e.*  $\kappa$ ) tend towards zero as  $T \rightarrow \infty$ . Therefore, no matter the target  $\kappa$  value in the continuous-time cost  $c_i^{\text{NC}}$  defined below, the corresponding  $\kappa$  in the discrete-time cost  $c_i^{\text{KS}}$  that approximates this will be less than 1 if the discretization is sufficiently fine-grained.

**Relation to Existing Continuous Time Model** The works by Chriss [8, 9, 7, 10] consider a continuous-time version of this problem, where the strategies  $\mathbf{h}_i$  are functions over some continuous time range (which they normalize to be  $[0, 1]$  without loss of generality). In this setting, we assume the strategies are defined by functions  $\mathbf{h}_i : [0, 1] \rightarrow \mathbb{R}$ , where  $\mathbf{h}_i(t)$  is their instantaneous trading rate at time  $t$ . We also define  $\mathbf{x}_i$  implicitly in terms of  $\mathbf{h}_i$  as the total accumulated position up to time  $t$ , which is mathematically given by

$$\mathbf{x}_i(t) = \int_0^t \mathbf{h}_i(l) dl.$$

Then, the assumed cost structure is

$$c_i^{\text{NC}}(\mathbf{h}_i; \mathbf{h}_{-i}) = \kappa \int_0^1 \mathbf{h}_i(t) \sum_{j=1}^n \mathbf{x}_i(t) dt + \int_0^1 \mathbf{h}_i(t) \sum_{j=1}^n \mathbf{h}_j(t) dt,$$

which is the continuous-time analogue of the cost structure based on decomposition into permanent-impact cost and temporary-impact cost mentioned above.<sup>2</sup>

Now, suppose we are given a problem instance of this continuous time model, with  $\kappa$  given, and time normalized into range  $[0, 1]$ . We can approximate this arbitrarily well with a discrete time model as  $T \rightarrow \infty$ , by letting the discrete time grid correspond to  $\{\frac{1}{T}, \frac{2}{T}, \dots, 1\}$  in continuous time. Specifically, we can do this as follows: suppose we are given collection of continuous-time strategies  $\mathbf{h}_1^c, \dots, \mathbf{h}_n^c$ , and define

$$\begin{aligned} \mathbf{x}_i^c(t) &= \int_0^t \mathbf{h}_i^c(l) dl && \text{(for continuous } t \in [0, 1]) \\ x_{i,t} &= \mathbf{x}_i^c\left(\frac{t}{T}\right) && \text{(for discrete } t \in [T]) \\ h_{i,t} &= x_{i,t} - x_{i,t-1} && \text{(for discrete } t \in [T]). \end{aligned}$$

---

<sup>2</sup>In reality, historically this continuous time model was the original one and the above discrete time version was introduced later, but we present in opposite order since our model is discrete-time.

Then, our discrete-time cost structure in terms of these strategy vectors  $\mathbf{h}_i$  will be given by

$$\begin{aligned}
c_i^{\alpha,\beta,T}(\mathbf{h}_i; \mathbf{h}_{-i}) &= \alpha \sum_{t=1}^T h_{i,t} \sum_{j=1}^n x_{j,t-1} + (\alpha + \beta) \sum_{t=1}^T h_{i,t} \sum_{j=1}^n h_{j,t} \\
&= \alpha \sum_{t=1}^T \left\{ \mathbf{x}_i^c\left(\frac{t}{T}\right) - \mathbf{x}_i^c\left(\frac{t-1}{T}\right) \right\} \sum_{j=1}^n \mathbf{x}_j^c\left(\frac{t-1}{T}\right) \\
&\quad + (\alpha + \beta) \sum_{t=1}^T \left\{ \mathbf{x}_i^c\left(\frac{t}{T}\right) - \mathbf{x}_i^c\left(\frac{t-1}{T}\right) \right\} \sum_{j=1}^n \left\{ \mathbf{x}_j^c\left(\frac{t}{T}\right) - \mathbf{x}_j^c\left(\frac{t-1}{T}\right) \right\} \\
&= \alpha \sum_{t=1}^T \frac{1}{T} \mathbf{h}_i^c\left(\frac{t - \gamma_{i,t}}{T}\right) \sum_{j=1}^n \mathbf{x}_j^c\left(\frac{t-1}{T}\right) \\
&\quad + (\alpha + \beta) \sum_{t=1}^T \frac{1}{T} \mathbf{h}_i^c\left(\frac{t - \gamma_{i,t}}{T}\right) \sum_{j=1}^n \frac{1}{T} \mathbf{h}_j^c\left(\frac{t - \gamma_{j,t}}{T}\right),
\end{aligned}$$

where the final line follows from the mean-value theorem, where  $\gamma_{i,t} \in (0, 1)$  for all  $i, t$ . Therefore, if we consider a sequence of discrete problem instances with  $\alpha = \kappa$  and  $\beta = T$ , we get

$$\begin{aligned}
\lim_{T \rightarrow \infty} c_i^{\kappa,T,T}(\mathbf{h}_i; \mathbf{h}_{-i}) &= \lim_{T \rightarrow \infty} \kappa \frac{1}{T} \sum_{t=1}^T \mathbf{h}_i^c\left(\frac{t - \gamma_{i,t}}{T}\right) \sum_{j=1}^n \mathbf{x}_j^c\left(\frac{t-1}{T}\right) \\
&\quad + \lim_{T \rightarrow \infty} \left( \frac{\kappa + T}{T} \right) \frac{1}{T} \sum_{t=1}^T \mathbf{h}_i^c\left(\frac{t - \gamma_{i,t}}{T}\right) \sum_{j=1}^n \mathbf{h}_j^c\left(\frac{t - \gamma_{j,t}}{T}\right) \\
&= \kappa \int_0^1 \mathbf{h}_i^c(t) \sum_{j=1}^n \mathbf{x}_j^c(t) dt + \int_0^1 \mathbf{h}_i^c(t) \sum_{j=1}^n \mathbf{h}_j^c(t) dt,
\end{aligned}$$

where first equality plugs in the above result with  $\alpha = \kappa$  and  $\beta = T$ , and the second follows from product of limits and the definition of the Riemann integral. Therefore, under appropriate re-normalization of the ratio  $\beta/\alpha$  as we make the discrete-time approximation more fine-grained, our model can approximate the existing continuous time cost structure considered in Chriss [8, 9, 7, 10] arbitrarily well if we let  $T \rightarrow \infty$ . Therefore, our model on the above restriction on idiosyncratic utilities, constraints, and exogenous actions subsumes theirs up to a vanishing discretization error.

## B Proofs for Section 3

### Proof of Lemma 1:

We first unroll the auto-regressive nature of the Walrasian price dynamic  $p_t^w$ . Observe that the following holds:

$$\begin{aligned} p_1^w &= p_0 + \alpha \sum_j h_{j,1} + \alpha s_1 ; \\ p_2^w &= p_0 + \alpha \sum_j h_{j,1} + \alpha s_1 + \alpha \sum_j h_{j,2} + \alpha s_2 ; \dots \end{aligned}$$

The execution price an agent pays is also influenced by the temporary impact. Combining this with the above, we can write the net cost an agent  $i$  faces as follows:  $u_i(\mathbf{h}_1, \dots, \mathbf{h}_n, \boldsymbol{\lambda})$

$$\begin{aligned} &= f_i(\mathbf{h}_i) - \sum_t h_{i,t} p_0 - \alpha \sum_{t=1}^T \sum_{\ell=1}^t h_{i,t} h_{i,\ell} - \alpha \sum_{t=1}^T \sum_{\ell=1}^t h_{i,t} \left( \sum_{j \neq i} h_{j,\ell} + s_\ell \right) - \beta \sum_{t=1}^T h_{i,t} \left( \sum_{j=1}^n h_{j,t} + s_t \right) \\ &= f_i(\mathbf{h}_i) - \underbrace{\alpha \sum_{t=1}^T \left( h_{i,t}^2 + h_{i,t} \sum_{\ell=1}^{t-1} h_{i,\ell} \right)}_{\text{quadratic terms}} - \beta \sum_{t=1}^T h_{i,t}^2 \\ &\quad - \underbrace{\alpha \sum_{t=1}^T h_{i,t} \sum_{\ell=1}^t \sum_{j \neq i} h_{j,\ell} - \beta \sum_{t=1}^T h_{i,t} \sum_{j \neq i} h_{j,t}}_{\text{linear terms } \propto \text{other agent}} - \underbrace{\alpha \sum_{t=1}^T h_{i,t} \sum_{\ell=1}^t s_\ell - \beta \sum_{t=1}^T h_{i,t} s_t - \sum_t p_0 h_{i,t}}_{\text{linear term } \propto \text{exogenous agent}} \end{aligned}$$

Focusing on the quadratic terms, it suffices to compute the Hessian, denoted by  $Q$ . Note that  $Q_{t,t} = 2\alpha + 2\beta$ . As for the off-diagonal values, these are composed entirely of  $\alpha$ . Indeed, for any  $t_1 \neq t_2$ , we have that  $Q_{t_1,t_2} = \alpha$ . Next, we consider the linear terms that are proportional to other agents. We wish to express it in the following form:  $\sum_{j \neq i} (A \mathbf{h}_j(\theta_j))^T \mathbf{h}_i(\theta_i)$ . For a given  $t$ , consider the first of the two linear terms proportional to others. For any  $j$ , observe that  $h_{i,t}(\theta_i)$  is multiplied by  $\alpha h_{j,1}(\theta_j), \dots, \alpha h_{j,t}(\theta_j)$ . As for the second term, it multiplies  $h_{i,t}(\theta_i)$  with  $\beta h_{j,t}(\theta_j)$ . Hence, we conclude that  $A$  is a lower-triangular matrix, whose diagonals are  $\alpha + \beta$  and the remaining values are  $\alpha$ . As for the linear term with respect to the exogenous agent, it follows a similar pattern, and we can express it as  $(A \mathbf{s})^T \mathbf{h}_i$ . We thus have the following expression for the matrices  $Q$  and  $A$ :

$$Q_{ij} = \begin{cases} \alpha & \text{if } i < j \\ 2\alpha + 2\beta & \text{if } i = j \\ \alpha & \text{if } i > j \end{cases} ; A_{ij} = \begin{cases} 0 & \text{if } i < j \\ \alpha + \beta & \text{if } i = j \\ \alpha & \text{if } i > j \end{cases}$$

Since  $Q$  is a symmetric matrix that can be written as  $Q = \alpha J + (\alpha + 2\beta)I$  where  $J$  is the all 1s matrix and  $I$  the identity matrix. Observe that for any  $\mathbf{x}$ , we have that:

$$\mathbf{x}^T Q \mathbf{x} = (\alpha + 2\beta)(\mathbf{x}^T I \mathbf{x}) + \alpha(\mathbf{x}^T J \mathbf{x}) = (\alpha + 2\beta)(\mathbf{x}^T \mathbf{x}) + \alpha(\mathbf{x}^T J \mathbf{x}) = (\alpha + 2\beta)\|\mathbf{x}\|_2^2 + \alpha \left( \sum_{i=1}^T x_i \right)^2 > 0$$

where the strict inequality holds since the parameters  $\alpha, \beta$  are non-negative and  $\mathbf{x} \neq 0$ . In other words, the  $Q$  matrix is positive definite and thus the utility of each agent, in terms of their own strategy, is a strictly concave function.

## Proof of Lemma 2

*Proof.* Consider an agent  $i$  and suppose all other agents are playing a mixed and possibly correlated strategy, denoted by  $\sigma_{-i}$ . Buyer  $i$  can choose to best-respond with a mixed strategy of their own, denoted  $p_i(h)$ . Without loss of generality, we set  $p_0 = 0$  and observe that:

$$\begin{aligned}
\text{br}_i(\sigma_{-i}) &= \arg \max_{p \in \Delta(G_i)} \int_{\mathbf{h}_i} p(\mathbf{h}_i) \left( f_i(\mathbf{h}_i) - sA\mathbf{h}_i - \int_{\mathbf{h}_{-i}} \sigma(\mathbf{h}_{-i}) [\mathbf{h}_i^T Q \mathbf{h}_i + \sum_{j \neq i} (A\mathbf{h}_j)^T \mathbf{h}_i] d\mathbf{h}_{-i} \right) d\mathbf{h}_i \\
&= \arg \max_{p \in \Delta(G_i)} \int_{\mathbf{h}_i} p(\mathbf{h}_i) \left[ f_i(\mathbf{h}_i) - \mathbf{h}_i^T Q \mathbf{h}_i - sA\mathbf{h}_i - \underbrace{\left( \int_{\mathbf{h}_{-i}} \sigma(\mathbf{h}_{-i}) \sum_{j \neq i} (A\mathbf{h}_j)^T d\mathbf{h}_{-i} \right) \mathbf{h}_i}_{\mathbf{v}_{-i}^T} \right] d\mathbf{h}_i \\
&= \arg \max_{p \in \Delta(G_i)} \int_{\mathbf{h}_i} p(\mathbf{h}_i) \left[ f_i(\mathbf{h}_i) - \mathbf{h}_i^T Q \mathbf{h}_i - \mathbf{v}_{-i}^T \mathbf{h}_i - sA\mathbf{h}_i \right] d\mathbf{h}_i \\
&= \arg \max_{\mathbf{h}_i \in G_i} \left[ f_i(\mathbf{h}_i) - \mathbf{h}^T Q \mathbf{h} - \mathbf{v}_{-i}^T \mathbf{h}_i - sA\mathbf{h}_i \right]
\end{aligned}$$

where the last equality follows from the linearity of expectation. The resulting maximization expression has a sole quadratic term:  $\mathbf{h}_i^T Q \mathbf{h}_i$ . From Lemma 1 we know  $Q$  is a PD matrix. Thus, the best-response optimization, to both pure or mixed strategies of others, is strictly concave, and there is a unique solution.

Interestingly, this result implies that in the complete information setting, any coarse correlated equilibrium (CCE) must also be a Nash. Indeed, for any  $\sigma$  to be a CCE, each agent's strategy must be pure (otherwise there is a deviating pure strategy that is a best response for any mixed strategy of others). A CCE with pure strategy is the same as Nash.  $\square$

### B.1 Proof of Theorem 1

*Proof. Uniqueness:* From 2, we know that each agent's best response is a concave optimization problem over a convex region  $G_i$ . In this proof, we shall express the agent's objective from a cost minimization perspective – i.e.  $c_i(\mathbf{h}_i; \mathbf{h}_{-i}, \boldsymbol{\lambda}) = -u_i(\mathbf{h}_i; \mathbf{h}_{-i}, \boldsymbol{\lambda})$ . As such, the best response objective will be convex. Formally (again setting  $p_0 = 0$  for ease of exposition):

$$\text{br}_i(\mathbf{h}_{-i}) = \arg \min_{\mathbf{h}_i \in G_i} \frac{1}{2} \mathbf{h}_i^T Q \mathbf{h}_i + \sum_{j \neq i} (A\mathbf{h}_j)^T \mathbf{h}_i + sA\mathbf{h}_i - f_i(\mathbf{h}_i) := \arg \min_{\mathbf{h}_i \in G_i} c_i(\mathbf{h}_i; \mathbf{h}_{-i}, \boldsymbol{\lambda})$$

At a best response for agent  $i$ ,  $\mathbf{h}_i^*$ , it must be that for all  $\forall \mathbf{h}_i \in G_i : \langle \nabla_{\mathbf{h}_i} c_i(\mathbf{h}_i^*; \mathbf{h}_{-i}), (\mathbf{h}_i - \mathbf{h}_i^*) \rangle \geq 0$ ; otherwise, the agent could move in that direction and decrease their net cost. This is a standard equivalence between a convex optimization and variational inequalities (see Rockafellar and Wets [23]). At a Nash Equilibrium, each buyer  $i$  must be playing their best response, given the strategies of other agents. Thus, we are looking for a set of trajectories  $(\mathbf{h}_1^{eq}, \dots, \mathbf{h}_n^{eq})$  such that:

$$\forall i, \forall (\mathbf{h}_1, \dots, \mathbf{h}_n) \in R_1 \times \dots \times R_n : \langle \nabla_{\mathbf{h}_i} c_i(\mathbf{h}_i^{eq}; \mathbf{h}_{-i}^{eq}), (\mathbf{h}_i - \mathbf{h}_i^{eq}) \rangle \geq 0 \quad (3)$$

Indeed, any tuple  $(\mathbf{h}_1^{eq}, \dots, \mathbf{h}_n^{eq})$  that satisfies the above must be a Nash Equilibrium. Observe that  $c_i$  is the sum of a quadratic function and a convex term. The gradient of  $c_i$  is then as follows:

$$\nabla_{\mathbf{h}_i} c_i(\mathbf{h}_i^*; \mathbf{h}_{-i}) = Q\mathbf{h}_i^* + \sum_{j \neq i} A\mathbf{h}_j + A\mathbf{s} - \nabla_{\mathbf{h}_i} f_i(\mathbf{h}_i)$$

Ignoring the  $\nabla_{\mathbf{h}_i} f_i(\mathbf{h}_i)$  term, the gradient is a linear function of all agent strategies. Denoting  $\mathbf{x} = [\mathbf{h}_1, \dots, \mathbf{h}_n]^T$  as the concatenation of agent strategies and  $\mathbf{s}$  as a constant (since this is not from a strategic agent), the variational inequality that characterizes the Nash Equilibrium can thus written with an operator  $F(\mathbf{x}) = M\mathbf{x} + b - [\nabla_{\mathbf{h}_1} f_1(\mathbf{h}_1), \dots, \nabla_{\mathbf{h}_n} f_n(\mathbf{h}_n)]$ . That is, a set of strategies  $\mathbf{x}^{eq} = [\mathbf{h}_1^{eq}, \dots, \mathbf{h}_n^{eq}]$  is a Nash Equilibrium if and only if, for all  $\mathbf{x} \in R_1 \times \dots \times R_n$ :  $\langle F(\mathbf{x}^{eq}), (\mathbf{x} - \mathbf{x}^{eq}) \rangle \geq 0$ . We note the following:

**Definition 9** ([23]). *An operator  $F$  is called strongly monotone on a set  $\mathcal{X}$  if and only if there exists a scalar  $c$  such that:*

$$\langle F(\mathbf{x}) - F(\mathbf{x}'), (\mathbf{x} - \mathbf{x}') \rangle \geq c \|\mathbf{x} - \mathbf{x}'\|^2, \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$$

From [23], we note that a variational inequality with a strongly monotone operator has a unique solution. Here, this implies a unique Nash Equilibrium. For our operator  $F$ :

$$\langle F(\mathbf{x}) - F(\mathbf{x}'), (\mathbf{x} - \mathbf{x}') \rangle = \langle M(\mathbf{x} - \mathbf{x}'), (\mathbf{x} - \mathbf{x}') \rangle - \sum_{i=1}^n \langle \nabla_{\mathbf{h}_i} f_i(\mathbf{h}_i) - \nabla_{\mathbf{h}'_i} f_i(\mathbf{h}'_i), (\mathbf{h}_i - \mathbf{h}'_i) \rangle$$

As such, it suffices to show that for positive  $c$ :  $\langle M(\mathbf{x} - \mathbf{x}'), (\mathbf{x} - \mathbf{x}') \rangle \geq c \|\mathbf{x} - \mathbf{x}'\|^2$  and for all  $i$ :  $-\langle \nabla_{\mathbf{h}_i} f_i(\mathbf{h}_i) - \nabla_{\mathbf{h}'_i} f_i(\mathbf{h}'_i), (\mathbf{h}_i - \mathbf{h}'_i) \rangle \geq 0$ . It is known that for convex functions, their gradient is a monotone operator. Thus,  $f_i$  being concave (and thus  $-f_i$  is convex) and the desired condition immediately holds. As such, it suffices to prove the strong monotonicity of the linear operator  $M$  and show that for any  $\mathbf{x}$ :  $\langle M\mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^T M \mathbf{x} \geq c \|\mathbf{x}\|^2$ . The matrix  $M$  is given by the following block matrix:

$$M = \begin{bmatrix} Q \in \mathbb{R}^{T \times T} & A \in \mathbb{R}^{T \times T} & \dots & A \in \mathbb{R}^{T \times T} \\ A \in \mathbb{R}^{T \times T} & Q \in \mathbb{R}^{T \times T} & \dots & A \in \mathbb{R}^{T \times T} \\ \vdots & \vdots & \ddots & \vdots \\ A \in \mathbb{R}^{T \times T} & A \in \mathbb{R}^{T \times T} & \dots & Q \in \mathbb{R}^{T \times T} \end{bmatrix} \quad (4)$$

We first note that while our matrix  $M$  may not symmetric, we can always write it as the sum of a symmetric component  $M_s = \frac{1}{2}(M + M^T)$  and a skew-symmetric component  $M_k = \frac{1}{2}(M - M^T)$ . By definition,  $M_k^T = -M_k$ . This means that for any  $\mathbf{x}$  where  $s = \mathbf{x}^T M_k \mathbf{x}$ :  $s = s^T = (\mathbf{x}^T M_k \mathbf{x})^T = -\mathbf{x}^T M_k \mathbf{x} = -s$ . Thus,  $\mathbf{x}^T M_k \mathbf{x} = 0$  and it suffices to only consider the symmetric component  $M_s$  for the strong monotonicity condition. Now suppose  $M_s$  is a positive definite matrix. Then we can always diagonalize it as  $P\Lambda P^T$ , where  $\Lambda$  is a diagonal matrix of positive eigenvalues and  $PP^T = P^T P = I$ . Then we can also express any  $\mathbf{x} = P\mathbf{y}$  for some  $\mathbf{y}$ . Then under this PD condition, we have:

$$\begin{aligned} \mathbf{x}^T M \mathbf{x} &= \mathbf{y}^T P^T P \Lambda P^T P \mathbf{y} = \mathbf{y}^T \Lambda \mathbf{y} \\ &\geq \lambda_{\min}(M_s) (P^T \mathbf{x})^T (P^T \mathbf{x}) = \lambda_{\min}(M_s) \mathbf{x}^T P P^T \mathbf{x} = \lambda_{\min}(M_s) \|\mathbf{x}\|^2 \end{aligned}$$

In other words, if  $M_s$  is positive definite, we can choose  $c = \lambda_{\min}(M_s) > 0$  and satisfy the conditions of strong monotonicity. The matrix  $M_s$  is an  $n \times n$  block matrix with  $Q$  on the diagonal and all other elements being  $A_s = \frac{1}{2}(A + A^T)$ . This can be succinctly represented using the Kronecker product (recall  $J_n$  is an  $n \times n$  all 1s matrix):

$$M_s = I_n \otimes (Q - A_s) + J_n \otimes A_s \quad (5)$$

Note that the all 1s matrix is positive-definite with one eigenvalue of  $(n-1)$  and all other eigenvalues 0. Therefore, we can write  $\Lambda_{J_n} = U^T J_n U$ , where  $\Lambda_{J_n} = \text{diag}(n, 0, \dots, 0)$ . Let  $P = U \otimes I_T$ , and note

that  $P^T P = (U^T \otimes I_T)(U \otimes I_T) = U^T U \otimes I_T = I_{nT}$ , where we use the mixed product property of Kronecker products. We shall be using  $P$  to diagonalize (in the block sense) the matrix  $M_s$ . Specifically, observe that due to the mixed product rule:

$$\begin{aligned} P^T M_s P &= P^T (I_n \otimes Q - A_s) P + P^T (J_n \otimes A_s) P \\ &= (U^T \otimes I_T)(I_n \otimes Q - A_s)(U \otimes I_T) + (U^T \otimes I_T)(J_n \otimes A_s)(U \otimes I_T) \\ &= (U^T I_n U \otimes I_T(Q - A_s)I_T) + (U^T J_n U \otimes I_T A_s I_T) \\ &= (I_n \otimes (Q - A_s)) + (\Lambda_{J_n} \otimes A_s) \end{aligned}$$

The first summand is a block diagonal matrix with  $Q - A_s$  in each entry, and the second summand is also block diagonal with  $nA_s$  in the first entry and 0 elsewhere. Therefore,  $P^T M_s P$  results in a block diagonal matrix  $\text{diag}(Q_s + (n-1)A_s, A_s, \dots, Q - A_s)$ . The eigenvalues of  $M_s$  are the eigenvalues of this block diagonal matrix, which in turn are the eigenvalues of each matrix in the diagonal. Thus, we need to show that  $Q + (n-1)A_s$  and  $Q - A_s$  both have positive eigenvalues. Note that  $Q = (\alpha + 2\beta)I_T + \alpha J_T$  and  $A_s = (\frac{\alpha}{2} + \beta)I_T + \frac{\alpha}{2}J_T$ . Thus, for any  $\mathbf{x} \in \mathbb{R}^T$ :

$$\begin{aligned} \mathbf{x}^T (Q - A_s) \mathbf{x} &= \mathbf{x}^T \left[ \left( \frac{\alpha}{2} + b \right) I_T + \frac{\alpha}{2} J_T \right] \mathbf{x} = \left( \frac{\alpha}{2} + b \right) \mathbf{x}^T \mathbf{x} + \frac{\alpha}{2} \left( \sum_{t=1}^T x_t \right)^2 > 0 \\ \mathbf{x}^T (Q + (n-1)A_s) \mathbf{x} &= \mathbf{x}^T \left[ (n+1) \left( \frac{\alpha}{2} + b \right) I_T + (n+1) \frac{\alpha}{2} J_T \right] \mathbf{x} \\ &= (n+1) \left( \frac{\alpha}{2} + b \right) \mathbf{x}^T \mathbf{x} + (n+1) \frac{\alpha}{2} \left( \sum_{t=1}^T x_t \right)^2 > 0 \end{aligned}$$

as long as either  $\alpha > 0$  or  $\beta > 0$ . Since these diagonal matrices are positive definite, they have positive eigenvalues, implying  $M_s$  has positive eigenvalues, implying strong monotonicity of the simultaneous best response operator, implying uniqueness of the equilibrium.

### Linear Convergence:

---

#### Algorithm 1: Extra-Gradient Algorithm

---

**Input:** Game Instance  $\mathcal{I}$ , Variational Operator  $F$ , step-size  $\eta$

Randomly Initialize a feasible joint strategy  $\mathbf{x}_0 = (\mathbf{h}_1, \dots, \mathbf{h}_n)$

**while**  $\|\mathbf{x}_r - \mathbf{x}_{r-1}\| \leq \varepsilon$  **do**

$\mathbf{x}_{r+1/2} = \text{Proj}_{G_1 \times G_n}(\mathbf{x}_r - \eta F(\mathbf{x}_r))$   
     $\mathbf{x}_{r+1} = \text{Proj}_{G_1 \times G_n}(\mathbf{x}_r - \eta F(\mathbf{x}_{r+1/2}))$

---

Theorem 3.4 of Wadia et al. [27] states that for any  $c$ -strongly monotone and  $L$ -Lipshcitz operator, the extragradient algorithm (Algorithm 1) with step-size  $\eta = \frac{1}{2(c+L)}$  converges to the fixed point at a linear rate of  $1 - \frac{c}{4L}$ . We have shown above that our given operator is  $c = \lambda_{\min}(M_s)$  strongly monotone. As for Lipschitzness, note that our operator can be decomposed as:  $F(\mathbf{x}) = M\mathbf{x} + b - J(\mathbf{x})$ , where  $J(\mathbf{x}) = [\nabla_{\mathbf{h}_1} f_1, \dots, \nabla_{\mathbf{h}_n} f_n]^T$ . Lipschitz constants for the sum of two maps add; thus, it suffices to solve for the Lipschitz constants for the linear operator  $M$ ,  $L_M$  and the gradient operator  $J$ ,  $L_J$ .

Any linear operator is lipschitz – in fact, the Lipschitz constant is just the 2-norm of the matrix  $M$ . For any matrix, the following is always true:  $\|M\|_2 \leq \sqrt{\|M\|_1 \|M\|_\infty}$ , where  $\|M\|_1$  is the

largest absolute column sum and  $\|M\|_\infty$  is the largest absolute row sum. In our specific matrix  $M$ , observe that:

$$\|M\|_1 = \|M\|_\infty = (2\alpha + 2\beta) + (T-1)\alpha + (n-1)[(T-1)\alpha + \alpha + \beta] = (nT+1)\alpha + (n+1)\beta \geq \|M\|_2$$

Thus,  $L_M = (nT+1)\alpha + (n+1)\beta$  is a suitable bound for the  $M$  operator Lipschitz constant. Secondly, for any  $i$ , observe that since  $f_i$  is concave, the operator  $\nabla_{\mathbf{h}_i} f_i$  is  $L_i = \sup_x \lambda_{\max}(-\nabla_{\mathbf{h}_i} f_i)$  lipschitz. Further, observe that:

$$\|J(\mathbf{x}) - J(\mathbf{x}')\|^2 = \sum_{i=1}^n \|\nabla_{\mathbf{h}_i} f_i(\mathbf{h}_i) - \nabla_{\mathbf{h}'_i} f_i(\mathbf{h}'_i)\|^2 \leq \max_i L_i \sum_{i=1}^n \|\mathbf{h}_i - \mathbf{h}'_i\|^2 = \max_i L_i \|\mathbf{x} - \mathbf{x}'\|^2$$

Therefore,  $L_J = \max_i L_i$  and the overall Lipschitz constant is  $L_J + L_M = \max_i \sup_x \lambda_{\max}(-\nabla_{\mathbf{h}_i} f_i) + (nT+1)\alpha + (n+1)\beta$ .  $\square$

## B.2 Proof of Theorem 2

*Proof.* Suppose there are  $n = 2$  agents and we have some constant values of  $\alpha, \beta$  – one can assume, without loss of generality, that  $\alpha = \beta = 1^3$ . Let the final position utility for both agents be given by the following linear function:  $u_i(\mathbf{h}_i) = r_i \sum_t h_{i,t}$ , where  $r_i$  can be interpreted as the reserve/fair-market price as perceived by agent  $i$ . Further, the two have box constraints on their cumulative position:  $V_i^- \leq \sum_i h_{i,t} \leq V_i^+$ . We shall assume the exogenous agent is not present – i.e.  $\mathbf{s} = \mathbf{0}$ .

For a positive constant  $x$ , let the initial price  $p_0 = x$  and the reserve prices for the agents be  $(r_1 = x, r_2 = x - \varepsilon)$ , where  $\varepsilon > 0$ . We first consider the equilibrium of this game without any constraints. Then each agent's best response is given by:

$$\text{br}_1(\mathbf{h}_2) = \arg \max_{\mathbf{h}_1} \left\{ -\frac{1}{2} \mathbf{h}_1^T Q \mathbf{h}_1 - (A \mathbf{h}_2)^T \mathbf{h}_1 \right\} \quad (6)$$

$$\text{br}_1(\mathbf{h}_2) = \arg \max_{\mathbf{h}_2} \left\{ -\varepsilon \mathbf{1}^T \mathbf{h}_2 - \frac{1}{2} \mathbf{h}_2^T Q \mathbf{h}_2 - (A \mathbf{h}_1)^T \mathbf{h}_2 \right\} \quad (7)$$

Observe that at the equilibrium of this unconstrained game, the gradient of both agents' best responses must be 0. Since this is a quadratic function, the gradient is linear, and the equilibrium can be uniquely specified by the following system of linear equalities:

$$\underbrace{\begin{bmatrix} Q & A \\ A & Q \end{bmatrix}}_{\text{Matrix } M \in \mathbb{R}^{2T \times 2T}} \underbrace{\begin{bmatrix} \mathbf{h}_1^{eq} \\ \mathbf{h}_2^{eq} \end{bmatrix}}_{\mathbf{z} \in \mathbb{R}^{2T}} = \underbrace{\begin{bmatrix} \mathbf{0} \\ -\varepsilon \end{bmatrix}}_{\mathbf{z} \in \mathbb{R}^{2T}}$$

Recall that the matrices  $Q, A$  are specified using only the terms  $\alpha, \beta$ . In lemma 2, we noted that  $Q$  is a positive-definite matrix and thus invertible. The matrix  $A$  is a lower triangular matrix with  $\alpha + \beta$  on the diagonals and is thus also invertible (insofar as  $\alpha > 0$  or  $\beta > 0$ ). As such, the matrix  $M$  above is invertible and the unconstrained equilibrium strategy is given by  $M^{-1} \mathbf{z}$ . Note that this does not depend on the value of  $x$ . Further, if  $V_i^- \leq -\|M^{-1} \mathbf{z}\|_1$  and  $V_i^+ \geq \|M^{-1} \mathbf{z}\|_1$ , then this unconstrained equilibrium is also an equilibrium in the original constrained game. As for the

<sup>3</sup>Insofar as  $\alpha, \beta$  are constants and not scaling with respect to the  $\varepsilon$  all results will hold.

strategy itself, let  $m_{ij}$  denote the values of  $-M^{-1}$  and note that  $m_{ij}$  can be seen as a scalar with respect to  $\varepsilon$ . Then we have that:

$$h_{1t} = \varepsilon \sum_{j=T}^{2T} m_{t,j} \quad \text{and} \quad h_{2t} = \varepsilon \sum_{j=T}^{2T} m_{T+t,j} \quad (8)$$

Given that the value of the final position is simply the product of the total amount bought and the reserve, the utility of buyer 1 (with reserve  $x$ ) is:

$$\begin{aligned} u_{eq}^1 &= x \varepsilon \underbrace{\sum_{t=1}^T \sum_{j=T}^{2T} m_{t,j}}_{\sum_t h_{1t}} - \sum_{t=1}^T \underbrace{\left[ \sum_{j=T}^{2T} m_{t,j} \varepsilon \left( x + \underbrace{\alpha \varepsilon \sum_{\tau=1}^t \sum_{j=T}^{2T} (m_{\tau,j} + m_{T+\tau,j})}_{\text{price } p_t} + \beta \varepsilon \sum_{j=T}^{2T} m_{t,j} + m_{T+t,j} \right) \right]}_{\text{price } p_t} \\ &= \left| \sum_{t=1}^T \sum_{j=T}^{2T} m_{t,j} \varepsilon^2 \left( \alpha \sum_{\tau=1}^t \sum_{j=T}^{2T} (m_{\tau,j} + m_{T+\tau,j}) + \beta \sum_{j=T}^{2T} (m_{t,j} + m_{T+t,j}) \right) \right| = \Theta(\varepsilon^2) \end{aligned}$$

where the absolute value in the second line follows, since utility at equilibrium will always be non-negative (the agents not trading would get utility 0, so utility at equilibrium must be at least 0). A similar analysis leads us to show that the utility of the second agent (with reserve  $x - \varepsilon$ ) is also bounded by  $\Theta(\varepsilon^2)$ , allowing us to conclude that the welfare at equilibrium is  $O(\varepsilon^2)$ . Formally:

$$u_2^{eq} = \left| -\varepsilon^2 \sum_{t=1}^T \sum_{j=T}^{2T} m_{T+t,j} - \sum_{t=1}^T \sum_{j=T}^{2T} m_{T+t,j} \varepsilon^2 \left( \alpha \sum_{\tau=1}^t \sum_{j=T}^{2T} (m_{\tau,j} + m_{T+\tau,j}) + \beta \sum_{j=T}^{2T} (m_{t,j} + m_{T+t,j}) \right) \right|$$

We now turn to characterizing the optimal welfare of this instance. For some  $\delta > 0$  (to be specified later), consider the following trajectories for each buyer (recall positive values mean buying):

$$\mathbf{h}_1 = [x, x, 0, \dots, 0] \quad \text{and} \quad \mathbf{h}_2 = [-x - \delta, -x - \delta, 0, \dots, 0] \quad (9)$$

Insofar as  $V_i^+ \geq 2x$  and  $V_i^- \leq -2x - \delta$ , the trajectories above are feasible. Under this strategy, it suffices to consider the prices at rounds  $t = 1, 2$ , for which we have that:  $p_1 = x - \alpha\delta - \beta\delta$  and  $p_2 = x - 2\alpha\delta - \beta\delta$ . Then the utilities for each buyer is given by:

$$\begin{aligned} u_1 &= x \cdot 2x - x(x - \alpha\delta - \beta\delta) - x(x - 2\alpha\delta - \beta\delta) = 3\alpha\delta x + 2\beta\delta x \\ u_2 &= (x - \varepsilon)(-2x - \delta) + (x + \delta)(x - \alpha\delta - \beta\delta) + (x + \delta)(x - 2\alpha\delta - \beta\delta) \\ &= 2\varepsilon x + 2\delta\varepsilon - 3\alpha\delta x - 3\alpha\delta^2 - 2\beta\delta x - 2\beta\delta^2 \\ \implies u_1^{opt} + u_1^{opt} &\geq 2\varepsilon x + 2\delta\varepsilon - 3\alpha\delta^2 - 2\beta\delta^2 = 2x\varepsilon + 2\delta\varepsilon - (3\alpha + 2\beta)\delta^2 \end{aligned}$$

This gives a concave quadratic (in the unspecified parameter  $\delta$ ) lower bound on the optimal utility. Maximizing it means choosing a  $\delta$  such that the gradient is 0:

$$\delta = \frac{\varepsilon}{3\alpha + 2\beta} \implies u_1^{opt} + u_1^{opt} \geq 2x\varepsilon + \frac{2\varepsilon^2}{3\alpha + 2\beta} - \frac{\varepsilon^2}{3\alpha + 2\beta} = 2x\varepsilon + \frac{\varepsilon^2}{3\alpha + 2\beta} = \Theta(x\varepsilon)$$

From here, it is evident that for any constants  $\alpha, \beta$  and  $x$ , we can construct an  $\varepsilon > 0$  parametrized instance  $\mathcal{I}_\varepsilon$  with box constraints  $V_i^- \leq \min(-\|M^{-1}\mathbf{z}\|, -2x - 2\delta)$  and  $V_i^+ \geq \max(\|M^{-1}\mathbf{z}\|, 2x)$  with the aforementioned  $\delta = \frac{\varepsilon}{3\alpha + 2\beta}$  such that:

$$\text{PoA}(\mathcal{I}_\varepsilon) = \frac{U_{opt}(\mathcal{I}_\varepsilon)}{U_{eq}(\mathcal{I}_\varepsilon)} \geq \frac{\Omega(\varepsilon)}{O(\varepsilon^2)} = \Omega\left(\frac{1}{\varepsilon}\right) \rightarrow \infty \quad \text{as } \varepsilon \rightarrow 0 \quad (10)$$

□



## C Proofs and Details For Section 4

### C.1 Proof of Lemma 3

*Proof.* In the most general sense, observe that agent  $i$ 's best response for a realized type  $\theta_i$  allows them to play a mixed strategy over all valid strategies:  $p_i(\mathbf{h}_i|\theta_i)$ , where  $\mathbf{h}_i$  is a vector in  $\mathbb{R}^T$  since the probability is already conditioned on  $\theta_i$ . Suppose the remaining agents are playing some mixed, possibly correlated strategy  $\sigma_{-i}$ , where  $\sigma_{-i}(\mathbf{h}_{-i}|\theta_{-i})$  denotes the probability that the remaining agents play strategy  $\mathbf{h}_{-i} \in R_{-i}$  when their joint type realization is some  $\theta_{-i}$ . We can then express agent  $i$ 's best response problem as follows (we use  $G_i$  to denote  $G_i(\theta_i)$ ):

$$\text{br}_i(\theta_i, \sigma_{-i}) = \arg \max_{p_i(\mathbf{h}_i|\theta_i) \in \Delta(G_i)} \int_{\mathbf{h}_i} p_i(\mathbf{h}_i|\theta_i) \sum_{\theta_{-i}} \int_{\mathbf{s}, \alpha, \beta} P(\theta_{-i}, \mathbf{s}, \alpha, \beta|\theta_i) \int_{\mathbf{h}_{-i}} \sigma_{-i}(\mathbf{h}_{-i}|\theta_{-i}) u_i(\mathbf{h}_i; \mathbf{h}_{-i}, \boldsymbol{\lambda})$$

The linearity of the integral and the fact that  $\int_{\mathbf{h}_i} p_i(\mathbf{h}_i|\theta_i) d\mathbf{h}_i = 1$  means that a maximum must exist at a vertex/pure strategy. If multiple pure strategies are optimal, then any linear combination (a mixed strategy) would also be a best-response. However, if there is a unique pure strategy maximizing this, then it means any mixed strategy must be strictly sub-optimal. In other words, it suffices to show that the pure-strategy best-response is unique even when others' play mixed and correlated strategies. This pure best-response problem is given by:

$$\begin{aligned} \text{br}_i(\theta_i, \sigma_{-i}) &= \arg \max_{\mathbf{h}_i \in G_i} \sum_{\theta_{-i}} \int_{\boldsymbol{\lambda}} P(\theta_{-i}, \boldsymbol{\lambda}|\theta_i) \int_{\mathbf{h}_{-i}} \sigma_{-i}(\mathbf{h}_{-i}|\theta_{-i}) u_i(\mathbf{h}_i; \mathbf{h}_{-i}, \boldsymbol{\lambda}) \\ &= \arg \max_{\mathbf{h}_i \in G_i} \sum_{\theta_{-i}} \int_{\boldsymbol{\lambda}} P(\theta_{-i}, \boldsymbol{\lambda}|\theta_i) \int_{\mathbf{h}_{-i}} \sigma_{-i}(\mathbf{h}_{-i}|\theta_{-i}) \left[ f_i(\mathbf{h}_i) - \mathbf{h}_i^T Q \mathbf{h}_i - \sum_{j \neq i} \mathbf{h}_j^T A \mathbf{h}_i - \mathbf{s} B \mathbf{h}_i \right] \\ &= \arg \max_{\mathbf{h}_i \in G_i} \int_{f_i \in F} f_i(\mathbf{h}_i) d\mu(f_i) - \mathbf{h}_i^T Q \mathbf{h}_i \\ &\quad - \underbrace{\left[ \sum_{\theta_{-i}} \int_{\mathbf{s}, \alpha, \beta} P(\theta_{-i}, \mathbf{s}, \alpha, \beta|\theta_i) \int_{\mathbf{h}_{-i}} \sigma_{-i}(\mathbf{h}_{-i}|\theta_{-i}) \sum_{j \neq i} \mathbf{h}_j^T A + \mathbf{s} B \right]}_{\mathbf{w}^T(\cdot)} \mathbf{h}_i \\ &= \arg \max_{\mathbf{h}_i \in G_i} f_i^*(\mathbf{h}_i) - \mathbf{h}_i^T Q \mathbf{h}_i - \mathbf{w}^T(\cdot) \mathbf{h}_i \end{aligned}$$

where  $\mu(f_i)$  is any finite non-negative measure on the function space  $F$ , and  $f^*$  is the result of the integral. The concavity of the function class  $F$  and non-negativity of measure  $\mu$  ensure that  $f^*$  is concave Rockafellar and Wets [23]. Next, we note that  $\mathbf{w}^T(\cdot)$  is a  $T$  dimensional vector that does not depend on the  $\mathbf{h}_i$ . Thus, the objective faced by buyer  $i$  is strictly concave (since  $Q$  is a PD matrix – see Lemma 2) and there is a unique solution. This immediately implies that a mixed strategy will always be a sub-optimal best-response.  $\square$

### C.2 Proof of Theorem 3

*Proof.* As in Theorem 1, we express the results from a minimization perspective. That is, each agent's best-response for type realization  $\theta_i$  is:  $\arg \min_{\mathbf{h}_i \in G_i} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda}|\theta_i} [c_i(\mathbf{h}_i, \mathbf{h}_{-i}, \boldsymbol{\lambda})]$ , where  $c_i(\mathbf{h}_i, \mathbf{h}_{-i}, \boldsymbol{\lambda}) = -u_i(\mathbf{h}_i, \mathbf{h}_{-i}, \boldsymbol{\lambda})$ . Also from 1, we note that the necessary and sufficient conditions for an  $n$  agent

BNE with  $k$  discrete types and pure strategies for each type, can be interpreted as follows:

$$\forall i \in [n], \forall \theta_i \in [k], \forall \mathbf{h}' \in \mathcal{H}_i : \mathbb{E}_{\boldsymbol{\theta}_{-i}, \boldsymbol{\lambda}} [c_i(\mathbf{h}_i^{eq}(\theta_i), \mathbf{h}_{-i}^{eq}(\boldsymbol{\theta}_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}', \mathbf{h}_{-i}^{eq}(\boldsymbol{\theta}_{-i}), \boldsymbol{\lambda})] \leq 0$$

Since expected utility is a smooth function, as in the Nash setting, the simultaneous conditions can be expressed as a variational inequality of the cost derivatives: there can exist no feasible direction at equilibrium at which cost is decreasing. Importantly, this characterization is exact even if the derivatives are scaled by a distinct constant. Formally, a set of strategies are at a BNE if and only if the following holds for any choice of  $\gamma_{i\ell}$  – we will choose  $\gamma_{i\ell} = P(\theta_i)$ , the marginal probability of an agent  $i$  being of type  $\theta_i \in [k]$  – recall  $\mathbf{h}_i(\theta_i) \in G_i(\theta_i)$  is the strategy used upon realization  $\theta_i$ :

$$\forall i \in [n], \forall \theta_i \in [k], \forall \mathbf{h}'(\theta_i) \in G_i : \langle \gamma_{i,\theta_i} \nabla_{\mathbf{h}_i(\theta_i)} \mathbb{E}_{\boldsymbol{\theta}_{-i}, \boldsymbol{\lambda}} [c_i(\mathbf{h}_i^{eq}(\theta_i), \mathbf{h}_{-i}^{eq}(\boldsymbol{\theta}_{-i}), \boldsymbol{\lambda})], (\mathbf{h}'(\theta_i) - \mathbf{h}_i^{eq}(\theta_i)) \rangle \geq 0$$

With our choice of scaling  $\gamma_{i,\ell}$ , and switching the order of gradients and expectation, we have that for any  $i, \theta_i$ ,  $\gamma_{i,\theta_i} \nabla_{\mathbf{h}_i(\theta_i)} \mathbb{E}_{\boldsymbol{\theta}_{-i}, \boldsymbol{\lambda}} [c_i(\mathbf{h}_i^{eq}(\theta_i), \mathbf{h}_{-i}^{eq}(\boldsymbol{\theta}_{-i}), \boldsymbol{\lambda})]$

$$\begin{aligned} &= P(\theta_i) \left( \sum_{\boldsymbol{\theta}_{-i}} \int_{\boldsymbol{\lambda}} Q_{\alpha,\beta} \mathbf{h}_i(\theta_i) P(\boldsymbol{\theta}_{-i}, \boldsymbol{\lambda} | \theta_i) \right. \\ &\quad \left. + \sum_{\boldsymbol{\theta}_{-i}} \int_{\boldsymbol{\lambda}} \left[ \sum_{j \neq i} A_{\alpha,\beta} \mathbf{h}_j(\theta_j) + B_{\alpha,\beta} \mathbf{s} \right] P(\boldsymbol{\theta}_{-i}, \boldsymbol{\lambda} | \theta_i) - \nabla_{\mathbf{h}_i(\theta_i)} \int_{\boldsymbol{\lambda}} f_i(\mathbf{h}_i(\theta_i)) d\mu(f_i | \theta_i) \right) \\ &= P(\theta_i) \left[ \int_{\alpha,\beta} Q_{\alpha,\beta} P(\alpha, \beta | \theta_i) \right] \mathbf{h}_i(\theta_i) + P(\theta_i) \sum_{j \neq i} \sum_{\theta_j} \int_{\alpha,\beta} A_{\alpha,\beta} \mathbf{h}_j(\theta_j) \sum_{\boldsymbol{\theta}_{-(i,j)}} \int_s P(\theta_j, \boldsymbol{\theta}_{-(i,j)}, \boldsymbol{\lambda} | \theta_i) \\ &\quad + \underbrace{\int_{\alpha,\beta,\mathbf{s}} B_{\alpha,\beta} \mathbf{s} P(\boldsymbol{\lambda}, \theta_i) - P(\theta_i) \nabla_{\mathbf{h}_i(\theta_i)} \int_{f_i \in F} f_i(\mathbf{h}_i(\theta_i)) d\mu(f_i | \theta_i)}_{b_{i,\theta_i}} \\ &= P(\theta_i) \left[ \underbrace{\int_{\alpha,\beta} Q_{\alpha,\beta} P(\alpha, \beta | \theta_i)}_{Q_{\alpha,\beta}^*, \theta_i \in \mathbb{R}^{T \times T}} \right] \mathbf{h}_i(\theta_i) + \sum_{j \neq i} \sum_{\theta_j} P(\theta_j, \theta_i) \left[ \underbrace{\int_{\alpha,\beta} A_{\alpha,\beta} P(\alpha, \beta | \theta_j, \theta_i)}_{A_{i,\theta_i,j,\theta_j}^* \in \mathbb{R}^{T \times T}} \right] \mathbf{h}_j(\theta_j) \\ &\quad + b_{i,\theta_i} - P(\theta_i) \cdot \nabla_{\mathbf{h}_i(\theta_i)} f_{i,\theta_i}^*(\mathbf{h}_i(\theta_i)) \end{aligned}$$

where in the last transition, we use the fact that:

$$P(\theta_j, \alpha, \beta | \theta_i) \cdot P(\theta_i) = P(\alpha, \beta, \theta_j, \theta_i) = P(\alpha, \beta | \theta_i, \theta_j) P(\theta_i, \theta_j)$$

We note that  $\mu(f_i | \theta_i)$  is a finite non-negative measure on the function space  $F$ , and  $f_{i,\theta_i}^*$  is the result of the functional integral. The concavity of the function class  $F$  and non-negativity of measure  $\mu$  ensure that  $f_{i,\theta_i}^*$  is a concave function Rockafellar and Wets [23]. Next, let  $k_p = \prod_{i=1}^n k_i$  and define  $\mathbf{x} = [\mathbf{h}_{1,1}, \dots, \mathbf{h}_{1,k_1}, \dots, \mathbf{h}_{n,1}, \dots, \mathbf{h}_{n,k_n}] \in \mathbb{R}^{n k_p}$  denote a complete strategy profile (strategy for each agent for each type). At a high-level, our goal is to show that this operator, denoted by  $F$ , is strictly monotone, which implies the uniqueness of the solution to the equilibrium variational inequality. That is, we want to show that for all  $\mathbf{x}, \langle F(\mathbf{x}) - F(\mathbf{x}'), (\mathbf{x} - \mathbf{x}') \rangle \geq m \|\mathbf{x} - \mathbf{x}'\|^2$ .

We can write this operator as follows:  $F(\mathbf{x}) = M_{\text{bayes}} \mathbf{x} + \mathbf{b} - J(\mathbf{x})$ , where  $\mathbf{b} \in \mathbb{R}^{k_p T}$  has  $b_{i,\theta_i} \in \mathbb{R}^T$  as index  $(i, \theta_i)$ . Observe that this vector is a constant with respect to the agent strategies. Similarly,  $J(\mathbf{x}) \in \mathbb{R}^{k_p T}$ , where at index  $i, \theta_i$ , we have  $P(\theta_i) \nabla_{\mathbf{h}_i(\theta_i)} f_{i,\theta_i}^*(\mathbf{h}_i(\theta_i)) \in \mathbb{R}^T$ . As for the

$M_{\text{bayes}} \in \mathbb{R}^{k_p T \times k_p T}$ , we can write this as an  $n \times n$  block matrix, where each block ( $i \in [n], j \in [n]$ ) is a  $k_i T \times k_j T$  matrix), defined as follows:

$$M_{\text{bayes}}^{ii} = \begin{bmatrix} P(\theta_i = 1)Q_{i,1}^* & 0 & \dots & 0 \\ 0 & P(\theta_i = 2)Q_{i,2}^* & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & P(\theta_i = k)Q_{i,k}^* \end{bmatrix}$$

$$M_{\text{bayes}}^{ij} = \begin{bmatrix} P(\theta_i = 1, \theta_j = 1)A_{i,1,j,1}^* & P(\theta_i = 1, \theta_j = 2)A_{i,1,j,2}^* & \dots & P(\theta_i = 1, \theta_j = k_j)A_{i,1,j,k}^* \\ P(\theta_i = 2, \theta_j = 1)A_{i,2,j,1}^* & P(\theta_i = 2, \theta_j = 2)A_{i,2,j,2}^* & \dots & P(\theta_i = 2, \theta_j = k_j)A_{i,2,j,k}^* \\ \vdots & \vdots & \dots & \vdots \\ P(\theta_i = k_i, \theta_j = 1)A_{i,k,j,1}^* & P(\theta_i = k_i, \theta_j = 2)A_{i,k,j,2}^* & \dots & P(\theta_i = k_i, \theta_j = k_j)A_{i,k,j,k}^* \end{bmatrix}$$

Observe that each  $f_{i,\theta_i}^*$  is a concave function. Since for all convex functions, their gradient is a monotone operator, it is immediate that  $-\langle J(\mathbf{x}) - J(\mathbf{x}') \rangle \geq 0$ . And since  $\mathbf{b}$  is a constant, it suffices to show that the matrix  $M$  is positive definite. That is, we want to show that  $\mathbf{x}^T M \mathbf{x} \geq m \|\mathbf{x}\|^2$ . Observe that:

$$\begin{aligned} \mathbf{x}^T M_{\text{bayes}} \mathbf{x} &= \sum_{i=1}^n \sum_{\theta_i} P(\theta_i) \mathbf{h}_i^T(\theta_i) Q_{i,\theta_i}^* \mathbf{h}_i(\theta_i) + \sum_{i \neq j} \sum_{\theta_i, \theta_j} P_{ij}(\theta_i, \theta_j) \mathbf{h}_i^T(\theta_i) A_{i,\theta_i,j,\theta_j} \mathbf{h}_j(\theta_j) \\ &= \sum_{i=1}^n \sum_{\theta_i} \int_{\alpha, \beta} \mathbf{h}_i^T(\theta_i) Q_{\alpha, \beta} \mathbf{h}_i(\theta_i) \sum_{\boldsymbol{\theta}_{-i}} P(\theta_i, \boldsymbol{\theta}_{-i}, \alpha, \beta) \\ &\quad + \sum_{i \neq j} \sum_{\theta_i, \theta_j} \int_{\alpha, \beta} \mathbf{h}_i^T(\theta_i) A_{\alpha, \beta} \mathbf{h}_j(\theta_j) \sum_{\boldsymbol{\theta}_{-(i,j)}} P(\theta_i, \theta_j, \boldsymbol{\theta}_{-(i,j)}, \alpha, \beta) \\ &= \sum_{\boldsymbol{\theta}} \int_{\alpha, \beta} P(\boldsymbol{\theta}, \alpha, \beta) \left[ \sum_{i=1}^n \mathbf{h}_i^T(\theta_i) Q_{\alpha, \beta} \mathbf{h}_i(\theta_i) + \sum_{j \neq i} \mathbf{h}_i^T(\theta_i) A_{\alpha, \beta} \mathbf{h}_j(\theta_j) \right] \\ &= \mathbb{E}_{\boldsymbol{\theta}, \alpha, \beta} \left[ \sum_{i=1}^n \mathbf{z}_{i, \theta_i}^T Q_{\alpha, \beta} \mathbf{z}_{i, \theta_i} + \sum_{i \neq j} \mathbf{z}_{i, \theta_i}^T A_{\alpha, \beta} \mathbf{z}_{j, \theta_j} \right] = \mathbb{E}_{\boldsymbol{\theta}, \alpha, \beta} [\mathbf{z}_{\boldsymbol{\theta}}^T M_{\alpha, \beta} \mathbf{z}_{\boldsymbol{\theta}}] \end{aligned}$$

where  $\mathbf{z}_{i, \theta_i} = \sum_{\ell} \mathbb{1}[\theta_i = \ell] \mathbf{h}_i(\ell)$  is a random vector of length  $T$  and for a realization  $\boldsymbol{\theta} \in [k]^n$ ,  $\mathbf{z}_{\boldsymbol{\theta}} = [\mathbf{z}_{1, \theta_1}, \dots, \mathbf{z}_{n, \theta_n}]^T$  is a concatenation of these  $n$  random vectors (of size  $nT$ ). Further,  $M_{\alpha, \beta}$  is a random matrix which, for any realization of  $\alpha, \beta$ , is the same as the  $M$  matrix used in the complete information setting. From Theorem 1, we also note that for any  $\alpha, \beta$ , the symmetric component of  $M_{\alpha, \beta}$ , denoted  $M_{\alpha, \beta}^s$  is positive definite; thus, by choosing  $c = \lambda_{\min}(M_{\alpha, \beta}^s)$  ensures the strong monotonicity condition on the operator  $M_{\alpha, \beta}$ . Thus, for any  $\mathbf{z}_{\boldsymbol{\theta}}$  and any realization of  $(\alpha, \beta)$ , there exists a  $c_{\alpha, \beta}$  such that  $\mathbf{z}_{\boldsymbol{\theta}}^T M_{\alpha, \beta} \mathbf{z}_{\boldsymbol{\theta}} \geq c_{\alpha, \beta} \|\mathbf{z}_{\boldsymbol{\theta}}\|^2$ , when  $\mathbf{z}_{\boldsymbol{\theta}} \neq \mathbf{0}$ .

To determine a uniform bound on  $c$  across the randomness of  $(\boldsymbol{\theta}, \alpha, \beta)$ , let each agent's type realization  $\theta_i = \ell$  occur with non-zero probability<sup>4</sup>. Then letting  $P_{\min} = \min_{i, \ell} P(\theta_i = \ell)$  be the smallest probability, and  $c_{\min} = \min_{\alpha, \beta} \lambda_{\min}(M_{\alpha, \beta}^s)$  the smallest eigenvalue possible in the

<sup>4</sup>Note that if a type realization occurs with probability 0, it can be removed from the support without loss of generality.

distribution support of  $\alpha, \beta$ :

$$\begin{aligned} \mathbf{x}^T M_{\text{bayes}} \mathbf{x} &= \mathbb{E}_{\boldsymbol{\theta}, \alpha, \beta} [\mathbf{z}_{\boldsymbol{\theta}}^T M_{\alpha, \beta} \mathbf{z}_{\boldsymbol{\theta}}] = \sum_{\boldsymbol{\theta} | \mathbf{z}_{\boldsymbol{\theta}} \neq \mathbf{0}} \int_{\alpha, \beta} P(\boldsymbol{\theta}, \alpha, \beta) c_{\alpha, \beta} \|\mathbf{z}_{\boldsymbol{\theta}}\|^2 \\ &\geq c_{\min} \sum_{i=1}^n \sum_{\ell=1}^k P(\theta_i = \ell) \|\mathbf{z}_{i, \theta_i}\|^2 = \underbrace{c_{\min} P_{\min}}_m \|\mathbf{x}\|^2 \end{aligned}$$

We recall from Theorem 1 that for any  $c$ -strongly monotone and  $L$ -Lipshietz operator  $F$ , the extra gradient algorithm converges linearly to the unique solution of the variational inequality. We have already shown the operator to be  $c$ -strongly monotone. Further, since the operator is of the form  $M_{\text{bayes}} \mathbf{x} + b - J(\mathbf{x})$ , it suffices to show Lipschitzness of each term. The linear operator  $M_{\text{bayes}}$  is always Lipschitz, with the constant depending on the norm of this matrix. Since each  $f \in F$  is smooth,  $J(\mathbf{x})$  is composed of the gradient of some smooth concave function. Therefore, this is also Lipschitz, with the constant depending on the Hessian of this function.  $\square$

### C.3 Experimental Setup

Our experimental setting for the Bayesian Simulations is as follows. There are 2 agents and 3 possible types for each agent. The type of an agent  $i$ ,  $\theta_i$  is a positive real number that is equal to the constraint. That is, an agent  $i$  of type  $\theta_i$  has a constraint  $-\theta_i \leq \mathbf{1}^T \mathbf{h}_i(\theta_i) \leq \theta_i$ . We have  $\theta_1 \in [10, 15, 20]$  and  $\theta_2 \in [20, 25, 30]$ . The joint type distribution  $P(\theta_1, \theta_2)$  is uniform over the 9 possible outcomes.

Each agent's idiosyncratic utility  $f_i$  is a linear function:  $f_i(\mathbf{h}_i(\theta_i)) = r_i \mathbf{1}^T \mathbf{h}_i(\theta_i)$ . The linearity of this function means it suffices to consider  $\mathbb{E}[r_i | \theta_i]$ . For agent 1, the type conditioned expected reserves are (3, 5, 7), and for agent 2, we set (6, 8, 10).

Lastly, the variational inequality characterizing the BNE has linear dependence on the  $\alpha, \beta$ . As such, it suffices to consider the expected value of these market parameters, conditioned on the joint type realization. We set  $\mathbb{E}[\alpha | \theta_1, \theta_2] = 0.1$  and  $\mathbb{E}[\beta | \theta_1, \theta_2] = \frac{c}{400}(\theta_1 + \theta_2)$ , where we have  $c = 1, 10, 100$  for the left, middle and right panel. These numbers were chosen to ensure the  $\beta$  values were in a comparable range to those used in the experiments for Section 3. This exact setup, with  $c = 10$ , is used for the online learning experiments for Section 5.

## D Proofs and Details for Section 5

In what follows we primarily use the cost notation  $c_i(\cdot)$  (recall that  $c_i(\mathbf{h}_i(\theta_i); \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda}) = -u_i(\mathbf{h}_i(\theta_i); \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})$ ).

### D.1 Proof of Theorem 4

**Proposition 1.** *Let  $P$  be a joint distribution over game instances  $\mathcal{I}$ . Let  $\sigma \in \Delta(\mathcal{H}_1 \times \dots \times \mathcal{H}_n)$  be a distribution over strategy profiles. Suppose  $\sigma$  satisfies for all  $i$ , for all  $\theta_i$ , for all  $\mathbf{h}'_i(\theta_i)$ :*

$$\mathbb{E}_{\mathbf{h} \sim \sigma} \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{\lambda} \sim P} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}'_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \leq \epsilon.$$

Then,  $\sigma$  is an approximate Bayesian coarse correlated equilibrium, satisfying for all  $i$ , for all  $\theta_i$ , for all  $\mathbf{h}'_i(\theta_i)$ :

$$\mathbb{E}_{\mathbf{h} \sim \sigma} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta_i} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}'_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \leq \frac{\epsilon}{\Pr(\theta_i)}.$$

Notice that when  $\sigma$  is a singleton distribution, this corresponds to an approximate Bayesian Nash equilibrium.

*Proof.* We show the contrapositive. Suppose for some agent  $i$ , there is a type  $\theta'_i$  and action  $\mathbf{h}'_i(\theta'_i)$  such that

$$\mathbb{E}_{\mathbf{h} \sim \sigma} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta'_i} [c_i(\mathbf{h}_i(\theta'_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}'_i(\theta'_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] > \frac{\epsilon}{\Pr(\theta'_i)}.$$

For each  $\theta_i$ , define

$$\mathbf{h}_i^*(\theta_i) \in \arg \min_{\mathbf{h}_i} \mathbb{E}_{\mathbf{h} \sim \sigma} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta_i} [c_i(\mathbf{h}_i, \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})].$$

By optimality of  $\mathbf{h}_i^*(\theta'_i)$ ,

$$\mathbb{E}_{\mathbf{h} \sim \sigma} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta'_i} [c_i(\mathbf{h}_i^*(\theta'_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \leq \mathbb{E}_{\mathbf{h} \sim \sigma} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta'_i} [c_i(\mathbf{h}'_i(\theta'_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})].$$

Thus,

$$\mathbb{E}_{\mathbf{h} \sim \sigma} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta'_i} [c_i(\mathbf{h}_i(\theta'_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}_i^*(\theta'_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] > \frac{\epsilon}{\Pr(\theta'_i)}.$$

Now consider the gain by a unilateral deviation to  $\mathbf{h}_i^*(\theta_i)$  for all  $\theta_i$ :

$$\begin{aligned} & \mathbb{E}_{\mathbf{h} \sim \sigma} \mathbb{E}_{\theta, \boldsymbol{\lambda} \sim P} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}_i^*(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \\ &= \sum_{\theta_i} \Pr(\theta_i) \mathbb{E}_{\mathbf{h} \sim \sigma} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta_i} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}_i^*(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})]. \end{aligned}$$

By optimality of  $\mathbf{h}_i^*(\theta_i)$  for every  $\theta_i$ , each summand is non-negative. Furthermore, the summand corresponding to  $\theta'_i$  exceeds  $\Pr(\theta'_i) \cdot \frac{\epsilon}{\Pr(\theta'_i)} = \epsilon$ . Hence the whole sum is  $> \epsilon$ , contradicting the hypothesis.  $\square$

**Lemma 4.** For all  $i \in [n]$ , let  $V_i(\mathbf{h}) = \nabla_{\mathbf{h}_i} \ell_i(\mathbf{h}_i, \mathbf{h}_{-i}; P) = \nabla_{\mathbf{h}_i} \mathbb{E}_{\theta, \boldsymbol{\lambda} \sim P} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \in \mathbb{R}^{k_i \times T}$  and  $V(\mathbf{h}) = (V_1(\mathbf{h}), \dots, V_n(\mathbf{h})) \in \mathbb{R}^{n \times k \times T}$ . The operator  $V$  is  $m$ -strongly monotone, i.e.  $\langle V(\mathbf{h}') - V(\mathbf{h}), \mathbf{h}' - \mathbf{h} \rangle \geq m \|\mathbf{h}' - \mathbf{h}\|^2$  for all  $\mathbf{h}, \mathbf{h}'$ , where  $m$  is the strong monotonicity constant of Theorem 3. Consequently, for all  $i$ ,  $\ell_i(\mathbf{h}_i, \mathbf{h}_{-i}; P)$  is  $m$ -strongly convex in  $\mathbf{h}_i$ .

*Proof.* Recall that Theorem 3 shows that the operator  $W(\mathbf{h}) \in \mathbb{R}^{n \times k \times T}$ , defined by:

$$W_i(\theta_i)(\mathbf{h}) = \Pr(\theta_i) \cdot \nabla_{\mathbf{h}_i(\theta_i)} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta_i} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \in \mathbb{R}^T$$

in the entry corresponding to agent  $i$  and type  $\theta_i$ , is  $m$ -strongly monotone for some positive  $m$ .

Now, for every  $i$ , we can write:

$$V_i(\mathbf{h}) = \nabla_{\mathbf{h}_i} \left( \sum_{\theta_i} \Pr(\theta_i) \cdot \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta_i} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \right)$$

Fix a agent  $i$  and a type  $\theta_i^*$ . Since each agent has finitely many types, we can write  $V$  as a vector of size  $nkT$ , where the entry of  $V$  corresponding to agent  $i$  and type  $\theta_i^*$  is:

$$\begin{aligned}
V_i(\theta_i^*)(\mathbf{h}) &= \nabla_{\mathbf{h}_i(\theta_i^*)} \left( \sum_{\theta_i} \Pr(\theta_i) \cdot \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P|\theta_i} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \right) \\
&= \nabla_{\mathbf{h}_i(\theta_i^*)} \left( \Pr(\theta_i^*) \cdot \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P|\theta_i^*} [c_i(\mathbf{h}_i(\theta_i^*), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \right) \\
&\quad \text{(since all terms not involving } \theta_i^* \text{ can be treated as constants)} \\
&= \Pr(\theta_i^*) \cdot \nabla_{\mathbf{h}_i(\theta_i^*)} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P|\theta_i^*} [c_i(\mathbf{h}_i(\theta_i^*), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \\
&= W_i(\theta_i^*)(\mathbf{h})
\end{aligned}$$

Thus  $V = W$ , and  $V$  is  $m$ -strongly monotone, i.e.  $\langle V(\mathbf{h}') - V(\mathbf{h}), \mathbf{h}' - \mathbf{h} \rangle \geq m \|\mathbf{h}' - \mathbf{h}\|^2$  for all  $\mathbf{h}, \mathbf{h}'$ . For every  $i$ ,  $m$ -strong convexity then follows by definition, by considering  $\mathbf{h}, \mathbf{h}'$  that are the same in all coordinates except  $i$ .  $\square$

*Proof of Theorem 4.* For each  $i$  and  $\mathbf{h}'_i$ , by the regret guarantees of  $\mathbf{Alg}_i$ :

$$\frac{1}{R} \sum_{r=1}^R (\ell_i^r(\mathbf{h}_i^r) - \ell_i^r(\mathbf{h}'_i)) = \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\theta, \boldsymbol{\lambda} \sim P} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}'_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \leq \epsilon_i(R)$$

Applying Proposition 1, we can conclude that for all  $\theta_i$  and all  $\mathbf{h}'_i(\theta_i)$ :

$$\mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P|\theta_i} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}'_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \leq \frac{\epsilon_i(R)}{\Pr(\theta_i)}$$

$\square$

## D.2 Algorithm Details

Here we present the algorithm of Jordan et al. [14] and state its guarantees.

---

**Algorithm 2:** AdaOGD (Algorithm 1 of Jordan et al. [14])

---

**Input:** Strategy space  $\mathcal{H}_i$

Initialize  $\mathbf{h}_i^1 \in \mathcal{H}_i$

Let  $z_0 = \frac{1}{\log(R+10)}$

**for**  $r = 1, \dots, R$  **do**

    Sample  $M^r \sim \text{Geometric}(z_0)$

    Let  $\eta^{r+1} = \frac{r+1}{\sqrt{1+\max\{M^1, \dots, M^r\}}}$

    Update  $\mathbf{h}_i^{r+1} = \arg \min_{\mathbf{h}_i \in \mathcal{H}_i} \{(\mathbf{h}_i - \mathbf{h}_i^r)^\top \tilde{\nabla}_i^r + \frac{\eta^{r+1}}{2} \|\mathbf{h}_i - \mathbf{h}_i^r\|^2\}$

---

**Theorem 6** (Theorem 3.7 of Jordan et al. [14]). *Consider a game  $\mathcal{G}$  among  $n$  agents, each with a convex and bounded action set  $\mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  and a cost function  $\ell_i : \prod_{i=1}^n \mathcal{H}_i \rightarrow \mathbb{R}$  satisfying: (i)  $\ell_i(\mathbf{h}_i, \mathbf{h}_{-i})$  is continuous in  $(\mathbf{h}_i, \mathbf{h}_{-i})$  and continuously differentiable in  $\mathbf{h}_i$ ; (ii)  $\nabla_{\mathbf{h}_i} \ell_i(\mathbf{h}_i, \mathbf{h}_{-i})$  is continuous in  $(\mathbf{h}_i, \mathbf{h}_{-i})$ ; (iii)  $\|\mathbf{h} - \mathbf{h}'\| \leq D$  for all  $\mathbf{h}, \mathbf{h}' \in \prod_{i=1}^n \mathcal{H}_i$ ; and (iv)  $\mathcal{G}$  is  $m$ -strongly monotone. Suppose at every round  $r \in [R]$ , each agent observes an unbiased and bounded gradient  $\tilde{\nabla}_{\mathbf{h}_i^r}^r$*

satisfying  $\mathbb{E}[\tilde{\nabla}_{\mathbf{h}_i^r}^r | \mathbf{h}^r] = \nabla_{\mathbf{h}_i^r} \ell_i(\mathbf{h}_i^r, \mathbf{h}_{-i}^r)$  and  $\mathbb{E}[\|\tilde{\nabla}_{\mathbf{h}_i^r}^r\|^2 | \mathbf{h}^r] \leq M$ . Then, if all agents run Algorithm 2, the final iterate satisfies:

$$\mathbb{E}[\|\mathbf{h}^R - \mathbf{h}^*\|^2] \leq O\left(\frac{D^2 M (1 + \exp(1/(m^2 \log R))) \log(nR) \log^2(R)}{R}\right)$$

where  $\mathbf{h}^*$  is the Nash equilibrium of  $\mathcal{G}$ , i.e. for all  $i \in [n]$ , for all  $\mathbf{h}_i \in \mathcal{H}_i$ ,  $\ell_i(\mathbf{h}_i^*, \mathbf{h}_{-i}^*) \leq \ell_i(\mathbf{h}_i, \mathbf{h}_{-i}^*)$ .

### D.3 Proof of Theorem 5

*Proof.* We define the game  $\mathcal{G}$  where each agent  $i$  chooses a strategy map  $\mathbf{h}_i \in \mathcal{H}_i$  and suffers cost:

$$\ell_i(\mathbf{h}_i, \mathbf{h}_{-i}; P) = \mathbb{E}_{\theta, \boldsymbol{\lambda} \sim P}[c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})]$$

We verify the conditions of Theorem 6 on this game  $\mathcal{G}$ . Since for all type profiles  $\theta$ ,  $c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})$  is continuous in  $(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}))$  and continuously differentiable in  $\mathbf{h}_i(\theta_i)$ , we have that  $\ell_i(\mathbf{h}_i, \mathbf{h}_{-i}; P)$  is continuous in  $(\mathbf{h}_i, \mathbf{h}_{-i})$  and continuously differentiable in  $\mathbf{h}_i$ . Under Assumption 3,  $\|\mathbf{h} - \mathbf{h}'\| \leq B\sqrt{nk}$  for all  $\mathbf{h}, \mathbf{h}'$ . Furthermore, by Lemma 4,  $\mathcal{G}$  is  $m$ -strongly monotone, where  $m$  is the strong monotonicity of Theorem 3.

To apply Theorem 6, it remains to establish that agents observe unbiased and bounded gradient feedback. Recall at each round  $r \in [R]$ , agent  $i$  receives as feedback:  $\tilde{\nabla}_{\mathbf{h}_i}^r = \nabla_{\mathbf{h}_i} c_i(\mathbf{h}_i^r(\theta_i^r), \mathbf{h}_{-i}^r(\theta_{-i}^r), \boldsymbol{\lambda}^r)$ . Since  $\theta^r, \boldsymbol{\lambda}^r \sim P$  are sampled independently from the strategies chosen at round  $r$ , we have that  $\mathbb{E}[\tilde{\nabla}_{\mathbf{h}_i}^r | \mathbf{h}^r] = \mathbb{E}[\tilde{\nabla}_{\mathbf{h}_i}^r] = \mathbb{E}[\nabla_{\mathbf{h}_i} c_i(\mathbf{h}_i^r(\theta_i^r), \mathbf{h}_{-i}^r(\theta_{-i}^r), \boldsymbol{\lambda}^r)] = \nabla_{\mathbf{h}_i} \ell_i(\mathbf{h}_i^r, \mathbf{h}_{-i}^r; P)$ , i.e. the gradient is unbiased. Moreover, we can compute, for any  $\mathbf{h}_i, \mathbf{h}_{-i}, \theta, \boldsymbol{\lambda}$ :

$$\begin{aligned} & \|\nabla_{\mathbf{h}_i} c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})\| \\ &= \|p_0 \cdot \mathbf{1}_T + \alpha J \mathbf{h}_i(\theta_i) + \alpha M \left( \sum_{j \neq i} \mathbf{h}_j(\theta_j) - \mathbf{s} \right) + \beta \left( 2\mathbf{h}_i(\theta_i) + \sum_{j \neq i} \mathbf{h}_j(\theta_j) - \mathbf{s} \right) - \nabla_{\mathbf{h}_i} f_i(\mathbf{h}_i(\theta_i))\| \\ &\leq |p_0| \sqrt{T} + \alpha T B + \alpha T ((n-1)B + S) + \beta ((n+1)B + S) + U' \quad (\text{by Assumption 3}) \end{aligned}$$

where  $J \in \mathbb{R}^{T \times T}$  is the matrix with  $J_{tt} = 2$  for all  $t \in [T]$  and 1 everywhere else, and  $M \in \mathbb{R}^{T \times T}$  is the matrix with  $M_{ts} = 1$  for  $s \leq t$  and 0 everywhere else. Hence,

$$\begin{aligned} \mathbb{E}[\|\tilde{\nabla}_{\mathbf{h}_i}^r\|^2 | \mathbf{h}^r] &= \mathbb{E}[\|\tilde{\nabla}_{\mathbf{h}_i}^r\|^2] \\ &= \mathbb{E}[\|\nabla_{\mathbf{h}_i} c_i(\mathbf{h}_i^r(\theta_i^r), \mathbf{h}_{-i}^r(\theta_{-i}^r), \boldsymbol{\lambda}^r)\|^2] \\ &\leq \left( p_{0_{\max}} \sqrt{T} + \alpha_{\max} T B + \alpha_{\max} T ((n-1)B + S) + \beta_{\max} ((n+1)B + S) + U' \right)^2 \\ &= \text{poly}(n, T, \alpha, \beta, p_0, B, S, U') \end{aligned}$$

Above,  $\alpha_{\max} = \arg \max_{\alpha \in \text{supp}(P)} \{\alpha\}$ ,  $\beta_{\max} = \arg \max_{\beta \in \text{supp}(P)} \{\beta\}$ , and  $p_{0_{\max}} = \arg \max_{p_0 \in \text{supp}(P)} \{|p_0|\}$

Therefore, by Theorem 6, the final iterate produced by Algorithm 2 satisfies:

$$\mathbb{E}[\|\mathbf{h}^R - \mathbf{h}^*\|^2] \leq O\left(\frac{D^2 M (1 + \exp(1/m^2 \log R)) \log(nR) \log^2(R)}{R}\right)$$

where  $\mathbf{h}^*$  is the Nash equilibrium of  $\mathcal{G}$ , and the expectation is taken over the randomness of the algorithm. Here, we have  $D = \text{poly}(n, k, B)$  and  $M = \text{poly}(n, T, \alpha, \beta, p_0, B, S, U')$ .

Next we show that  $\ell_i$  is Lipschitz in the  $\ell_2$  norm, which will allow us to argue that since  $\mathbf{h}^R$  and  $\mathbf{h}^*$  are close in  $\ell_2$  distance, they must also incur similar cost. Observe that by Assumption 3, for any  $j \neq i$ , for any  $\mathbf{h}_i, \mathbf{h}_{-i}, \theta, \boldsymbol{\lambda}$ :

$$\|\nabla_{\mathbf{h}_j} c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})\| = \|\alpha M^\top \mathbf{h}_i(\theta_i) + \beta \mathbf{h}_i(\theta_i)\| \leq (\alpha T + \beta)B$$

and so:

$$\begin{aligned} & \sup_{\mathbf{h}} \|\nabla_{\mathbf{h}} c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})\| \\ & \leq \underbrace{|p_0|\sqrt{T} + \alpha TB + \alpha T((n-1)B + S) + \beta((n+1)B + S) + U' + n(\alpha T + \beta)B}_{=: L'(p_0, \alpha, \beta)} \end{aligned}$$

Therefore for all  $\mathbf{h}, \mathbf{h}', \theta, \boldsymbol{\lambda}$ ,  $c_i$  is  $L'(p_0, \alpha, \beta)$ -Lipschitz in  $\mathbf{h}$ , i.e.:

$$|c_i(\mathbf{h}'_i(\theta_i), \mathbf{h}'_{-i}(\theta_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})| \leq L' \|\mathbf{h}' - \mathbf{h}\|$$

Taking expectations, we have that  $\ell_i$  is  $L$ -Lipschitz in  $\mathbf{h}$ , i.e. for all  $\mathbf{h}, \mathbf{h}'$ :

$$\begin{aligned} |\ell_i(\mathbf{h}'_i, \mathbf{h}'_{-i}; P) - \ell_i(\mathbf{h}_i, \mathbf{h}_{-i}; P)| &= |\mathbb{E}_{\theta, \boldsymbol{\lambda} \sim P}[c_i(\mathbf{h}'_i(\theta_i), \mathbf{h}'_{-i}(\theta_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})]| \\ &\leq \mathbb{E}_{\theta, \boldsymbol{\lambda} \sim P}[|c_i(\mathbf{h}'_i(\theta_i), \mathbf{h}'_{-i}(\theta_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})|] \\ &\leq L \|\mathbf{h}' - \mathbf{h}\| \end{aligned}$$

where  $L \leq \max_{p_0, \alpha, \beta} L'(p_0, \alpha, \beta) = \text{poly}(n, T, \alpha, \beta, p_0, B, S, U')$ .

Thus, the cost evaluated at  $\mathbf{h}^R$  is close to the cost evaluated at  $\mathbf{h}^*$ :

$$\begin{aligned} \mathbb{E} [\ell_i(\mathbf{h}_i^*, \mathbf{h}_{-i}^*; P) - \ell_i(\mathbf{h}_i^R, \mathbf{h}_{-i}^R; P)] &\leq L \cdot \mathbb{E} [\|\mathbf{h}^R - \mathbf{h}^*\|] \quad (\text{by } L\text{-Lipschitzness}) \\ &\leq L \cdot O \left( \sqrt{\frac{D^2 M (1 + \exp(1/m^2 \log R)) \log(nR) \log^2(R)}{R}} \right) \end{aligned}$$

In the second inequality, we use that fact that  $\mathbb{E} [\|\mathbf{h}^R - \mathbf{h}^*\|^2] \leq \mathbb{E} [\|\mathbf{h}^R - \mathbf{h}^*\|^2]$  by Jensen's inequality. Furthermore, since the entries of  $\|\mathbf{h}^R - \mathbf{h}^*\|$  are non-negative, we also have that for any  $\mathbf{h}_i \in \mathcal{H}_i$ :

$$\begin{aligned} \mathbb{E} [\ell_i(\mathbf{h}_i, \mathbf{h}_{-i}^R; P) - \ell_i(\mathbf{h}_i, \mathbf{h}_{-i}^*; P)] &\leq L \cdot \mathbb{E} [\|\mathbf{h}_{-i}^R - \mathbf{h}_{-i}^*\|] \quad (\text{by } L\text{-Lipschitzness}) \\ &\leq L \cdot \mathbb{E} [\|\mathbf{h}^R - \mathbf{h}^*\|] \\ &\leq L \cdot O \left( \sqrt{\frac{D^2 M (1 + \exp(1/m^2 \log R)) \log(nR) \log^2(R)}{R}} \right) \end{aligned}$$

Combining the above, we can show that  $\mathbf{h}^R$  is an approximate Nash equilibrium of  $\mathcal{G}$ . In



particular, for any  $i$ , for any  $\mathbf{h}_i \in \mathcal{H}_i$ :

$$\begin{aligned}
& \mathbb{E} [\ell_i(\mathbf{h}_i^R, \mathbf{h}_{-i}^R; P) - \ell_i(\mathbf{h}_i, \mathbf{h}_{-i}^R; P)] \\
& \leq \mathbb{E} [\ell_i(\mathbf{h}_i^*, \mathbf{h}_{-i}^*; P) - \ell_i(\mathbf{h}_i, \mathbf{h}_{-i}^R; P)] + L \cdot O \left( \sqrt{\frac{D^2 M (1 + \exp(1/m^2 \log R)) \log(nR) \log^2(R)}{R}} \right) \\
& \leq \mathbb{E} [\ell_i(\mathbf{h}_i^*, \mathbf{h}_{-i}^*; P) - \ell_i(\mathbf{h}_i, \mathbf{h}_{-i}^*; P)] + 2L \cdot O \left( \sqrt{\frac{D^2 M (1 + \exp(1/m^2 \log R)) \log(nR) \log^2(R)}{R}} \right) \\
& \leq 2L \cdot O \left( \sqrt{\frac{D^2 M (1 + \exp(1/m^2 \log R)) \log(nR) \log^2(R)}{R}} \right) \\
& \hspace{15em} \text{(by the fact that } \mathbf{h}^* \text{ is a Nash equilibrium)}
\end{aligned}$$

Thus, applying Proposition 1, we can conclude that  $\mathbf{h}^R$  is an approximate Bayesian Nash equilibrium. Specifically, for all  $i$ , for all  $\theta_i$ , and for all  $\mathbf{h}_i(\theta_i)$ :

$$\begin{aligned}
& \mathbb{E} [\mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P|\theta_i} [c_i(\mathbf{h}_i^R(\theta_i), \mathbf{h}_{-i}^R(\theta_{-i}), \boldsymbol{\lambda}) - c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}^R(\theta_{-i}), \boldsymbol{\lambda})]] \\
& \leq \frac{2L}{\Pr(\theta_i)} \cdot O \left( \sqrt{\frac{D^2 M (1 + \exp(1/m^2 \log R)) \log(nR) \log^2(R)}{R}} \right) \quad \text{(by Proposition 1)} \\
& \leq O \left( \frac{\text{poly}(n, T, k, \alpha, \beta, p_0, B, S, U')}{\Pr(\theta_i)} \cdot \frac{\log^{3/2}(R)}{\sqrt{R}} \right)
\end{aligned}$$

as desired.  $\square$

## D.4 Experimental Details

The online learning experimental setup follows that of Section 4. Here we provide more details on the conditional distributions of agent types and market parameters used. Recall that  $\theta_1 \in [10, 15, 20]$  and  $\theta_2 \in [20, 25, 30]$ , and the joint type distribution  $P(\theta_1, \theta_2)$  is uniform over the 9 possible type profiles. Agent 1's linear utility coefficient  $r_1$  is drawn from a conditional Gaussian distribution:  $r_1|\theta_1 \sim \mathcal{N}(\mu(\theta_1), 1)$  with  $\mu(10) = 3, \mu(15) = 5$ , and  $\mu(20) = 7$ . Similarly, Agent 2's linear utility coefficient  $r_2$  is drawn from a conditional Gaussian distribution:  $r_2|\theta_2 \sim \mathcal{N}(\mu(\theta_2), 1)$  with  $\mu(20) = 6, \mu(25) = 8$ , and  $\mu(30) = 10$ . Thus the type conditioned expected reserves are  $(3, 5, 7)$  for agent 1 and  $(6, 8, 10)$  for agent 2. Finally, we fix  $p_0 = 0$ ,  $\alpha = 0.1$ ,  $\beta = \frac{1}{40}(\theta_1 + \theta_2)$ , and draw  $\mathbf{s}$  from the Gaussian distribution: for all  $t \in [T]$ ,  $s_t \sim \mathcal{N}(0, 1)$ .

## E Learning in the Bayesian Game Without Stochastic Feedback

Here we relax the assumption that agents have access to online algorithms with no-regret guarantees under *stochastic gradient* feedback. Instead, we assume access to no-regret algorithms that, given cost function feedback, can learn over an agent's strategy space conditional on any type. Recall that  $G_i(\theta_i)$  is the set of feasible strategies  $\mathbf{h}_i(\theta_i)$  for agent  $i$  and type  $\theta_i$ . Given a sequence of costs

$c_i^r$ , an algorithm  $\mathbf{Alg}$  achieves average regret bounded by  $\epsilon(R)$  if:

$$\frac{1}{R} \sum_{r=1}^R c_i^r(\mathbf{h}_i^r(\theta_i)) - \min_{\mathbf{h}_i(\theta_i) \in G_i(\theta_i)} \frac{1}{R} \sum_{r=1}^R c_i^r(\mathbf{h}_i(\theta_i)) \leq \epsilon(R)$$

Mirroring Theorem 4, we show how agents can converge to Bayesian coarse correlated equilibrium by running separate instances of such no-regret algorithms, one for each type. The procedure is described in Algorithm 3 and mirrors the setup used by Hartline et al. [13].

---

**Algorithm 3:** No-Regret Learning Protocol Without Stochastic Feedback

---

**Input:** No-regret algorithms  $\mathbf{Alg}_i$

**Output:** Joint distribution of strategy profiles  $\sigma \in \Delta(\mathcal{H}_1 \times \dots \times \mathcal{H}_n)$

Each agent  $i$  initializes an instance of  $\mathbf{Alg}_i$  over the action space  $G_i(\theta_i)$  for every  $\theta_i \in \Theta_i$ .

We denote the instance corresponding to  $\theta_i$  by  $\mathbf{Alg}_i(\theta_i)$ .

**for**  $r = 1, \dots, R$  **do**

**for**  $i = 1, \dots, n$  **do**

        For each  $\theta_i$ , let  $\mathbf{h}_i^r(\theta_i) \in G_i(\theta_i)$  be the output of  $\mathbf{Alg}_i(\theta_i)$ .

        Observe  $\theta_i^r$  and take action  $\mathbf{h}_i^r(\theta_i^r)$ .

        Receive cost function  $c_i^r$  and update  $\mathbf{Alg}_i(\theta_i^r)$  with  $c_i^r$ . Update all other  $\mathbf{Alg}_i(\theta_i)$ ,  $\theta_i \neq \theta_i^r$ , with  $c_i^r = 0$ .

Output empirical distribution over strategy profiles  $\{\mathbf{h}^1, \dots, \mathbf{h}^R\}$ .

---

**Theorem 7.** Fix a joint distribution  $P$  over types  $\theta$  and instances  $\lambda$ . Fix  $\delta \in (0, 1)$ . For every  $i \in [n]$ , suppose there is an algorithm  $\mathbf{Alg}_i$  that, given any  $G_i(\theta_i)$ , and against any sequence  $c_i^1, \dots, c_i^R$ , obtains average regret bounded by  $\epsilon_i(R)$  after  $R$  rounds. Let  $\sigma^R$  be the output of the Bayesian no-regret learning protocol (Algorithm 3), when given as input algorithms  $\mathbf{Alg}_i$ . Then, for every agent  $i$ , for every type  $\theta_i \in \Theta_i$ , and every  $\mathbf{h}_i'(\theta_i) \in G_i(\theta_i)$ , with probability at least  $1 - \delta$ :

$$\mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\theta_{-i}, \lambda \sim P | \theta_i} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \lambda) - c_i(\mathbf{h}_i'(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \lambda)] \leq \frac{\epsilon_i(R) + 2H \sqrt{\frac{2 \ln \frac{2}{\delta}}{R}}}{\Pr(\theta_i)}$$

Here,  $H = B p_{0_{\max}} \sqrt{T} + \alpha_{\max} B T(nB + S) + \beta_{\max} B(nB + S) + U$ , where  $\alpha_{\max} = \arg \max_{\alpha \in \text{supp}(P)} \{\alpha\}$ ,  $\beta_{\max} = \arg \max_{\beta \in \text{supp}(P)} \{\beta\}$ , and  $p_{0_{\max}} = \arg \max_{p_0 \in \text{supp}(P)} \{p_0\}$ .

First, we show in the following lemma a concentration bound: on any sequence a type  $\theta_i^*$  was observed, the agent  $i$ 's cost under the empirically observed types and game instances concentrate around their expected cost.

**Lemma 5.** Fix a agent  $i$  and a type  $\theta_i^* \in \Theta_i$ . Fix  $\delta \in (0, 1)$ . Suppose costs are bounded between  $[-H, H]$  uniformly over all types and strategies. Let  $\mathbf{h}^1, \dots, \mathbf{h}^R$  be any sequence of strategy profiles, and let  $\sigma^R$  denote the empirical distribution over  $\mathbf{h}^1, \dots, \mathbf{h}^R$ . Then, with probability at least  $1 - \delta$ :

$$\left| \frac{1}{R} \sum_{r=1}^R \mathbf{1}[\theta_i^r = \theta_i^*] c_i(\mathbf{h}_i^r(\theta_i^r), \mathbf{h}_{-i}^r(\theta_{-i}^r), \lambda) - \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\theta, \lambda \sim P} [\mathbf{1}[\theta_i = \theta_i^*] c_i(\mathbf{h}_i(\theta_i^*), \mathbf{h}_{-i}(\theta_{-i}), \lambda)] \right| \leq H \sqrt{\frac{2 \ln \frac{1}{\delta}}{R}}.$$

*Proof.* For convenience, let  $Y_i(\mathbf{h}, \theta_{-i}, \boldsymbol{\lambda}) = c_i(\mathbf{h}_i(\theta_i^*), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})$  and  $I_i^r = \mathbf{1}[\theta_i^r = \theta_i^*]$ . Hence we can write:

$$\frac{1}{R} \sum_{r=1}^R \mathbf{1}[\theta_i^r = \theta_i^*] c_i(\mathbf{h}_i^r(\theta_i^r), \mathbf{h}_{-i}^r(\theta_{-i}^r), \boldsymbol{\lambda}) = \frac{1}{R} \sum_{r=1}^R I_i^r Y_i(\mathbf{h}^r, \theta_{-i}^r, \boldsymbol{\lambda}^r).$$

and, letting  $p(\theta_i^*) = \Pr_P[\theta_i^*]$ :

$$\begin{aligned} \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{\lambda} \sim P} [\mathbf{1}[\theta_i = \theta_i^*] c_i(\mathbf{h}_i(\theta_i^*), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] &= \frac{1}{R} \sum_{r=1}^R \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{\lambda} \sim P} [\mathbf{1}[\theta_i = \theta_i^*] c_i(\mathbf{h}_i^r(\theta_i^*), \mathbf{h}_{-i}^r(\theta_{-i}), \boldsymbol{\lambda})] \\ &= \frac{1}{R} \sum_{r=1}^R p(\theta_i^*) \mathbb{E}_{\boldsymbol{\theta}_{-i}, \boldsymbol{\lambda} \sim P | \theta_i^*} [Y_i(\mathbf{h}^r, \theta_{-i}, \boldsymbol{\lambda})]. \end{aligned}$$

Thus we want to bound the quantity  $\left| \frac{1}{R} \sum_{r=1}^R I_i^r Y_i(\mathbf{h}^r, \theta_{-i}^r, \boldsymbol{\lambda}^r) - \frac{1}{R} \sum_{r=1}^R p(\theta_i^*) \mathbb{E}_{\boldsymbol{\theta}_{-i}, \boldsymbol{\lambda} \sim P | \theta_i^*} [Y_i(\mathbf{h}^r, \theta_{-i}, \boldsymbol{\lambda})] \right|$ .

Let  $F_{\leq r}$  denote the sequence  $\{I_i^s Y_i(\mathbf{h}^s, \theta_{-i}^s, \boldsymbol{\lambda}^s)\}_{s \leq r}$ . Since types are drawn independently each round,  $I_i^r$  is independent of  $F_{\leq r-1}$ . Moreover,  $\mathbf{h}^r$  is chosen prior to the draw of  $\boldsymbol{\theta}^r, \boldsymbol{\lambda}^r$ , so  $\boldsymbol{\theta}^r, \boldsymbol{\lambda}^r$  are independent of  $\mathbf{h}^r$ . Thus:

$$\begin{aligned} \mathbb{E}[I_i^r Y_i(\mathbf{h}^r, \theta_{-i}^r, \boldsymbol{\lambda}^r) | F_{\leq r-1}] &= \mathbb{E}[\mathbb{E}[I_i^r Y_i(\mathbf{h}^r, \theta_{-i}^r, \boldsymbol{\lambda}^r) | F_{\leq r-1}, \theta_i^r] | F_{\leq r-1}] \\ &= \mathbb{E}[\mathbf{1}[\theta_i^r = \theta_i^*] \mathbb{E}[Y_i(\mathbf{h}^r, \theta_{-i}^r, \boldsymbol{\lambda}^r) | F_{\leq r-1}, \theta_i^r] | F_{\leq r-1}] \\ &= \mathbb{E}[\mathbf{1}[\theta_i^r = \theta_i^*] \mathbb{E}_{\mathbf{h}_{-i}, \boldsymbol{\lambda} \sim P | \theta_i^r} [Y_i(\mathbf{h}^r, \theta_{-i}, \boldsymbol{\lambda}) | F_{\leq r-1}] | F_{\leq r-1}] \\ &= \mathbb{E}[\mathbf{1}[\theta_i^r = \theta_i^*] \mathbb{E}_{\mathbf{h}_{-i}, \boldsymbol{\lambda} \sim P | \theta_i^r} [Y_i(\mathbf{h}^r, \theta_{-i}, \boldsymbol{\lambda})] | F_{\leq r-1}] \\ &= \mathbb{E}_{\mathbf{h}_{-i}, \boldsymbol{\lambda} \sim P | \theta_i^*} [Y_i(\mathbf{h}^r, \theta_{-i}, \boldsymbol{\lambda})] \cdot \Pr[\mathbf{1}[\theta_i^r = \theta_i^*] | F_{\leq r-1}] \\ &= \mathbb{E}_{\mathbf{h}_{-i}, \boldsymbol{\lambda} \sim P | \theta_i^*} [Y_i(\mathbf{h}^r, \theta_{-i}, \boldsymbol{\lambda})] \cdot \Pr[\mathbf{1}[\theta_i^r = \theta_i^*]] \\ &= p(\theta_i^*) \mathbb{E}_{\mathbf{h}_{-i}, \boldsymbol{\lambda} \sim P | \theta_i^*} [Y_i(\mathbf{h}^r, \theta_{-i}, \boldsymbol{\lambda})]. \end{aligned}$$

By Azuma's inequality:

$$\Pr \left[ \left| \frac{1}{R} \sum_{r=1}^R I_i^r Y_i(\mathbf{h}^r, \theta_{-i}^r, \boldsymbol{\lambda}^r) - \frac{1}{R} \sum_{r=1}^R \mathbb{E}[I_i^r Y_i(\mathbf{h}^r, \theta_{-i}^r, \boldsymbol{\lambda}^r) | F_{\leq r-1}] \right| \geq m \right] \leq 2 \exp \left( \frac{-m^2 R}{2H^2} \right).$$

Plugging in  $m \geq H \sqrt{\frac{2 \ln \frac{1}{\delta}}{R}}$ , we have that with probability at least  $1 - \delta$ :

$$\begin{aligned} &\left| \frac{1}{R} \sum_{r=1}^R I_i^r Y_i(\mathbf{h}^r, \theta_{-i}^r, \boldsymbol{\lambda}^r) - \frac{1}{R} \sum_{r=1}^R \mathbb{E}[I_i^r Y_i(\mathbf{h}^r, \theta_{-i}^r, \boldsymbol{\lambda}^r) | F_{\leq r-1}] \right| \\ &= \left| \frac{1}{R} \sum_{r=1}^R I_i^r Y_i(\mathbf{h}^r, \theta_{-i}^r, \boldsymbol{\lambda}^r) - \frac{1}{R} \sum_{r=1}^R p(\theta_i^*) \mathbb{E}_{\boldsymbol{\theta}_{-i}, \boldsymbol{\lambda} \sim P | \theta_i^*} [Y_i(\mathbf{h}^r, \theta_{-i}, \boldsymbol{\lambda})] \right| \\ &= \left| \frac{1}{R} \sum_{r=1}^R \mathbf{1}[\theta_i^r = \theta_i^*] c_i(\mathbf{h}_i^r(\theta_i^r), \mathbf{h}_{-i}^r(\theta_{-i}^r), \boldsymbol{\lambda}) - \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{\lambda} \sim P} [\mathbf{1}[\theta_i = \theta_i^*] c_i(\mathbf{h}_i(\theta_i^*), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \right| \\ &\leq H \sqrt{\frac{2 \ln \frac{1}{\delta}}{R}}, \end{aligned}$$

as claimed.  $\square$

Now we prove the theorem.

*Proof.* Let  $\sigma^R$  be the empirical distribution over  $\mathbf{h}^1, \dots, \mathbf{h}^R$ , the history of strategy profiles output by the learning protocol. Consider a agent  $i$ , and fix a type  $\theta_i^* \in \Theta_i$  and any action  $\mathbf{h}'_i(\theta_i) \in G_i(\theta_i)$ . By the regret guarantee of  $\mathbf{Alg}_i(\theta_i^*)$ , we have that:

$$\frac{1}{R} \sum_{r=1}^R c_i(\mathbf{h}_i^r(\theta_i^*), \mathbf{h}_{-i}^r(\theta_{-i}^*), \boldsymbol{\lambda}) - \frac{1}{R} \sum_{r=1}^R c_i(\mathbf{h}'_i(\theta_i), \mathbf{h}_{-i}^r(\theta_{-i}^*), \boldsymbol{\lambda}) \leq \epsilon_i(R).$$

By construction, on the rounds where  $\theta_i^r \neq \theta_i^*$ ,  $c_i^r = 0$  for all actions in  $G_i(\theta_i)$ , and so we equivalently have:

$$\underbrace{\frac{1}{R} \sum_{r=1}^R \mathbf{1}[\theta_i^r = \theta_i^*] c_i(\mathbf{h}_i^r(\theta_i^*), \mathbf{h}_{-i}^r(\theta_{-i}^*), \boldsymbol{\lambda})}_{(1)} - \underbrace{\frac{1}{R} \sum_{r=1}^R \mathbf{1}[\theta_i^r = \theta_i^*] c_i(\mathbf{h}'_i(\theta_i), \mathbf{h}_{-i}^r(\theta_{-i}^*), \boldsymbol{\lambda})}_{(2)} \leq \epsilon_i(R).$$

Before analyzing this expression, we bound the magnitude of costs. For any  $\mathbf{h}_i, \mathbf{h}_{-i}, \theta, \boldsymbol{\lambda}$ , we have, by applying Cauchy-Schwarz and Assumption 3:

$$\begin{aligned} & |c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})| \\ & \leq |p_0| \|\mathbf{h}_i(\theta_i)\|_1 + \alpha \|M \mathbf{h}_i(\theta_i)\| \left\| \sum_{j=1}^n \mathbf{h}_j(\theta_j) - \mathbf{s} \right\| + \beta \|\mathbf{h}_i(\theta_i)\| \left\| \sum_{j=1}^n \mathbf{h}_j(\theta_j) - \mathbf{s} \right\| - f_i(\mathbf{h}_i(\theta_i)) \\ & \leq B|p_0|\sqrt{T} + \alpha BT(nB + S) + \beta B(nB + S) + U, \end{aligned}$$

where  $M$  is the lower triangular matrix. We set  $H = Bp_{0_{\max}}\sqrt{T} + \alpha_{\max}BT(nB + S) + \beta_{\max}B(nB + S) + U$ .

Then, to analyze term (1): using Lemma 5 and the fact that:

$$\mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{\lambda} \sim P} [\mathbf{1}[\theta_i = \theta_i^*] c_i(\mathbf{h}_i(\theta_i^*), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] = p(\theta_i^*) \cdot \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{\lambda} \sim P | \theta_i^*} [c_i(\mathbf{h}_i(\theta_i^*), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})],$$

we have that with probability at least  $1 - \frac{\delta}{2}$ :

$$\begin{aligned} & \left| \frac{1}{R} \sum_{r=1}^R \mathbf{1}[\theta_i^r = \theta_i^*] c_i(\mathbf{h}_i^r(\theta_i^*), \mathbf{h}_{-i}^r(\theta_{-i}^*), \boldsymbol{\lambda}) - \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{\lambda} \sim P} [\mathbf{1}[\theta_i = \theta_i^*] c_i(\mathbf{h}_i(\theta_i^*), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \right| \\ & = \left| \frac{1}{R} \sum_{r=1}^R \mathbf{1}[\theta_i^r = \theta_i^*] c_i(\mathbf{h}_i^r(\theta_i^*), \mathbf{h}_{-i}^r(\theta_{-i}^*), \boldsymbol{\lambda}) - p(\theta_i^*) \cdot \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{\lambda} \sim P | \theta_i^*} [c_i(\mathbf{h}_i(\theta_i^*), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \right| \\ & \leq H \sqrt{\frac{2 \ln \frac{2}{\delta}}{R}}. \end{aligned}$$

Similarly, for term (2): we apply Lemma 5 on the sequence where for all  $r \in [R]$ ,  $\mathbf{h}_i^r(\theta_i) = \mathbf{h}'_i(\theta_i)$

for all  $\theta_i \in \Theta_i$ , and  $\mathbf{h}_{-i}^r$  remains unchanged. We have that with probability at least  $1 - \frac{\delta}{2}$ :

$$\begin{aligned}
& \left| \frac{1}{R} \sum_{r=1}^R \mathbf{1}[\theta_i^r = \theta_i^*] c_i(\mathbf{h}_i'(\theta_i), \mathbf{h}_{-i}^r(\theta_{-i}^r), \boldsymbol{\lambda}) - \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{\lambda} \sim P} [\mathbf{1}[\theta_i = \theta_i^*] c_i(\mathbf{h}_i'(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \right| \\
&= \left| \frac{1}{R} \sum_{r=1}^R \mathbf{1}[\theta_i^r = \theta_i^*] c_i(\mathbf{h}_i'(\theta_i), \mathbf{h}_{-i}^r(\theta_{-i}^r), \boldsymbol{\lambda}) - p(\theta_i^*) \cdot \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{\lambda} \sim P | \theta_i^*} [c_i(\mathbf{h}_i'(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \right| \\
&\leq H \sqrt{\frac{2 \ln \frac{2}{\delta}}{R}}.
\end{aligned}$$

Thus, we can conclude, with probability at least  $1 - \delta$ :

$$\begin{aligned}
& p(\theta_i^*) \cdot \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta_i^*} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] - p(\theta_i^*) \cdot \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta_i^*} [c_i(\mathbf{h}_i'(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \\
&\leq \frac{1}{R} \sum_{r=1}^R \mathbf{1}[\theta_i^r = \theta_i^*] c_i(\mathbf{h}_i^r(\theta_i^r), \mathbf{h}_{-i}^r(\theta_{-i}^r), \boldsymbol{\lambda}) + \frac{1}{R} \sum_{r=1}^R \mathbf{1}[\theta_i^r = \theta_i^*] c_i(\mathbf{h}_i'(\theta_i), \mathbf{h}_{-i}^r(\theta_{-i}^r), \boldsymbol{\lambda}) + 2H \sqrt{\frac{2 \ln \frac{2}{\delta}}{R}} \\
&\leq \epsilon_i(R) + 2H \sqrt{\frac{2 \ln \frac{2}{\delta}}{R}},
\end{aligned}$$

and:

$$\begin{aligned}
& \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta_i^*} [c_i(\mathbf{h}_i(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] - \mathbb{E}_{\mathbf{h} \sim \sigma^R} \mathbb{E}_{\theta_{-i}, \boldsymbol{\lambda} \sim P | \theta_i^*} [c_i(\mathbf{h}_i'(\theta_i), \mathbf{h}_{-i}(\theta_{-i}), \boldsymbol{\lambda})] \\
&\leq \frac{\epsilon_i(R) + 2H \sqrt{\frac{2 \ln \frac{2}{\delta}}{R}}}{\Pr(\theta_i^*)}.
\end{aligned}$$

This completes the proof. □