Convergence Rates in Stochastic Stackelberg Games with Smooth Algorithmic Agents

Eric Frankel, Kshitij Kulkarni, Dmitriy Drusvyatskiy, Sewoong Oh, Lillian J. Ratliff

August 24, 2024

Abstract

Mechanisms are frequently designed to influence downstream competitive agent behavior, but face two critical challenges. First, decision-makers might not *a priori* know the agents' objectives. Second, agents might *learn* their response, introducing stochasticity into the decision-making process. The hierarchical interaction between the decision-maker and agents naturally abstracts to a Stackelberg game. When the decision-maker's action induces a strongly monotone game, we provide convergence rates on several fundamental settings. Together, these analyses demonstrate how convergence changes in low information environments as the decision maker progresses from oblivious to naïve to strategic. First, when the decision-maker sequentially selects actions oblivious to the agents, we identify regimes governed by the drift-to-noise ratio and characterize agents' optimal play. Second, when the decision-maker naïvely employs repeated retraining, we define a novel equilibrium concept—performatively stable Stackelberg equilibrium—that enables us to characterize outcomes and convergence rates. Finally, when the decision-maker is cognizant of the reactive nature of the agents yet does not know their objectives, we present an episodic derivative-free method and characterize the rate of convergence to the true Stackelberg equilibrium. Numerical experiments illustrate our theoretical results and their application to real-world scenarios.

1 Introduction

In many practical applications, a decision-maker takes actions that shape the behavior of a group of potentially competitive agents (Cai et al., 2015; Dean et al., 2024; Kim and Perdomo, 2023; Ratliff and Fiez, 2020; Westenbroek et al., 2019). For example, recommendation systems deploy predictive models of engagement to encourage particular user interactions (Calvano and Polo, 2021; Hardt et al., 2022). Similarly, in crowdsourcing markets, appropriately-priced incentives are often used to elicit high-quality responses (Dasari et al., 2020; Hu et al., 2018; Shah and Zhou, 2016; Xie et al., 2014). Fundamental to these applications is the assumption that agent preferences are *fully known* and *stationary*; that is, the decision-maker knows how agents will react and expects them to react the same over time (Fiez et al., 2020). In practice, however, this assumption is frequently violated, particularly when competitive agents *adapt* their response to a decision-maker's action (Fiez et al., 2021; Nar et al., 2017; Perdomo et al., 2020). These scenarios become even more challenging when the decision-maker has limited knowledge about agents' objectives or ability to react.

A natural abstraction for such an interaction between the decision-maker and agents is a Stackelberg game, where the decision-maker takes the role of *leader* and the agents take the role of *followers* (Li and Sethi, 2016; Von Stackelberg, 2010). Stackelberg games are a bilevel stochastic optimization problem from the perspective of the decision-maker wherein the decision-maker seeks to optimize its objective subject to the agents collective behavior (Colson et al., 2007; Von Stackelberg, 2010). Within this setting, we focus on a decision-maker that seeks to sequentially minimize a cost function that is evolving according to unknown, possibly stochastic dynamics, which arise from the agents tracking an equilibrium induced by the decision-maker's deployed actions. If the decision-maker faces a single agent, then this setting is reminiscent of many problems in machine learning, optimization, and control, such as performative prediction, concept drift, stochastic tracking and adaptive control (Cutler et al., 2021; Hardt et al., 2016; Perdomo et al., 2020). We discuss further relevance of these topics in the related work Section 1.2. Incorporating *multiple agents* and the structure of the *game* into the dynamics remains an open challenge that we seek to resolve.

Motivated by practical considerations, in this work, we study three scenarios in which a decision-maker might choose actions to solve their stochastic Stackelberg game: (1) the decision-maker that obliviously deploys a sequence of actions, (2) the naïve decision-maker that engages in repeated retraining, and (3) a strategic decision-maker that leverages a derivative-free gradient estimate to optimize through the agents dynamic behavior. These scenarios reflect the decision-maker's ability to reason about the agents' behavior. We also include numerical experiments that illustrate our theoretical results and their application to real-world scenarios via semi-synthetic simulations in Appendix B.

1.1 Contributions

We analyze three important settings wherein a decision-maker faces multiple, potentially competing, dynamically adapting agents and provide relevant theory and algorithms. Motivated by many practical settings, the algorithms are epoch-based, meaning that the decision-maker updates their action on a slower "time-scale" than the agents. The agents update their actions by playing one of several (stochastic) algorithms—which we discuss in Appendix F.

- ***Oblivious Decision-Maker***: The oblivious decision-maker simply deploys a sequence of actions u_t that induce *drift* in the agents' stochastic learning algorithms and induce a sequence of games and corresponding equilibria. In Section 3, we identify regimes governed by the drift-to-noise ratio and characterize the agents' optimal play via non-asymptotic convergence guarantees. In particular, we bound the equilibrium tracking error in each of the regimes, and we provide high probability tracking bounds that provide guarantees that hold in settings with irreversible drift such as learning with adaptive agents that strategically respond to deployed actions of both the decision-maker and other agents. These efficiency estimates expose how the equilibrium error decouples from the noise in the agents' learning process and the time drift in the game.
- ***Naïve Decision-Maker***: The naïve decision-maker recognizes that there a is distribution shift. In Section 4.1, we define an appropriate notion of performatively stable Stackelberg equilibrium, characterize its existence, and upper bound the *performative gap* between the performatively stable and true Stackelberg equilibrium. When the decision-maker employs a stochastic gradient method to repeatedly retrain, we show convergence to an approximate performatively stable Stackelberg equilibrium in $\mathcal{O}(\log(1/\varepsilon) + \sigma^2/\varepsilon)$ iterations, where σ^2 is the variance of the gradient estimator.
- ***Strategic Decision-Maker***: The strategic decision-maker recognizes that the agents are dynamically responding to their actions, yet does not know the agents objectives and hence cannot directly optimize their loss. In Section 4.2, we devise a derivative-free method that converges to an approximate Stackelberg equilibrium of the underlying game between the learning agents and the decision-maker in $\mathcal{O}((d^2/\varepsilon^2)\log(1/\varepsilon))$ iterations, where d is the dimension of the decision-maker's action space; this matches the optimal rate for a single player stochastic decision problem, up to logarithmic factors (Wood and Dall'Anese, 2022).

1.2 Related Work

Asymptotic equilibrium tracking is a long studied problem in single player stochastic optimization and stochastic approximation; see (Borkar, 2009; Kusher and Yin, 1997) and references therein. Our work focuses on obtaining convergence rates when the decision-maker faces a time-varying stochastic optimization problem subject to equilibrium constraints that are themselves varying in time. Below we highlight the most relevant work in this broad field, focusing on recent developments.

Static Performative Prediction. The decision-maker's problem is analogous to the setting of performative prediction, first introduced in (Perdomo et al., 2020), in the sense that the decision-maker faces a stochastic optimization problem where the distribution describing the environment depends on the actions of the decision-maker. Performative prediction, in turn, shares many features with the earlier work on stochastic optimization with decision-dependent probabilities (Hellemo et al., 2018) and strategic classification (Hardt et al., 2016; Mendler-Dünner et al., 2020). Numerous recent papers have developed algorithms and convergence guarantees in different performative prediction settings (Brown et al., 2022; Cutler et al., 2022, 2024; Drusvyatskiy and Xiao, 2023; Maheshwari et al., 2022; Mendler-Dünner et al., 2020; Miller et al., 2021; Narang et al., 2022). In particular, Mendler-Dünner et al. (2020) develops the first stochastic optimization algorithms within the performative prediction setting. The subsequent work by Drusvyatskiy and Xiao (2023) reveals that all the typical stochastic optimization algorithms used in the classical static setting extend directly to the performative setting with no loss in efficiency. The work Cutler et al. (2024) moreover shows that the basic stochastic gradient method asymptotically achieves the best possible sample complexity among any estimation procedures. Recent work by Conger et al. (2023) extends the specific sub-problem known as strategic classification to functional spaces by way of optimal transport in order to analyze the effects of the entire distribution (as compared to the mean) as a function of the decision-maker's action. Another interesting direction is explored in Narang et al. (2022) wherein the authors extend the performative prediction problem to multiple players and characterize the Nash equilibrium.

Time-Varying Stochastic Optimization & Performative Prediction. Of the recent work on performative prediction, the most closely related work focuses on performative prediction problems that change dynamically over time in response to exogenous changes in the environment. (Brown et al., 2022) introduced the notion of dynamics in the performative prediction problem through repeated risk minimization. Ray et al. (2022) introduce novel epoch-based algorithms for performative prediction when the environment is subject to geometrically decaying dynamics. There has also been a recent surge on the empirical front in related fields such as recommendation design when the decision-maker recognizes that the user pool may be reactive (Cen et al., 2024). Cutler et al. (2023) provides convergence rates for gradient-based stochastic optimization methods over time-varying decision-dependent distributions. Wood and Dall'Anese (2023) develop a similar analysis for zero sum games, and provide bounds on tracking stochastic saddle point equilibrium. Finally, (Cen et al., 2024) studies performative and strategic effects in recommendation systems, and provides a theoretical model to study user strategization along with an empirical study. Of these, the analysis in (Cutler et al., 2023) is most closely related to our work, especially in the oblivious decision-maker setting. We extend the analysis in that paper to strongly monotone games. Further, none of these works considers the Stackelberg setting in which the decision-maker (leader) faces multiple competing agents (followers) who are themselves learning and adapting.

Bilevel Optimization & Stackelberg Games. There is vast work on bilevel optimization and Stackelberg games (Başar and Olsder, 1998; Bracken and McGill, 1973; Colson et al., 2007; Dempe and Zemkoho, 2020; Stackelberg et al., 1952); the specific work most related to this paper focuses on settings where the agent problem is an equilibrium problem or variational inequality. In this setting, the literature is specialized to mathematical programming with equilibrium constraints. Prominent examples include settings where a leader optimizes over the outcome of a Cournot game (Sherali et al., 1983), or Stackelberg congestion games (Wardrop, 1952). Typically it is assumed that the decision-maker has full knowledge of the agent game or can control the agent game through multiple specialized queries such as in recent work (Li et al., 2023; Maheshwari et al., 2023). There has also been work on incentive design when facing multiple adaptive agents such as (Ratliff and Fiez, 2020; Yang et al., 2020, 2022); however, the majority of this work makes the assumption that the decision-maker can estimate the preferences of the agents, can compute the *a priori* optimal solution to use as a benchmark, gives asymptotic convergence guarantees, or provides empirical results. In contrast, our work does not assume that the decision-maker has any knowledge of the agent preferences or update methods beyond belonging to a broad contractive update class.

2 Preliminaries

Throughout, we use $\mathbb{R}^d$ to denote a d-dimensional space with inner product $\langle \cdot, \cdot \rangle$ and the corresponding induced norm is given by $\|x\| = \sqrt{\langle x, x \rangle}$. For any set $\mathcal{X} \subset \mathbb{R}^d$, we denote the projection of a vector y onto $\mathcal{X}$ as $\mathrm{proj}_{\mathcal{X}}(y) = \mathrm{argmin}_{x \in \mathcal{X}} \|x - y\|$. Finally, for a convex set $\mathcal{X}$, we denote its normal cone at $x \in \mathcal{X}$ as $N_{\mathcal{X}}(x) = \{v \in \mathbb{R}^d : \langle v, y - x \rangle \leq 0 \;\; \forall\, y \in \mathcal{X}\}$. To simplify notation, we set $[n] := \{1, \ldots, n\}$.

Throughout the paper, we focus on a *Stackelberg game.* That is, a decision-maker takes actions u lying in some closed convex set $\mathcal{U} \subseteq \mathbb{R}^d$, which influence the behavior of n competitive agents. In the next two sections, we describe in turn how the agents compete based on the action u and how the decision maker chooses the action u in the first place.

2.1 Induced Game Amongst Agents

Given a decision-maker's action $u \in \mathcal{U}$, each player $i \in [n]$ seeks to solve the problem

$$\min_{x_i \in \mathcal{X}_i} f_i^u(x_i, x_{-i}).$$

Here, $\mathcal{X}_i \subseteq \mathbb{R}^{m_i}$ denotes the set of actions that agent i can take and $f_i^u(x_i, x_{-i})$ denotes a C^2-smooth loss function of agent i that is induced by the decision-maker's action u. For simplicity, we use the notation $x := (x_i, x_{-i})$, where x_i denotes the action of agent i and x_{-i} denotes the action of all agents other than i. We will moreover interchangeably use the notation $f_i^u(x) = f_i(x, u) = f_i(x_i, x_{-i}, u)$, whenever convenient. We let $\mathcal{X} := \prod_i \mathcal{X}_i$ denote the agents' joint action space and we set $m := \sum_{i=1}^n m_i$.

Together, the functions $\mathcal{G}_u := (f_1^u, \ldots, f_n^u)$ define the *game* induced by $u \in \mathcal{U}$. We say that $\mathcal{G}_u$ is a C^1*-smooth convex game* if, for each $i \in [n]$, the set $\mathcal{X}_i$ is closed and convex, the function $f_i^u(\cdot, x_{-i})$ is convex in x_i for all fixed $(u, x_{-i}) \in \mathcal{U} \times \mathcal{X}_{-i}$, and the partial gradient $\nabla_i f_i^u(x)$ (with respect to x_i) exists and is continuous. Furthermore, a C^1-smooth convex game is called μ*-strongly monotone* for $\mu > 0$ if the inequality

$$\langle \omega_u(x) - \omega_u(x'), x - x' \rangle \geq \mu \|x - x'\|^2 \quad \text{holds for all} \quad x, x' \in \mathcal{X} \subseteq \mathbb{R}^m,$$

where the map $\omega_u(x) := (\nabla_1 f_1^u, \ldots, \nabla_n f_n^u)$ is called the *game Jacobian*. Examples of strongly monotone games include quadratic games, Kelly auctions, and Bertrand competition; proofs are given in Appendix A.

The most classical solution concept for the induced game is a *Nash equilibrium*. Given a fixed action $u \in \mathcal{U}$, a strategy $x^* \in \mathcal{X}$ is a *Nash equilibrium for* $\mathcal{G}_u$ if the condition holds:

$$f_i^u(x_i^*, x_{-i}^*) \leq f_i^u(x_i, x_{-i}^*) \qquad \text{for all } x_i \in \mathcal{X}_i \text{ and all } i \in [n].$$

This inequality can succinctly be summarized by the inclusion $0 \in \omega_u(x^*) + N_{\mathcal{X}}(x^*)$. Denote the set of Nash equilibria for $\mathcal{G}_u$ as $\mathtt{Eq}(\mathcal{G}_u)$. We impose the following assumption throughout.

Assumption 1. We assume that for each action $u \in \mathcal{U}$, the following are true: (i) the induced game $\mathcal{G}_u$ is μ–strongly monotone, (ii) the mappings $x_i \mapsto \nabla_i f_i^u(x_i, x_{-i})$ are L_i–Lipschitz continuous, and (iii) $\|\nabla_u \omega_u(x)\|_{\mathtt{op}}$ is bounded.

By standard results on the theory of variational inequalities (Dontchev et al., 2009; Rockafellar and Wets, 2009), Assumption 1.i implies that the Nash equilibrium of the induced game $\mathcal{G}_u$ is unique (Rosen, 1965). Furthermore, Assumptions 1.i and iii allows us to conclude (by way of the implicit function theorem (Dontchev et al., 2009, Theorem 2F.10)) that $x^*(\cdot)$ is Lipschitz continuous; we hereafter refer to this Lipschitz parameter as $L_{\mathtt{eq}}$. We provide further discussion on Assumption 1 in Appendix E; e.g., one sufficient condition for (i) is that the game Jacobian $\nabla_x \omega_u(x)$ is positive definite and this could be incorporated as an additional constraint on u as in prior literature (Ratliff and Fiez, 2020).

2.2 Decision-Maker's Stochastic Optimization Problem

The decision-maker seeks to minimize a loss $\ell : \mathcal{U} \times \mathcal{X} \to \mathbb{R}$ that depends on their action and the induced equilibrium $x^*(u)$ of the game $\mathcal{G}_u$. The most salient solution concept for the decision-maker is a *Stackelberg equilibrium*:

$$u^* \in \underset{u \in \mathcal{U}}{\operatorname{argmin}} \{\ell(u, x^*(u)) |\ x^*(u) \in \mathtt{Eq}(\mathcal{G}_u)\}. \tag{1}$$

However, there are two potential sources of uncertainty: (i) the stochasticity of the environment which causes the observations of the decision maker to be noisy, and (ii) the agents are trying to learn the induced equilibrium $x^*(u)$ and may be employing a stochastic algorithm to do so. We model these noise sources through the observation model $z = X + \xi \sim \mathcal{D}(u)$ where $X \in \mathcal{X}$ is a random variable representing the joint action which is the output of the (potentially stochastic) algorithms of the agents (thereby capturing (ii)) and $\xi \sim \mathcal{D}_o$ is a zero-mean, finite variance random variable representing the environment stochasticity (thereby capturing (i)). Throughout, let $\mathcal{L}(u) := \mathbb{E}_{z\sim\mathcal{D}(u)}[\ell(u,z)]$ denote the decision-maker's expected loss where the expectation

The difficulty for the decision maker, however, is that *they do not* a priori *have access to the agents' loss functions nor algorithms they may be employing.* This means that, while the decision-maker would like to optimize $\mathcal{L}$, it cannot compute the full gradient of $\mathcal{L}$. That is, the second term in

$$\nabla\mathcal{L}(u) = \mathop{\mathbb{E}}_{z\sim\mathcal{D}(u)} \nabla_u \ell(u,z) + \frac{d}{dv} \mathop{\mathbb{E}}_{z\sim\mathcal{D}(v)} \ell(u,z)|_{v=u}$$

is unavailable, whereas the first can be easily estimated from sampling.

2.3 Smooth Algorithmic Agents

In practice, the agents may not be able to instantaneously compute a Nash equilibrium for each deployed action u, an observation noted in related settings such as minimax optimization (Fiez et al., 2021). Instead, the agents are more likely to employ a learning algorithm that approximates an equilibrium. In such a setting, we assume that the decision-maker deploys an action u_t for a number of iterations $\tau_t \in \mathbb{N}$ within epoch t. Thus, for a fixed action u_t, each agent $i \in [n]$ independently updates their actions according to an algorithm $\bar{\mathcal{A}}_i$—i.e., $x_{i,t}^{k+1} = \bar{\mathcal{A}}_i(x_t^1, x_t^2, \ldots x_t^k, u_t)$ for $k \in [\tau_t]$, where $\tau_t \geq 1$ is the duration for which u_t is fixed. Set $x_t^0 := x_{t-1}^{\tau}$, $x_{t+1} := x_t^{\tau}$, and $x_0^0 := x_0$, abusing notation slightly. Let x_{t+1} be the agents' joint response after running $\bar{\mathcal{A}} = (\bar{\mathcal{A}}_1, \ldots, \bar{\mathcal{A}}_n)$ for τ_t timesteps. We consider a broad class of algorithms for the agents that adhere to the following definition.

Definition 1 (**ρ-contracting stochastic algorithm**)**.** For constants $\rho \in (0,1)$ and $\sigma_{\mathtt{a}} \in [0,\infty)$, the update rule employed by an agent is called a ρ-contracting stochastic algorithm if the following holds:

$$\mathbb{E}\,\|x_t^{k+1} - x^*(u_t)\|^2 \;\leq\; \rho^2\,\mathbb{E}\,\|x_t^k - x^*(u_t)\|^2 + \rho^2(c\sigma_{\mathtt{a}})^2,$$

for some $c > 0$, where $x^*(u_t) \in \mathtt{Eq}(\mathcal{G}_{u_t})$.

Note that *deterministic algorithms* are characterized by $\sigma_a = 0$. There are many examples of algorithms that satisfy Definition 1, including stochastic gradient play, asynchronous stochastic gradient play, best response dynamics, and even momentum-based gradient play in strongly convex-strongly concave zero-sum games. Below we hightlight a few common updates; further examples along with proofs appear in Appendix F.

Example 1 (Stochastic Gradient Play)**.** Consider first players updating according to

$$x_{t+1} = \mathop{\mathrm{proj}}_{\mathcal{X}}(x_t - \gamma\widehat{\omega}(x_t)), \quad \text{where} \quad \mathbb{E}[\widehat{\omega}(x_t)] = \omega(x_t).$$

As long as there exists a constant $\sigma_{\mathtt{a}} > 0$ satisfying $\mathbb{E}[\|\widehat{\omega}(x_t) - \omega(x_t)\|^2] \leq \sigma_{\mathtt{a}}^2$, then stochastic gradient play satisfies (cf. Lemma 5)

$$\mathbb{E}\,\|x_{t+1} - x^*\|^2 \leq \frac{1}{1+\mu\gamma}\mathbb{E}\|x_t - x^*\|^2 + \frac{2\gamma^2\sigma_{\mathtt{a}}^2}{1+\gamma\mu},$$

so that this update is ρ-contracting with $\rho^2 = \frac{1}{1+\mu\gamma}$ and $c = \sqrt{2}\gamma$.

Example 2 (Asynchronous Stochastic Gradient Play)**.** Asynchronous stochastic gradient play is a slight modification of stochastic gradient play wherein players update asynchronously. In each iteration, the probability that player i receives new information and therefore updates its action is $p_i \in (0,1]$. The update in this case is

$$x_{i,t+1} = \begin{cases} \mathop{\mathrm{proj}}_{\mathcal{X}_i}\left(x_{i,t} - \gamma\widehat{\omega}_i(x_{i,t}, x_{-i.t})\right), & \text{w.p. } p_i \\ x_{i,t}, & \text{w.p. } (1-p_i) \end{cases}$$

or compactly $x_{t+1} = \mathrm{proj}_{\mathcal{X}}(x_t - \gamma P\widehat{\omega}(x_t))$, where $P = \mathrm{diag}(p_1 I_{m_1}, \ldots, p_n I_{m_n})$ with $x_i \in \mathbb{R}^{m_i}$. This update is also ρ–contracting. The analysis in Lemma 5 (Appendix F) does not change much; the primary difference is that the Lipschitz constant $L_{\mathtt{a}}$ is rescaled by $p_{\mathtt{max}} := \max\{(p_1, \ldots, p_n)\}$ and the strong monotonicity constant μ is rescaled by $p_{\mathtt{min}} := \min\{(p_1, \ldots, p_n)\}$. The reason this works out is that we can simply perform the exact same analysis using a modified inner product as has been performed in prior literature—i.e., we simply perform the analysis in the inner product $[x,y] = \langle P^{-1}x, y\rangle$

Example 3 (Best Response)**.** Best response updates are another common update in the learning in games literature. The best response update is given by

$$x_{i,t+1} = \operatorname*{argmin}_{x_i \in \mathcal{X}_i} f_i(x_i, x_{-i,t}) \quad \forall\ i \in [n].$$

In Lemma 6 (cf. Appendix F) we show that the best response update is ρ–contracting with $\rho := \frac{L_{\mathtt{a}}\sqrt{n-1}}{\mu}$ and $\sigma_{\mathtt{a}} = 0$ as long as the players are in the regime where $\rho < 1$. This lemma also proves that a Nash equilibrium exists for the game $\mathcal{G} = (f_1, \dots, f_n)$.

Example 4 (Momentum-Based Updates in Zero-Sum Games)**.** Strongly convex, strongly concave zero sum game $(f, -f)$ where player one seeks to minimize $f(x_1, x_2)$ with respect to x_1 and player two seeks to maximize f with respect to x_2 are known to be strongly monotone. The family of momentum-based updates—such as optimistic gradient descent-ascent (OGDA) and negative momentum—are ρ–contracting for such games. This family of updates is given by

$$x_{t+1} = (1+\beta)x_t - \beta x_{t-1} - \gamma((1+\alpha)\omega(x_t) - \alpha\omega(x_{t-1})), \tag{2}$$

where α is the extrapolation parameter, β is the momentum parameter, and $\omega(x) = (\nabla_1 f(x), -\nabla_2 f(x))$ with $x = (x_1, x_2)$. For example, standard gradient descent-ascent is equivalent to setting $(\alpha, \beta) = (0, 0)$. OGDA is given by $(\alpha, \beta) = (1, 0)$ and negative momentum is given by $(\alpha, \beta) = (0, \beta)$ for some $\beta < 1$.

2.4 Understanding the Challenges to Equilibrium Tracking & Convergence

To gain some intuition for the results and the technical challenges, we decompose the *equilibrium tracking error* into three salient components. Consider a sequence of decision-maker actions $\{u_t\}$. From the perspective of the agents, they are trying to learn a time-varying equilibrium $x_t^\star \in \mathtt{Eq}(\mathcal{G}_{u_t})$. As we will formally show, the tracking error decomposes as

$$\mathbb{E}\,\|x_t - x_t^\star\|^2 \;\lesssim\; \rho^{2t}\|x_0 - x_0^\star\|^2 + \frac{\sigma_{\mathtt{a}}^2}{1-\rho^2} + \left(\frac{\Delta_{\mathtt{a}}}{1-\rho^2}\right)^2, \tag{3}$$

where $\Delta_{\mathtt{a}} := \max_k\{\|x_k^\star - x_{k-1}^\star\|^2\}$ is the *drift* and $\sigma_{\mathtt{a}}$ characterizes the *noise*. The first term in (3) is exponentially decaying so that, as $t \to \infty$, we are left with the drift and noise terms. This raises the following challenge:

> Challenge 1: *Given the decision-maker induced drift in the game, can we identify the drift-to-noise regimes and characterize their optimal choice of contraction rate ρ and corresponding target accuracy?*

This is a non-trivial extension of the single player setting since we can no longer appeal to the function value to make the argument as in (Cutler et al., 2023). In Corollary 1 and Proposition 2, we characterize these regimes and optimal contraction rate for stochastic gradient play, which allows for the agent to equilibriate to even the worst case drift.

From the decision-maker's point of view, the agents are employing an unknown (stochastic) algorithm. This means the decision-maker's problem is time-varying as well. Beyond the oblivious deployment of $\{u_t\}$, the decision-maker hopes to control the drift induced amongst the agents to ensure that $\|u_t^\star - u_{t-1}^\star\|^2 \to 0$. For example, if the decision-maker's loss ℓ is α-strongly convex in u and she employs a stochastic gradient-based algorithm with stepsize η, then

$$\mathbb{E}\,\|u_t - u_t^\star\|^2 \;\lesssim\; \left(1 - \frac{\eta\alpha}{2}\right)^t \|u_0 - u_0^\star\|^2 + \frac{\eta\sigma^2}{\alpha} + \left(\frac{\Delta}{\alpha\eta}\right)^2, \tag{4}$$

where σ^2 is the gradient estimator variance and $\Delta := \max_k \|u_k^\star - u_{k-1}^\star\|^2$ is the drift. Moreover, $u_t^\star \in \operatorname{argmin}_u\{\mathbb{E}_{\xi\sim\mathcal{D}_o}\,\ell(u, x_t + \xi)\}$, where x_t is the agents' response from a (potentially stochastic) algorithm $\mathcal{A}$. There are two natural settings: the decision-maker (i) naïvely employs repeated retraining (common in practice) that does not account for the dependence of z on u, and (ii) strategically employs a method that accounts for the aforementioned dependence. For (i), we introduce the *performatively stable Stackelberg equilibrium*—which mirrors the notion of a performatively stable point in performative prediction and a performatively stable Nash equilibrium in the multiplayer setting. For (ii), a Stackelberg equilibrium is natural since the decision-maker is accounting for the reaction of the agents in its algorithm. For these settings, we address the following:

> Challenge 2: *Can we design algorithms (or show when they exist) such that $u^\star_t \to u^\star$ where $(u^\star, x^*(u^\star))$ is an (appropriate) equilibrium?*

In Theorems 3, and 4, we give convergence results for attaining an approximate performatively stable Stackelberg equilibrium or approximate Stackelberg equilibrium. Algorithms for obtaining performatively stable Stackelberg equilibrium are much more sample efficient as compared to those that obtain to Stackelberg equilibrium. This poses a problem for the decision-maker: how should they assess the tradeoff between performance degradation—reaching a sub-optimal equilibrium like performatively stable Stackelberg equilibrium—and sample complexity? More explicitly, consider the scenario shown in the left of Figure 1. Here, a decision-maker could use a derivative-free optimization method to reach within $\varepsilon > 0$ of the true Stackelberg equilibrium u^*, *or* use the more efficient repeated gradient method to reach within $\varepsilon' = \varepsilon - \|u^{\mathsf{ps}} - u^*\|$ of the performatively stable Stackelberg equilibrium u^{ps}; in either case, the *worst-case* expected distance from u^* is the same. With this in mind, we aim to answer the following:

> Challenge 3: *What is the performance gap between the performatively stable Stackelberg equilibrium versus the true Stackelberg equilibrium?*

In Theorem 2, we first show when performatively stable Stackelberg exist in terms of problem parameters. Then, in Proposition 4, we characterize the performance gap in terms of properties of the game—in particular, how dynamic agents are in reaction to the decision-maker. To gain some intuition consider the right side of Figure 1 reveals that depending on how *reactive*—as measured by the Lipschitz constant of $x^*(\cdot)$—the agents are, taking into consideration performativity may or may not be beneficial.[1] When the reactivity (L_{eq}) is small, the repeated gradient method is not only more sample efficient but also obtains a performatively stable equilibrium that is near the Stackelberg equilibrium. As L_{eq} grows, the performative gap $\|u^{\mathsf{ps}} - u^*\|$ grows so that ε' has to shrink in order to for $\varepsilon' + \|u^{\mathsf{ps}} - u^*\|$ to remain inside the ε-ball around u^*. Therefore as the reactivity increases, there is a decreasing marginal gain in terms of sample complexity from running repeated gradient descent as compared to the derivative free method.

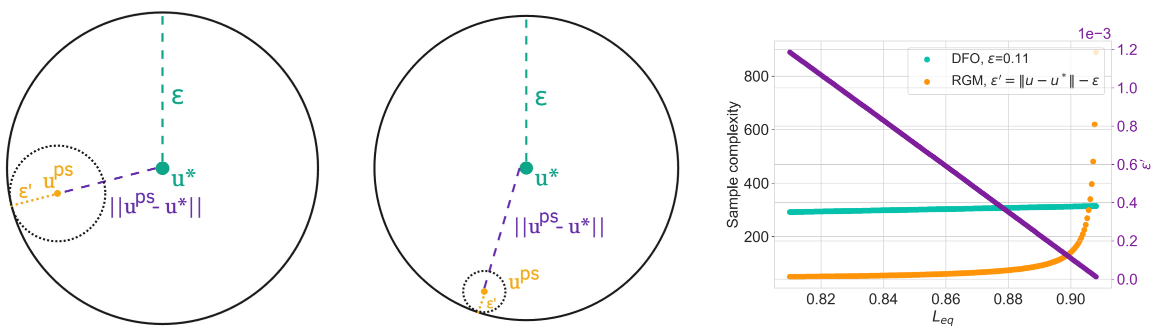


Figure 1: (Left/Center) A scenario where the performatively stable equilibrium u^{ps} is within $\varepsilon > 0$ of u^*. By the triangle inequality, the quality of solution within ε of u^* is similar to a solution within $\varepsilon' = \varepsilon - \|u^{\mathsf{ps}} - u^*\|$ of u^{ps}. As the reactivity L_{eq} grows, $\|u^{\mathsf{ps}} - u^*\|$ grows so that ε' has to shrink for u^{ps} to stay within ε of u^*. (Right) Big-$\mathcal{O}$ sample complexity versus L_{eq} approximated from Theorems 3, 4. Depending on how reactive the agents are, using the repeated gradient method within this scenario can become more costly in sample complexity than using a derivative-free method; on the other hand, when the reactivity is low the repeated gradient method returns an ε'–performatively stable equilibrium that is closer to the Stackelberg equilibrium than the derivative free method in fewer samples.

[1] We take a particular Stackelberg game and use the big-O sample complexity bounds to compare the methods. More details on how we created this plot including code are included in the supplement.

3 Oblivious Decision-Maker

An *oblivious decision-maker* deploys a sequence of actions $\{u_s\}_{s=1}^t$ and passively observes how the agents respond according to some collective set of algorithms $\mathcal{A}$ as described in Section 2. This includes pricing or recommendations where repeated or periodic changes are made to interventions. The changing decision-maker's actions lead to drifting equilibrium: $x_t^\star := x^*(u_t)$. Given a bounded worst-case discrepancy, $\Delta := \max_t\{\|u_t - u_{t+1}\|\}$, *can we bound the time to reach a target equilibrium tracking error,* $\mathbb{E}\|x_t - x_t^\star\|^2$*?* We start by first proving that the expected equilibrium tracking error is bounded. In many practical settings we may not be able to run the algorithm repeatedly which is required for interpreting the expected error bounds; indeed, in game theoretic settings users can learn through repeated interaction due to their inherent strategic nature and therefore there may be irreversible drift. To address this, we state a tracking error bounds which hold with high probability.

3.1 Expected Equilibrium Tracking Error

To address this, we extend recent work in the single player stochastic optimization setting, namely (Cutler et al., 2023), to monotone games.

Suppose the decision-maker deploys actions u_t such that $x_t^\star \in \mathtt{Eq}(\mathcal{G}_{u_t})$ is the induced equilibrium for the μ–strongly monotone game $\mathcal{G}_{u_t}$. Under Assumption 1, for any t, we have that $\mathbb{E}\,\|x_{t+1}^\star - x_t^\star\| \leq L_{\mathsf{eq}}\,\mathbb{E}\,\|u_{t+1} - u_t\|$, so the agents' problem is drifting in time depending on the sequence of actions u_t. As long as the agents employ a ρ-contracting stochastic method as in Definition 1, then it is possible to bound the expected equilibrium tracking error with a notable dependence on σ_a.

Proposition 1 (Informal)**.** Under Assumption 1, if agents employ a ρ-contracting stochastic algorithm in the regime $\rho \in [0, 1)$. Then, the expected equilibrium tracking error satisfies

$$\mathbb{E}\|x_t - x_t^\star\|^2 \;\lesssim\; \left(1 - \frac{(1-\rho^2)}{2}\right)^t \|x_0 - x_0^\star\|^2 + \frac{(c\sigma_{\mathsf{a}})^2}{1-\rho^2} + \left(\frac{\Delta_{\mathsf{a}}}{1-\rho^2}\right)^2,$$

where $\Delta_{\mathsf{a}} := \max_t\{\|x_{t+1}^\star - x_t^\star\|\}$ and ρ, c, and σ_{a} are from Definition 1.

Proposition 1, proved in Appendix G, gives a *worst case* t-step tracking error as it depends on $\Delta_{\mathsf{a}} := \max\{\|x_{t+1}^\star - x_t^\star\|^2\}$. A natural question is whether or not algorithms exist for the agents that have last iterate convergence guarantees in this time-varying setting. Here, we focus on stochastic gradient play (SGP) since it provides deeper insight and intuition for the difficulty of the problem:

$$x_{t+1} = \operatorname*{proj}_{\mathcal{X}}(x_t - \gamma\widehat{\omega}_t) \quad \text{where} \quad \widehat{\omega}_t := (\widehat{\nabla}_1 f_1^{u_t}(x_t), \ldots, \widehat{\nabla}_n f_n^{u_t}(x_t)). \tag{SGP}$$

Proposition 1 reduces to the following corollary.

Corollary 1 (Informal)**.** Under the assumptions of Proposition 1, suppose that the agents are running SGP with stepsize $\gamma \leq \mu/(2L_{\mathsf{a}}^2)$, where $L_{\mathsf{a}} := \max_{i\in[n]} L_i$, and an unbiased gradient estimator $\widehat{\omega}$ with $\mathbb{E}[\|\widehat{\omega}_t - \mathbb{E}_t[\widehat{\omega}_t]\|^2] \leq \sigma_{\mathsf{a}}^2$. Then $\rho^2 = 1/(1+\gamma\mu)$ and $c = \sqrt{2}\gamma$ so that

$$\mathbb{E}_t\|x_t - x_t^\star\|^2 \lesssim \left(1 - \frac{\mu\gamma}{4}\right)^t \|x_0 - x_0^\star\|^2 + \frac{\gamma\sigma_{\mathsf{a}}^2}{\mu} + \left(\frac{L_{\mathsf{eq}}\Delta}{\gamma\mu}\right)^2.$$

Letting $t \to \infty$, the term depending on $\|x_0 - x_0^\star\|^2$ tends to zero, leaving only the remaining noise and drift terms. Optimizing these terms over γ leads to the optimal learning rate and asymptotic tracking error, respectively:

$$\gamma_\star := \min\left\{\frac{\mu}{2L_{\mathsf{a}}^2}, \left(\frac{2(L_{\mathsf{eq}}\Delta)^2}{\mu\sigma_{\mathsf{a}}^2}\right)^{1/3}\right\} \quad \text{and} \quad \varepsilon_\star := \min_{\gamma\in(0,\mu/(2L_{\mathsf{a}}^2)]}\left\{\frac{\gamma\sigma_{\mathsf{a}}^2}{\mu} + \left(\frac{L_{\mathsf{eq}}\Delta^2}{\mu\gamma}\right)^2\right\}.$$

The learning rate $\gamma_\star$ determines the interesting regimes. Indeed, setting $\mu/(2L_{\mathsf{a}}^2) = (2(L_{\mathsf{eq}}\Delta)^2/(\mu\sigma_{\mathsf{a}}^2))^{1/3}$ and rearranging, we have two regimes for the drift-to-noise ratio: we are in the *low* regime if $\Delta/\sigma_{\mathsf{a}} < \mu^2/(4L_{\mathsf{eq}}L_{\mathsf{a}}^2)$ and otherwise the *high* regime.

In the high drift-to-noise regime, if agents run stochastic gradient play with $\gamma_\star \asymp \mu/(2L_{\mathtt{a}}^2)$,[2] we have that

$$\mathbb{E}\,\|x_t - x_t^\star\|^2 \lesssim \varepsilon_\star \quad \text{in} \quad t \lesssim \frac{L_{\mathtt{a}}^2}{\mu^2}\log\left(\frac{\|x_0 - x_0^\star\|^2}{\varepsilon_\star}\right) \quad \text{time steps.}$$

This case matches the learning rate from the deterministic setting (see, e.g., (Chasnov et al., 2020a; Rosen, 1965)). The low drift-to-noise regime is decidedly more interesting: setting the step-size $\gamma_\star \asymp (2(\Delta L_{\mathtt{eq}})^2 \cdot \frac{1}{\mu\sigma_{\mathtt{a}}^2})^{1/3}$, stochastic gradient play produces a point x_t such that

$$\mathbb{E}\,\|x_t - x_t^\star\|^2 \lesssim \varepsilon_\star \quad \text{in} \quad t \lesssim \frac{\sigma_{\mathtt{a}}^2}{\mu^2\varepsilon_\star}\log\left(\frac{\|x_0 - x_0^\star\|^2}{\varepsilon_\star}\right) \quad \text{time steps.}$$

The following proposition shows that if the agents employ stochastic gradient play in stages, then much like the single player time-invariant optimization problem (Kulunchakov and Mairal, 2019), the agents tracking error can be greatly improved.

Proposition 2 (Informal)**.** Suppose that induced time-varying agent problem is in the low drift-to-noise regime. There is an algorithm that proceeds by running stochastic gradient play (SGP) in K stages with T_k steps in each of the $k \in [K]$ stages such that the total time $T = T_0 + \cdots + T_{K-1}$ satisfies

$$T \lesssim \frac{L_{\mathtt{a}}^2}{\mu^2}\log\left(\frac{2\|x_0 - x_0^\star\|^2}{\varepsilon_\star}\right)^+ + \frac{\sigma_{\mathtt{a}}^2}{\mu^2\varepsilon_\star},$$

and the expected tracking error satisfies $\mathbb{E}\,\|x_K - x_K^\star\|^2 \lesssim \varepsilon_\star$.

In Appendix G, we detail the construction of this algorithm and its parameters. Essentially, the agents employ a stage-based algorithm that repeatedly runs stochastic gradient play in stages by reinitializing the algorithm at the previous iteration and adjusting the stepsize by a factor of 2^{-k}, which progressively reduces the bias. The iteration complexity is then shown to be composed of two terms: the classical complexity $\mathcal{O}((L_{\mathtt{a}}^2/\mu^2)\log(1/\varepsilon_\star))$ of the gradient play for deterministic objectives, and another term which is the optimal complexity for stochastic gradient play in $\mathcal{O}(\sigma_{\mathtt{a}}^2/(\mu^2\varepsilon_\star))$. Crucially, the regime is controlled by the decision-maker's algorithm (due to the dependence on Δ) *as well as* the agents' step-size. As a result, the decision-maker is incentivized to design an algorithm where the agents' problem is in the low drift-to-noise regime, which can lead to a tracking error within a constant factor of $\varepsilon_\star$. More generally, even in time-varying games, the decision-make and agents can both choose algorithm design choices that lead to optimal (within a constant factor) asymptotic tracking error.

Remark 1 (Beyond Worst Case Expected Equilibrium Tracking Bounds)**.** The preceding results focus on the *worst-case equilibrium tracking error*. In Appendix G.2, we also give a time-varying equilibrium tracking error that assumes a particular contraction rate on the decision-maker's sequence $\{u_t\}$ and then characterizes the regimes and corresponding convergence rates for the agents' actions.

3.2 High Probability Bounds on the Tracking Error

In this section, we characterize high probability bounds on the tracking error; more formal details are given in Appendix G.3.

The expected equilibrium tracking results are characterized in terms of the expected tracking error; accordingly, characterizing the guarantees of the algorithm are only meaningful if it is run *multiple* times. In game theoretic settings, due to the strategic nature of users and their ability to learn, it is unlikely to repeat the same experiment over and over many many times. Instead, if the algorithm is deployed in *real-time with irreversible drift*, high-probability efficiency results are more meaningful since they characterize the performance of the algorithm if it were executed only once.

We require the following tail assumptions on the equilibrium drift and gradient noise.

Assumption 2 (Sub-Gaussian drift and noise)**.** There exist constants $\Delta_{\mathtt{a}}, \sigma_{\mathtt{a}} > 0$ such that the following two conditions hold for all $t \geq 0$:

[2] Here, $\asymp$ and $\lesssim$ indicate an equality and inequality, respectively, holding up to a constant.

(a) The drift $\Delta^2_{\mathtt{a},t}$ is sub-exponential conditioned on $\mathcal{F}_t$ with parameter $\Delta^2_{\mathtt{a}}$:

$$\mathbb{E}[\exp(\lambda\Delta^2_{\mathtt{a},t})|\ \mathcal{F}_t] \le \exp(\lambda\Delta^2_{\mathtt{a}}) \quad \text{for all} \quad 0 \le \lambda \le \Delta^2_{\mathtt{a}}$$

(b) The gradient noise ξ_t is norm sub-Gaussian conditioned on $\mathcal{F}_t$ with parameter $\sigma_a/2$:

$$\mathbb{P}(\|\xi_t\| \ge \zeta|\ \mathcal{F}_t) \le 2\exp(-2\zeta^2/\sigma^2_{\mathtt{a}}) \quad \text{for all} \quad \zeta > 0.$$

Note that Assumption 11 implies Assumption 10 under the with the same constants $\Delta_{\mathtt{a}}$, $\sigma_{\mathtt{a}}$. We need a slightly modified (in fact simpler) version of Proposition 29 from (Cutler et al., 2023), which is an extension of Claim D.1 from (Harvey et al., 2019).

Proposition 3. Consider a scalar stochastic process $\{V_t, D_t, X_t\}$ on a probability space with filtration $\mathcal{H}_t$ such that V_t is nonnegative and $\mathcal{H}_t$-measurable, and satisfies

$$V_{t+1} \le \alpha_t V_t + X_t$$

for some deterministic constant $\alpha_t \in (-\infty, 1]$. Suppose that the moment generating functions of X_t conditioned on $\mathcal{H}_t$ satisfies

$$\mathbb{E}[\exp(\lambda X_t)|\ \mathcal{H}_t] \le \exp(\lambda\nu_t) \quad \forall\ 0 \le \lambda \le 1/\nu_t,$$

for some constants $\sigma_t, \nu_t > 0$. Then the inequality

$$\mathbb{E}[\exp(\lambda V_{t+1})] \le \exp(\lambda\cdot\nu_t)\mathbb{E}\left[\exp\left(\lambda\alpha_t V_t\right)\right],$$

holds for all $0 \le \lambda \le \frac{1}{2\nu_t}$.

This proposition enables us to prove the following high probability tracking bound.

Theorem 1 (High probability tracking error.)**.** Suppose that Assumptions 1, 9, and 11 hold and that the decision-maker deploys a sequence $\{u_s\}_{s=0}^t$ satisfying $\frac{\Delta}{\sigma_{\mathtt{a}}} < \frac{\mu^2}{4\sqrt{3}L_{\mathtt{eq}}L^3_{\mathtt{a}}}$ so that the agents are in the low drift-to-noise regime. Let $\{x_t\}$ be the iterates produced by the agents running stochastic gradient play (Algorithm 3) with $\gamma \le \frac{\mu}{2L^2_{\mathtt{a}}}$. Then there exists an absolute constant $c > 0$ such that for any specified $t \in \mathbb{N}$ and $\delta \in (0, 1)$, the following estimate holds with probability at least $1-\delta$:

$$\|x_t - x^\star_t\|^2 \le \left(1 - \frac{\mu\gamma}{4}\right)^t \|x_0 - x^\star_0\|^2 + c\left(\frac{\sigma^2_{\mathtt{a}}\gamma}{\mu} + \left(\frac{\Delta_{\mathtt{a}}}{\mu\gamma}\right)^2\right)\log\left(\frac{e}{\delta}\right). \tag{5}$$

The proof of the preceding theorem is given in Appendix G.3. This theorem translates to a time-to-track high probability result.

Corollary 2 (Time to track with high probability.)**.** Suppose that the assumptions of Theorem 1 hold so that we are in the low drift-to-noise regime. Let $\{x_t\}$ be the iterates produced by the agents running stochastic gradient play (Algorithm 3) with $\gamma \le \frac{\mu}{2L^2_{\mathtt{a}}}$. Moreover suppose that we have Moreover, suppose that there is a constant R available such that $R \ge \|x_0 - x^\star_0\|^2$. Further, set constants

$$K = 1 + \left\lceil \log_2\left(\frac{\mu}{L^2_{\mathtt{a}}}\cdot\left(\frac{\sigma^2_{\mathtt{a}}\mu}{\Delta^2_{\mathtt{a}}}\right)^{1/3}\right)\right\rceil$$

and

$$\gamma_0 = \frac{\mu}{2L^2_{\mathtt{a}}}, \quad T_0 = \left\lceil \frac{8L^2_{\mathtt{a}}}{\mu^2}\log\left(\frac{L^2_{\mathtt{a}}R}{\sigma^2_{\mathtt{a}}}\right)^+\right\rceil, \quad \gamma_k = \frac{\gamma_{k-1}+\gamma_\star}{2}, \quad T_k = \left\lceil\frac{4\log(12)}{\mu\gamma_k}\right\rceil, \quad \text{for all} \quad k \ge 1.$$

Consider running stochastic gradient play (Algorithm 3) in $k = 0, \ldots, K-1$ stages. Then $T = T_0 + \cdots + T_{K-1}$ satisfies

$$T \lesssim \frac{L^2_{\mathtt{a}}}{\mu^2}\log\left(\frac{L^2_{\mathtt{a}}R}{\sigma^2_{\mathtt{a}}}\right)^+ + \frac{\sigma^2_{\mathtt{a}}}{\mu^2\varepsilon_\star} \le \frac{L^2_{\mathtt{a}}}{\mu^2}\log\left(\frac{R}{\varepsilon_\star}\right)^+ + \frac{\sigma^2_{\mathtt{a}}}{\mu^2\varepsilon_\star},$$

and for any given $\delta \in (0, 1)$, the expected tracking error satisfies $\mathbb{E}\,\|x_K - x^\star_K\|^2 \lesssim \varepsilon_\star \log\left(\frac{e}{\delta}\right)$ with probability at least $1-\delta$.

As noted, these results give us practically relevant theoretical bounds that apply in game theoretic settings where induced drift (e.g., by a decision-maker) is likely irreversible.

4 Convergence to Stackelberg Solutions: Drift-to-Noise Control

Recall from Section 2 that the decision-maker seeks to solve a stochastic optimization problem with both uncertainty due to the response of the agents and stochasticity of the environment—namely, they seek to minimize the loss $\mathcal{L}(u) = \mathbb{E}_{z\sim\mathcal{D}(u)}[\ell(u,z)]$ with respect to u where $z \sim \mathcal{D}(u)$ is the stochastic observation that the decision-maker receives from the environment and abstracts the agents' (stochastic) decision process. In the ideal setting, the agents would be playing in equilibrium for each deployed u, and the decision maker could then optimize for the Stackelberg solution for the "leader"—i.e., $z = x^*(u) + \xi$ for some zero-mean, finite-variance random variable $\xi \sim \mathcal{D}_o$. However, this is not the case in practice, as the agents are learning and adapting themselves.

4.1 Naïve Decision-Maker

We first consider a decision-maker who naïvely recognizes that the environment is reactive but does not have *a priori* access to the agents' response mapping. In particular, the underlying decision-dependence is unknown. A common approach in machine learning systems is a variant of *repeated retraining*. We introduce a novel equilibrium concept for such methods, devise a provably convergent epoch-based algorithm, and characterize its iteration complexity.

4.1.1 Performatively Stable Stackelberg

In the naïve setting, the decision-maker is pursuing what is known as a *performatively stable point*; however, the decision-dependence in this case is determined by an *equilibrium problem* which begs for a new equilibrium concept which we introduce.

Definition 2. A point $(u^{\mathsf{ps}}, x(u^{\mathsf{ps}}))$ is a *performatively stable Stackelberg equilibrium* if

$$u^{\mathsf{ps}} \in \operatorname*{argmin}_{u\in\mathcal{U}} \mathop{\mathbb{E}}_{z\sim\mathcal{D}(u^{\mathsf{ps}})}[\ell(u,z)] \quad \text{and} \quad x(u^{\mathsf{ps}}) \in \mathtt{Eq}(\mathcal{G}_{u^{\mathsf{ps}}}).$$

To analyze a performatively stable Stackelberg equilibrium, some regularity assumptions on ℓ are needed.

Assumption 3. The following hold:

a. The loss $\ell(\cdot,z)$ is C^1-smooth and α–strongly convex for any z;

b. The map $(u,z) \mapsto \nabla_{u,z}\ell(u,z)$ is L_ℓ–Lipschitz continuous, where $\nabla_{u,z}\ell \equiv (\nabla_u\ell, \nabla_z\ell)$.

Theorem 2 (Existence & Uniqueness of Performatively Stable Stackelberg Equilibrium)**.** Under Assumptions 1 and 3 and when $1 < \alpha/(L_{\mathsf{eq}}L_\ell)$, there exists a unique performatively stable Stackelberg equilibrium.

The proof, which is contained in Appendix H.1, follows from Banach's fixed point theorem. Once we know the regime in which performatively stable Stackelberg exists, it is natural to ask what the relationship is between such equilibrium and the *true* Stackelberg equilibrium (cf. (1)).

As it turns out, we can also bound the *performative gap*—i.e., the gap between these two equilibrium concepts.

Proposition 4 (Performative Equilibrium Gap.)**.** Under the assumptions of Theorem 2, if $\ell(u,z)$ is L_{z} Lipschitz continuous in z, then

$$\|u^* - u^{\mathsf{ps}}\| + \|x^*(u^*) - x^*(u^{\mathsf{ps}})\| \leq (1 + L_{\mathsf{eq}})L_{\mathsf{z}}L_{\mathsf{eq}}/(\alpha - L_\ell L_{\mathsf{eq}}).$$

Depending on the problem parameters, running the naïve repeated gradient method can lead to suboptimal Stackelberg equilibrium. We provide more detail in Appendix H, but the gap may be small. Moreover, as it turns out, the sample complexity of finding the true Stackelberg equilibrium (in low information environments with unknown agent preferences) is significantly worse. Therefore, it is interesting to examine regimes in which the performative gap is relatively small and obtaining a true Stackelberg is *not* worth the extra sample complexity (cf. Figure 1).

Algorithm 1 Epoch-Based Drift-to-Noise Control

Input: `Alg`, T, η_t, initial parameter $x_1 \in \mathcal{X}$, query radius $\delta > 0$ (if `Alg` = `DFM`)
for $s = 1, \dots, T$ **do**
  if `Alg` = `DFM` **then**
    set $\tilde{u}_t = u_t + \delta v_t$ where $v_t \sim \mathbb{S}^d$;
  end if
  if `Alg` = `RGM` **then**
    set $\tilde{u}_t = u_t$
  end if
  for $k = 1, \dots, \tau_t$ **do**
    query agents with $\tilde{u}_t$
    observe $z_t = \mathcal{A}(x_{t-1}, \tilde{u}_t) + \xi$ where $\xi \sim \mathcal{D}_o$
  end for
  if `Alg` = `DFM` **then**
    set $g_t = \frac{d}{\delta}\ell(\tilde{u}_t, z_t)v_t$ and $\widetilde{\mathcal{U}} = (1-\delta)\mathcal{U}$;
  end if
  if `Alg` = `RGM` **then**
    set $g_t = \nabla_u \ell(\tilde{u}_t, z_t)$ and $\widetilde{\mathcal{U}} = \mathcal{U}$;
  end if
  update $u_{t+1} = \mathrm{proj}_{\widetilde{\mathcal{U}}}(x_t - \eta_t \widehat{g}_t)$
end for

4.1.2 Stochastic Repeated Gradient Method

For a decision-maker with loss ℓ, the repeated stochastic gradient method is given by

$$u_{t+1} = \underset{\mathcal{U}}{\mathrm{proj}}(u_t - \eta g_t), \tag{RGM}$$

where $g_t := \nabla_u \ell(u_t, z_t)$ is the gradient of ℓ with respect to the u argument, and z_{t+1} is the stochastic observation of the agents response given the algorithms they are using employ a ρ-contracting stochastic update as defined in Section 2 and detailed in Appendix F.

Assumption 4 (Finite Variance)**.** Suppose there exists a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ such that $\mathcal{F}_0 = \{\emptyset, \Omega\}$, g_t is $\mathcal{F}_{t+1}$-measurable, and there exists a constant $\sigma > 0$ satisfying $\mathbb{E}_t\|g_t - \mathbb{E}_t[g_t]\|^2 \leq \sigma^2$ where $\mathbb{E}_t = \mathbb{E}[\cdot|\ \mathcal{F}_t]$ denotes the conditional expectation.

Recall from Section 2.4 that when the decision-maker deploys a sequence of actions $\{u_t\}$, it produces a drift in the agents equilibrium problem. In order to control that drift, we need to have a bound on the equilibrium tracking error for each deployed u_t. To this end, suppose the agents run a ρ-contracting algorithm $\mathcal{A}$ with $\rho \in [0, 1)$ and $\sigma_{\mathsf{a}} \in [0, \infty)$, and that the decision-maker fixes its action u_t in round t for τ time-steps. The agents response is them $\mathcal{A}(x_{t-1}, u_t)$—i.e., the agents run $\mathcal{A}$ for τ time-steps where they initialize at x_{t-1}. There are two scenarios we consider:

1. **Deterministic agents with stochastic observations**: agents employ ρ-contracting deterministic learning algorithms $\mathcal{A}$ where $\sigma_a = 0$, and the decision-maker observes a noisy sample of this response—i.e., $\mathcal{A}(x, u) + \xi$, where $\xi \sim \mathcal{D}_o$ for some distribution $\mathcal{D}_o$;
2. **Stochastic agents**: agents employ stochastic gradient play—possibly asynchronous—and the decision makers observes a noisy sample of the response.

The first setting avoids the constant bias that shows up when agents employ a stochastic gradient method (cf. Appendix F), and the second setting is controlled via the type of algorithm the agents are employing—see, e.g., Corollary 5 which proves stage-based stochastic gradient play hits a specified target accuracy in a certain number of iterations.[3] The following theorem provides a one-step contraction.

[3] Asynchronous stochastic gradient play also converges to a given target accuracy with a slight modification of constants—see Appendix F for more detail.

Theorem 3 (Informal)**.** Suppose that Assumptions 1, 3, and 4 hold, that we are in the regime where $1 < \frac{\alpha}{L_{\mathtt{eq}} L_\ell}$, that $\sup_{(u,x)\in\mathcal{U}\times\mathcal{X}} \mathbb{E}[\|\nabla_u \ell(u, x+\xi)\|] \le L_{\mathtt{u}}$, and that we have available a constant $R > \|x_0 - x^*(u_0)\|$. Set $C := \rho\|x_0 - x^*(u_0)\| + \eta L_{\mathtt{eq}} L_{\mathtt{u}}/(1-\rho)$. Further, suppose the decision-maker runs Algorithm 1 with $\mathtt{Alg} := \mathtt{RGM}$ using step-size $\eta \le \alpha/(4L_\ell^2(1+L_{\mathtt{eq}}^2))$, and the agents employ a ρ–contracting algorithm $\mathcal{A}$ with $\rho \in [0,1)$. Consider either of the following settings:

1. If $\sigma_{\mathtt{a}} = 0$, set the epoch length $\tau \gtrsim \Omega\left(\log\left(\frac{L_\ell R}{\sqrt{\alpha\eta}\sigma}\right)/\log(\frac{1}{\rho})\right)$ and agent tolerance $\epsilon_\tau = C\rho^\tau$.
2. If $\sigma_{\mathtt{a}} \in (0,\infty)$ and $\mathcal{A} = \mathtt{SGP}$, set $\tau \asymp \mathcal{O}\left(\frac{L_{\mathtt{a}}^2}{\mu^2}\log\left(\frac{2R^2}{\epsilon_\tau}\right) + \frac{\sigma_{\mathtt{a}}^2}{\mu^2\epsilon_\tau}\right)$ where $\epsilon_\tau \asymp \eta^2\sigma^2$.

Then the estimate holds:

$$\mathbb{E}\|u_{t+1} - u^{\mathtt{ps}}\|^2 \le \left(1 - \frac{\alpha\eta}{2}\right)^t \|u_0 - u^{\mathtt{ps}}\|^2 + \frac{4\eta\sigma^2}{\alpha}.$$

The formal statement and proof are in Appendix H. Given this t-step bound, the decision-maker can then employ a standard stage-based method to obtain convergence to an approximate performatively stable Stackelberg equilibrium. The corollary follows from employing Algorithm 2 with $\mathcal{A} = \mathtt{RGM}$.

Corollary 3. Under the assumptions of Theorem 3, consider running the stochastic repeated gradient method in $k = 0, \ldots, K$ super-epochs, for T_k epochs each with constant step-size $\eta_k = 2^{-k}\eta_0$, and such that the last iterate of each epoch k is used as the first iterate in stage $k+1$. Fix a target accuracy $\varepsilon > 0$ and suppose the decision-maker has available $R \ge \|u_0 - u^{\mathtt{ps}}\|$. Set $\eta_0 := \frac{\alpha}{4L_\ell^2(1+L_{\mathtt{eq}}^2)}$, and

$$T_0 = \left\lceil \frac{2}{\alpha\eta_0}\log(\frac{2R^2}{\varepsilon})\right\rceil, \quad T_k = \left\lceil \frac{2\log(4)}{\alpha\eta_k}\right\rceil \quad \text{for } k \ge 1, \quad \text{and } K = \left\lceil 1 + \log_2\left(\frac{\sigma^2}{L_\ell^2(1+L_{\mathtt{eq}}^2)\varepsilon}\right)\right\rceil.$$

Then

$$\mathbb{E}\,\|u_T - u^{\mathtt{ps}}\|^2 \le \varepsilon \quad \text{and} \quad \mathbb{E}\,\|x_T - x^*(u^{\mathtt{ps}})\|^2 \le 2(\epsilon_\tau + L_{\mathtt{eq}}\varepsilon)$$

in a total number of epochs

$$T = \sum_{k=1}^K T_k \quad \text{that is at most} \quad \mathcal{O}\left(\frac{L_\ell^2(1+L_{\mathtt{eq}}^2)}{\alpha^2}\log\left(\frac{2R^2}{\varepsilon}\right) + \frac{\sigma^2}{\alpha^2\varepsilon}\right).$$

Note that since we run τ iterations within each epoch, the total number of iterations is $T \cdot \tau$. Analogous to Proposition 2, we progressively decrease the step-size to control the bias until the desired accuracy is achieved. The choice of epoch length in each case is selected to control the bias introduced—namely, $\mathbb{E}\,\|\mathcal{A}(x_t, u_t) - x^*(u_t)\|^2$. In the deterministic case, the only bias term is directly controllable by τ since it is simply $\mathbb{E}\,\|\mathcal{A}(x_t, u_t) - x^*(u_t)\|^2 \le C^2\rho^{2\tau}$. On the other hand, in the stochastic case, we have $\mathbb{E}\,\|x_t^{k+1} - x^*(u_t)\|^2 \le \rho^2\,\mathbb{E}\,\|x_t^k - x^*(u_t)\|^2 + \rho^2(c\sigma_{\mathtt{a}})^2$ so that the is a constant bias term due to the noise $\sigma_{\mathtt{a}}$; this requires having the agents deploy stochastic gradient play as in Corollary 5 (Appendix F). Thus, the epoch length is select to ensure that this algorithm hits the requisite target accuracy.

4.2 Strategic Decision-Maker

Finally, we consider a *strategic* decision-maker, who knows that the agents are responding to u_t, yet still does not know the objectives of the agents, and so employs a derivative-free method to estimate the full gradient of its loss function

$$\nabla\mathcal{L}(u) = \mathop{\mathbb{E}}_{z\sim\mathcal{D}(u)} \nabla_u \ell(u,z) + \frac{d}{dv}\mathop{\mathbb{E}}_{z\sim\mathcal{D}(v)} \ell(u,z)|_{v=u}.$$

The derivative free method is defined as follows (see, e.g., (Agarwal et al., 2010)). The decision-maker fixes a $\delta > 0$, samples $\xi \sim \mathcal{D}_o$ along with a uniformly sampled vector $v_t \sim \mathbb{S}^d$, and computes the estimate

$$g_t = \frac{d}{\delta}\ell(u_t + \delta v_t, \mathcal{A}(x_t, u_t + \delta v_t) + \xi)v_t.$$

Then, the decision-maker updates according to Algorithm 1 where $\texttt{Alg} = \texttt{DFM}$—i.e.,

$$u_{t+1} = \underset{(1-\delta)\mathcal{U}}{\text{proj}} (u_t - \eta_t g_t), \tag{DFM}$$

The estimate g_t is a one-point gradient estimate of the expected loss when the agents are at $\mathcal{A}(x_t, u_t+\delta v_t)$. We use a one-point gradient estimate since the agents are strategically responding and it may not be possible to query the same population of agents multiple times before they adapt their play. The challenge here compared to classical analysis of derivative-free methods is accounting for the bias introduced by the agents' *drift* and *noise*. To start, we invoke the following assumptions.

Assumption 5. The following hold:

a. The loss $|\ell(u)| < \infty$ and we set $\ell_* = \sup_{u\in\mathcal{U}, z\in\mathcal{Z}} |\ell(u,z)|$;

b. The map $u \mapsto \nabla^2\mathcal{L}(u)$ is $L_{\texttt{H}}$-Lipschitz continuous;

c. The expected loss $\mathcal{L}(u)$ is $\bar{\alpha}$–strongly convex;

d. There exists $b, B > 0$ satisfying $b\mathbb{B} \subseteq \mathcal{U} \subseteq B\mathbb{B}$ where $\mathbb{B} = \{u \in \mathbb{R}^d | \ \|u\| \leq 1\}$ is the unit ball centered at the origin.

Assumption 5.*d* is common; see, e.g., (Agarwal et al., 2010). What this part of the assumption is implying is that the convex set $\mathcal{U}$ is compact and has a non-empty interior since otherwise we can map $\mathcal{U}$ to a lower dimensional space. The apparently strong assumption is Assumption 5.*c*.

In the performative prediction literature (Miller et al., 2021; Perdomo et al., 2020; Ray et al., 2022), it is common to assume that the probability measure $\mathcal{D}(u)$ and the loss ℓ satisfy *mixture dominance*—i.e., for any $u \in \mathcal{U}$ and $s \in (0,1)$, we have that

$$\mathop{\mathbb{E}}_{z\sim\mathcal{D}(sv+(1-s)w)} \ell(u,z) \leq \mathop{\mathbb{E}}_{z\sim s\mathcal{D}(v)+(1-s)\mathcal{D}(w)} \ell(u,z) \quad \forall\, v, w \in U.$$

Further, it is assumed that $W_1(\mathcal{D}(u), \mathcal{D}(u')) \leq \beta\|u - u'\|$ for all $u, u' \in \mathcal{U}$. This is tântamount to $x^*(\cdot)$ being $L_{\texttt{eq}}$–Lipschitz continuous in our setting. Given that mixture dominance holds, under Assumptions 3 and 5.*a*, then it can be shown that $\mathcal{L}(u)$ is $\bar{\alpha} := (\alpha - 2\gamma L_\ell)$ strongly convex (cf. Theorem 3.1 (Miller et al., 2021)).

In our setting this means that the composition of the agents' response and the loss ℓ is strongly convex. We already have Assumption 3, which states that the loss ℓ is α strongly convex in u for any z. Hence, this the loss ℓ is strongly convex if $\ell(u,\cdot)$ is convex-nondecreasing and $x^*(\cdot)$ is convex, or $\ell(u,\cdot)$ is convex-nonincreasing and $x^*(\cdot)$ is concave (Boyd et al., 2004). Though similar assumptions have been made in the single agent setting (Dong et al., 2018), this admittedly restricts the set of equilibrium for which our results hold. However, convexity is allows for convergence to the *global optimum*; similar statements for local convergence hold without this assumption as do statements on convergence to *stationary points*—i.e. optimality only to first order with no curvature guarantees.

Analogous to the naïve setting, we consider both deterministic agents with noisy observations, and stochastic agents. Let u^* be the decision-maker's action in the Stackelberg equilibrium (i.e. optimal for $\mathcal{L}$) over $\mathcal{U}$ subject to $x^*(u) \in \texttt{Eq}(\mathcal{G}_u)$.

Theorem 4 (Informal)**.** Suppose that Assumptions 1, 3, and 5 hold, and that we have available a constant $R > \|x_0 - x^*(u_0)\|$. Further, suppose that we are in the regime where $\alpha > 2\gamma L_\ell$ and the decision-maker runs Algorithm 1 with $\texttt{Alg} := \texttt{DFM}$ using step-size $\eta_t = \frac{4}{\bar{\alpha}(t+1)}$, query radius $\delta < \min\{b, \frac{\bar{\alpha}}{L_{\texttt{H}}}\}$, and the agents employ a ρ–contracting algorithm $\mathcal{A}$ with $\rho \in [0,1)$. Set $L := L_\ell(1 + L_{\texttt{eq}})$. Consider either of the following settings:

1. If $\sigma_{\texttt{a}} = 0$, set the epoch length $\tau_t \gtrsim \Omega\left(\log\left(\frac{2\delta L_\ell R}{\sqrt{\eta_t}\bar{\alpha}\ell_* d}\right)\frac{1}{\log(1/\rho)}\right)$, constant $c = 32\ell_*^2 d^2$ and tolerance $\epsilon_t = C\rho^{\tau_t}$.
2. If $\sigma_{\texttt{a}} \in (0,\infty)$ and $\mathcal{A} = \texttt{SGP}$, set $\tau_t \asymp \mathcal{O}\left(\frac{L_{\texttt{a}}^2}{\mu^2}\log\left(\frac{2R^2}{\epsilon_t}\right) + \frac{\sigma_{\texttt{a}}^2}{\mu^2\epsilon_t}\right)$ where $\epsilon_t \asymp (\delta^2(t+1))^{-1}$ and constant $c = 16(\ell_*^2 d^2 + 1)$.

Then, the estimate holds:

$$\mathbb{E}\,\|u_t - u^*\|^2 \le \frac{\max\{2\bar{\alpha}^2\delta^2\|u_0 - u^*\|^2, c\}}{\delta^2\bar{\alpha}^2(t+1)} + 2\delta^2\left(\left(1+\frac{L}{\bar{\alpha}}\right)\|u^*\| + \frac{L}{\bar{\alpha}}\right).$$

The proof along with a more formal statement is contained in Appendix I. In the following corollary (the proof of which is contained in Appendix I), we show that given a fixed error tolerance ε, if the decision-maker runs Algorithm 1 with $\texttt{Alg} = \texttt{DFM}$ according to the preceding theorem, then we can characterize the iteration complexity to reach an approaximate Stackelberg equilibrium.

Corollary 4. Suppose the assumptions of Theorem 6 hold. Fix target accuracy $\varepsilon < 4b^2\left(\left(1+\frac{L}{\bar{\alpha}}\right)B + \frac{L}{\bar{\alpha}}\right)^2$, and set $\delta = \bar{\alpha}\sqrt{\varepsilon/4}/((\bar{\alpha}+L)B+L)$ and $\eta_t = 4/(\bar{\alpha}(t+1))$. The iterates (u_t, x_t) converge to an approximate Stackelberg equilibrium $(u^*, x^*(u^*))$—i.e., the estimates $\mathbb{E}[\|u_t - u^*\|^2] \le \varepsilon$ and $\mathbb{E}[\|x_t - x^*(u^*)\|^2] \le 2(\epsilon_0 + L_{\mathsf{eq}}\varepsilon)$ hold for all

$$t \ge \frac{\max\{16\bar{\alpha}^4\varepsilon B^2, 8c((\bar{\alpha}+L)B+L)^2\}}{\bar{\alpha}^4\varepsilon^2}.$$

Observe that ε can be selected arbitrarily small to control the agents locality relative to the equilibrium. This corollary provides a bound in terms of the number of epochs; the total iteration complexity is $\sum_{s=1}^t \tau_s = \mathcal{O}(\frac{d^2}{\varepsilon^2}\log(\frac{1}{\varepsilon}))$ as shown in Appendix I. This rate matches the rate of single point derivative-free convex optimization (Agarwal et al., 2010) up to log factors, where the log factor is precisely due to the lower level agents running their algorithms for τ_t time-steps. The rate for the derivative free method is decidedly worse than for the repeated gradient method, as expected, owing to the extra estimator bias. However, the repeated gradient method converges to a suboptimal Stackelberg equilibrium—we explore this numerically in Appendix B. Reflecting back on Figure 1, an interesting future direction is to fully understand how performativity affects the sample complexity and performance tradeoff and therefore, practical algorithm selection. Derivative free methods can be prohibitively expensive in practice and hence, deciding when to forego such a method would be of use.

5 Discussion

We consider a novel class of stochastic Stackelberg games, where updates from the decision-maker and the agents induces a time-varying game for both parties. We present finite-time efficiency estimates that are governed by the drift-to-noise ratio for the agents' updates for settings where the decision-maker sequentially deploys actions. We also identify two epoch-based algorithms that find two different notions of equilibria, the performatively stable Stackelberg equilibrium and the true Stackelberg equilibrium. We characterize the existence of the former equilibrium, and establish convergence rates. Illustrative numerical examples explore the theoretical assumptions and suggest many interesting directions for future work.

Indeed, the results motivate future work that captures the interplay between game theory, optimization, and learning. Better characterizing the tradeoffs in the performative gap, both in $\mathcal{U}$ as well as in cost, and in sample complexity is essential across a number of performative prediction settings. Additionally, having a better characterization the extent of performativity exists in a stochastic optimization system would enable decision-makers to determine which algorithmic approach (i.e., computationally expensive derivative free methods versus sub-optimal repeated stochastic methods) is beneficial given the reactivity of its user base.

Our theoretical results also depend on a number intrinsic parameters such as Lipschtiz constants which may not be readily available in practice. This suggests developing adaptive algorithms for learning in game theoretic settings such as the ones explored in this paper. Another interesting direction is to consider additional exogenous dynamic sources and to similar characterize the drift-to-noise regimes where drift may be due to either this exogenous source or endogenous learning process (the latter is what we have examined in this paper).

Acknowledgement

This work is supported by NSF grants 2019844 and 2112471. Ratliff and Frankel are supported in part by NSF Award 1844729. Ratliff is also supported in part by ONR Award N000142012571. Kulkarni is supported by the NSF DToD Fellowship. Research of Drusvyatskiy was supported by the NSF DMS-2306322, NSF CCF 1740551, and AFOSR FA9550-24-1-0092 awards.

References

Ralph Abraham, Jerrold E Marsden, and Tudor Ratiu. *Manifolds, tensor analysis, and applications*, volume 75. Springer Science & Business Media, 2012.

Alekh Agarwal, Ofer Dekel, and Lin Xiao. Optimal algorithms for online convex optimization with multipoint bandit feedback. In *Colt*, pages 28–40. Citeseer, 2010.

Tamer Başar and Geert Jan Olsder. *Dynamic noncooperative game theory*. SIAM, 1998.

Dimitris Bertsimas, Vishal Gupta, and Ioannis Ch Paschalidis. Data-driven estimation in equilibrium using inverse optimization. *Mathematical Programming*, 153:595–633, 2015.

Vivek S Borkar. *Stochastic approximation: a dynamical systems viewpoint*, volume 48. Springer, 2009.

Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Jerome Bracken and James T McGill. Mathematical programs with optimization problems in the constraints. *Operations Research*, 21(1):37–44, 1973.

Gavin Brown, Shlomi Hod, and Iden Kalemaj. Performative prediction in a stateful world. In *International Conference on Artificial Intelligence and Statistics*, pages 6045–6061. PMLR, 2022.

Yang Cai, Constantinos Daskalakis, and Christos Papadimitriou. Optimum statistical estimation with strategic data sources. In *Conference on Learning Theory*, pages 280–296. PMLR, 2015.

Emilio Calvano and Michele Polo. Market power, competition and innovation in digital markets: A survey. *Information Economics and Policy*, 54:100853, 2021.

Sarah H Cen, Andrew Ilyas, Jennifer Allen, Hannah Li, and Aleksander Madry. Measuring strategization in recommendation: Users adapt their behavior to shape future content. *arXiv preprint arXiv:2405.05596*, 2024.

Benjamin Chasnov, Lillian Ratliff, Eric Mazumdar, and Samuel Burden. Convergence analysis of gradient-based learning in continuous games. In *Uncertainty in artificial intelligence*, pages 935–944. PMLR, 2020a.

Benjamin Chasnov, Lillian J Ratliff, Eric Mazumdar, and Samuel A Burden. Convergence analysis of gradient-based learning with non-uniform learning rates in non-cooperative multi-agent settings. *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2020b.

Benoît Colson, Patrice Marcotte, and Gilles Savard. An overview of bilevel optimization. *Annals of operations research*, 153(1):235–256, 2007.

Lauren Conger, Franca Hoffmann, Eric Mazumdar, and Lillian Ratliff. Coupled gradient flows for strategic non-local distribution shift. *arXiv preprint arXiv:2307.01166*, 2023.

Samuel Coogan, Lillian J Ratliff, Daniel Calderone, Claire Tomlin, and S Shankar Sastry. Energy management via pricing in lq dynamic games. In *2013 American Control Conference*, pages 443–448. IEEE, 2013.

Joshua Cutler, Dmitriy Drusvyatskiy, and Zaid Harchaoui. Stochastic optimization under time drift: iterate averaging, step decay, and high probability guarantees. *arXiv preprint arXiv:2108.07356*, 2021.

Joshua Cutler, Mateo Díaz, and Dmitriy Drusvyatskiy. Stochastic approximation with decision-dependent distributions: asymptotic normality and optimality. *arXiv preprint arXiv:2207.04173*, 2022.

Joshua Cutler, Dmitriy Drusvyatskiy, and Zaid Harchaoui. Stochastic optimization under distributional drift. *Journal of Machine Learning Research*, 24(147):1–56, 2023.

Joshua Cutler, Mateo Díaz, and Dmitriy Drusvyatskiy. Stochastic approximation with decision-dependent distributions: asymptotic normality and optimality. *Journal of Machine Learning Research*, 25(90):1–49, 2024.

Venkat Surya Dasari, Burak Kantarci, Maryam Pouryazdan, Luca Foschini, and Michele Girolami. Game theory in mobile crowdsensing: A comprehensive survey. *Sensors*, 20(7):2055, 2020.

Sarah Dean, Mihaela Curmei, Lillian Ratliff, Jamie Morgenstern, and Maryam Fazel. Emergent specialization from participation dynamics and multi-learner retraining. In *International Conference on Artificial Intelligence and Statistics*, pages 343–351. PMLR, 2024.

Stephan Dempe and Alain Zemkoho. Bilevel optimization. In *Springer optimization and its applications. Vol. 161*. Springer, 2020.

Jinshuo Dong, Aaron Roth, Zachary Schutzman, Bo Waggoner, and Zhiwei Steven Wu. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 55–70, 2018.

Asen L Dontchev, R Tyrrell Rockafellar, and R Tyrrell Rockafellar. *Implicit functions and solution mappings: A view from variational analysis*, volume 616. Springer, 2009.

Salvatore d'Oro, Panayotis Mertikopoulos, Aris L Moustakas, and Sergio Palazzo. Interference-based pricing for opportunistic multicarrier cognitive radio systems. *IEEE Transactions on Wireless Communications*, 14(12):6536–6549, 2015.

Dmitriy Drusvyatskiy and Lin Xiao. Stochastic optimization with decision-dependent distributions. *Mathematics of Operations Research*, 48(2):954–998, 2023.

Benoit Duvocelle, Panayotis Mertikopoulos, Mathias Staudigl, and Dries Vermeulen. Multiagent online learning in time-varying games. *Mathematics of Operations Research*, 48(2):914–941, 2023.

Francisco Facchinei and Christian Kanzow. Generalized nash equilibrium problems. *4or*, 5(3):173–210, 2007.

Tanner Fiez, Benjamin Chasnov, and Lillian Ratliff. Implicit learning dynamics in stackelberg games: Equilibria characterization, convergence analysis, and empirical study. In *International Conference on Machine Learning*, pages 3133–3144. PMLR, 2020.

Tanner Fiez, Chi Jin, Praneeth Netrapalli, and Lillian J Ratliff. Minimax optimization with smooth algorithmic adversaries. *arXiv preprint arXiv:2106.01488*, 2021.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM conference on innovations in theoretical computer science*, pages 111–122, 2016.

Moritz Hardt, Meena Jagadeesan, and Celestine Mendler-Dünner. Performative power. *Advances in Neural Information Processing Systems*, 35:22969–22981, 2022.

Nicholas JA Harvey, Christopher Liaw, Yaniv Plan, and Sikander Randhawa. Tight analyses for non-smooth stochastic gradient descent. In *Conference on Learning Theory*, pages 1579–1613. PMLR, 2019.

Lars Hellemo, Paul I Barton, and Asgeir Tomasgard. Decision-dependent probabilities in stochastic programs with recourse. *Computational Management Science*, 15:369–395, 2018.

Zehong Hu, Yitao Liang, Jie Zhang, Zhao Li, and Yang Liu. Inference aided reinforcement learning for incentive mechanism design in crowdsourcing. *Advances in Neural Information Processing Systems*, 31, 2018.

Zhouyuan Huo and Heng Huang. Asynchronous mini-batch gradient descent with variance reduction for non-convex optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31 (1), 2017.

Chi Jin, Praneeth Netrapalli, Rong Ge, Sham M. Kakade, and Michael I. Jordan. A short note on concentration inequalities for random vectors with subgaussian norm, 2019.

Michael P Kim and Juan C Perdomo. Making decisions under outcome performativity. In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2023.

Andrei Kulunchakov and Julien Mairal. A generic acceleration framework for stochastic composite optimization. *Advances in Neural Information Processing Systems*, 32, 2019.

HJ Kusher and GG Yin. Stochastic approximation algorithms and applications, 1997.

Jiayang Li, Jing Yu, Boyi Liu, Yu Nie, and Zhaoran Wang. Achieving hierarchy-free approximation for bilevel programs with equilibrium constraints. In *International Conference on Machine Learning*, pages 20312–20335. PMLR, 2023.

Tao Li and Suresh P. Sethi. A review of dynamic stackelberg game models. *Discrete and Continuous Dynamical Systems-series B*, 22:125–159, 2016. URL https://api.semanticscholar.org/CorpusID:126323935.

Xiangru Lian, Yijun Huang, Yuncheng Li, and Ji Liu. Asynchronous parallel stochastic gradient for nonconvex optimization. *Advances in neural information processing systems*, 28, 2015.

Chinmay Maheshwari, Chih-Yuan Chiu, Eric Mazumdar, Shankar Sastry, and Lillian Ratliff. Zeroth-order methods for convex-concave min-max problems: Applications to decision-dependent risk minimization. In *International Conference on Artificial Intelligence and Statistics*, pages 6702–6734. PMLR, 2022.

Chinmay Maheshwari, S Shankar Sasty, Lillian Ratliff, and Eric Mazumdar. Convergent first-order methods for bi-level optimization and stackelberg games. *arXiv preprint arXiv:2302.01421*, 2023.

Celestine Mendler-Dünner, Juan Perdomo, Tijana Zrnic, and Moritz Hardt. Stochastic optimization for performative prediction. *Advances in Neural Information Processing Systems*, 33:4929–4939, 2020.

John P Miller, Juan C Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. In *International Conference on Machine Learning*, pages 7710–7720. PMLR, 2021.

Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach. In *International Conference on Artificial Intelligence and Statistics*, pages 1497–1507. PMLR, 2020.

Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996.

Kamil Nar, Lillian J Ratliff, and Shankar Sastry. Learning prospect theory value function and reference point of a sequential decision maker. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 5770–5775. IEEE, 2017.

Adhyyan Narang, Evan Faulkner, Dmitriy Drusvyatskiy, Maryam Fazel, and Lillian J Ratliff. Multiplayer performative prediction: Learning in decision-dependent games. *arXiv preprint arXiv:2201.03398*, 2022.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *International Conference on Machine Learning*, pages 7599–7609. PMLR, 2020.

Lillian J Ratliff and Tanner Fiez. Adaptive incentive design. *IEEE Transactions on Automatic Control*, 66 (8):3871–3878, 2020.

Lillian J Ratliff, Samuel Coogan, Daniel Calderone, and S Shankar Sastry. Pricing in linear-quadratic dynamic games. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1798–1805. IEEE, 2012.

Mitas Ray, Lillian J Ratliff, Dmitriy Drusvyatskiy, and Maryam Fazel. Decision-dependent risk minimization in geometrically decaying dynamic environments. 2022.

Benjamin Recht, Christopher Re, Stephen Wright, and Feng Niu. Hogwild!: A lock-free approach to parallelizing stochastic gradient descent. *Advances in neural information processing systems*, 24, 2011.

R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

R.T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

J Ben Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society*, pages 520–534, 1965.

Ernest K Ryu and Stephen Boyd. Primer on monotone operator methods. *Appl. comput. math*, 15(1):3–43, 2016.

Nihar B Shah and Dengyong Zhou. Double or nothing: Multiplicative incentive mechanisms for crowdsourcing. *Journal of Machine Learning Research*, 17(165):1–52, 2016.

Hanif D Sherali, Allen L Soyster, and Frederic H Murphy. Stackelberg-nash-cournot equilibria: characterizations and computations. *Operations Research*, 31(2):253–276, 1983.

Heinrich von Stackelberg et al. Theory of the market economy. 1952.

David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

Heinrich Von Stackelberg. *Market structure and equilibrium*. Springer Science & Business Media, 2010.

John Glen Wardrop. Road paper. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1(3):325–362, 1952.

Tyler Westenbroek, Roy Dong, Lillian J Ratliff, and S Shankar Sastry. Competitive statistical estimation with strategic data sources. *IEEE Transactions on Automatic Control*, 65(4):1537–1551, 2019.

Killian Wood and Emiliano Dall'Anese. Stochastic saddle point problems with decision-dependent distributions. *SIAM J. Optim.*, 33:1943–1967, 2022. URL https://api.semanticscholar.org/CorpusID:245827836.

Killian Reed Wood and Emiliano Dall'Anese. Online saddle point tracking with decision-dependent data. In *Learning for Dynamics and Control Conference*, pages 1416–1428. PMLR, 2023.

Hong Xie, John CS Lui, Joe Wenjie Jiang, and Wei Chen. Incentive mechanism and protocol design for crowdsourcing systems. In *2014 52nd annual Allerton conference on communication, control, and computing (Allerton)*, pages 140–147. IEEE, 2014.

Jiachen Yang, Ang Li, Mehrdad Farajtabar, Peter Sunehag, Edward Hughes, and Hongyuan Zha. Learning to incentivize other learning agents. *Advances in Neural Information Processing Systems*, 33:15208–15219, 2020.

Jiachen Yang, Ethan Wang, Rakshit Trivedi, Tuo Zhao, and Hongyuan Zha. Adaptive incentive design with multi-agent meta-gradient reinforcement learning. *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, 2022.

Guodong Zhang and Yuanhao Wang. On the suboptimality of negative momentum for minimax optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 2098–2106. PMLR, 2021.

Zhengyuan Zhou, Panayotis Mertikopoulos, Susan Athey, Nicholas Bambos, Peter Glynn, and Yinyu Ye. Learning in games with lossy feedback. In *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS)*, pages 1–11, 2018.

Appendix Organization

The appendix contains many sections, so here we provide a contents list to help the reader navigate the material.

A Examples of Monotone Games

In this section, we provide several examples of strongly monotone games.

A.1 Quadratic Games

Consider the game defined by costs

$$f_i(x_i, x_{-i}) = \frac{1}{2}\begin{bmatrix} x_i \\ x_{-i} \end{bmatrix}^\top \begin{bmatrix} A_i & B_i^\top \\ B_i & D_i \end{bmatrix} \begin{bmatrix} x_i \\ x_{-i} \end{bmatrix} + \begin{bmatrix} a_i \\ b_i \end{bmatrix}^\top \begin{bmatrix} x_i \\ x_{-i} \end{bmatrix}, \tag{6}$$

where $A_i \in \mathbb{R}^{d_i \times d_i}$, $D_i \in \mathbb{R}^{d_{-i} \times d_{-i}}$, $B_i \in \mathbb{R}^{d_{-i} \times d_i}$, $a_i \in \mathbb{R}^{d_i}$ and $b_i \in \mathbb{R}^{d_{-i}}$ with $A_i = A_i^\top$ and $D_i = D_i^\top$. Further, we assume that $A_i \succ 0$ for each $i = 1, 2$. The D_i matrices penalize player i based solely on x_{-i} and may often be negative or zero. Quadratic games are a useful approximation of the behavior of more complex games around an equilibrium. This game is strongly monotone if there exists $\mu \in (0, \infty)$ such that

$$\langle \omega(x) - \omega(x'), x - x' \rangle \geq \mu \|x - x'\|^2,$$

where

$$\omega(x) = (\nabla_1 f_1(x), \ldots, \nabla_n f_n(x)) \quad \text{with} \quad \nabla_i f_i(x_i, x_{-i}) = A_i x_i + B_i^\top x_{-i} + a_i.$$

A sufficient condition for strong monotonicity is

$$J(x) = \nabla\omega(x) = \begin{bmatrix} A_1 & B_{12} & \cdots & B_{1n} \\ B_{21} & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & B_{(n-1)n} \\ B_{n1} & \cdots & B_{n(n-1)} & A_n \end{bmatrix} \succ 0, \quad \text{where} \quad B_i = \begin{bmatrix} B_{i1} \\ \vdots \\ B_{i(i-1)} \\ B_{i(i+1)} \\ \vdots \\ B_{in} \end{bmatrix}.$$

There are many important examples of quadratic games in economics and control theory. Below we highlight a few.

A.1.1 Open Loop Linear Quadratic Dynamic Games

One important class in control theory is that of linear quadratic dynamic games in open loop strategies. For simplicity, we write out the details for a $n = 2$ player game; however, these derivations easily extend to arbitrary but finite n. To that end, consider a two player linear quadratic dynamic game with open loop policies $\mathbf{v}_i = (v_{i,0}, \ldots, v_{i,T-1})$ with costs

$$\begin{aligned} f_i(\mathbf{v}_1, \mathbf{v}_2) &= \sum_{t=0}^{T-1} \frac{1}{2} z_t^\top Q_i z_t + \frac{1}{2} v_{i,t}^\top R_i v_{i,t} + v_{i,t}^\top R_{i,-i} v_{-i,t} + \frac{1}{2} z_T^\top Q_{i,f} z_T \\ z_{t+1} &= F z_t + G_1 v_{1,t} + G_2 v_{2,t}, \ \ z_t \in \mathbb{R}^m. \end{aligned}$$

Unfolding the dynamics and letting $Z = [z_0^\top, \ldots, z_T^\top]^\top$, we have that $Z = W_1 \mathbf{v}_1 + W_2 \mathbf{v}_2 + \mathbf{F} z_0$ where

$$W_i = \begin{bmatrix} 0 & \cdots & & & 0 \\ G_i & 0 & \cdots & & 0 \\ FG_i & G_i & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ F^{T-2}G_i & F^{T-3}G_i & \cdots & G_i & 0 \\ F^{T-1}G_i & F^{T-2}G_i & \cdots & FG_i & G_i \end{bmatrix}, \quad i = 1, 2,$$

and $\mathbf{F} = \begin{bmatrix} I & F^\top & \cdots & (F^{T-1})^\top & (F^T)^\top \end{bmatrix}^\top$. Define the following cost matrices:

$$\begin{aligned} \mathbf{Q}_i &:= \mathrm{diag}(Q_i, \ldots, Q_i, Q_{i,f}) \in \mathbb{R}^{m(T+1) \times m(T+1)}, \\ \mathbf{R}_i &:= \mathrm{diag}(R_i, \ldots, R_i) \in \mathbb{R}^{d_i T \times d_i T}, \\ \mathbf{R}_{i,-i} &:= \mathrm{diag}(R_{i,-i}, \ldots, R_{i,-i}) \in \mathbb{R}^{d_i T \times d_{-i} T}. \end{aligned}$$

Player i's cost is

$$f_i(\mathbf{v}_i, \mathbf{v}_{-i}) = \tfrac{1}{2} \mathbf{v}_i^\top \mathbf{R}_i \mathbf{v}_i + \mathbf{v}_i^\top \mathbf{R}_{i,-i} \mathbf{v}_{-i} + \tfrac{1}{2} (W_1 \mathbf{v}_1 + W_2 \mathbf{v}_2 + \mathbf{F} z_0)^\top \mathbf{Q}_i (W_1 \mathbf{v}_1 + W_2 \mathbf{v}_2 + \mathbf{F} z_0).$$

Mapping back to the original quadratic cost form in (6), we have that

$$\begin{aligned} &A_i = \mathbf{R}_i + W_i^\top \mathbf{Q}_i W_i, \quad B_i = \left(\mathbf{R}_{i,-i} + W_i^\top \mathbf{Q}_i W_{-i}\right)^\top, \quad D_i = W_{-i}^\top \mathbf{Q}_i W_{-i} \\ &a_i^\top = z_0^\top \mathbf{F}^\top \mathbf{Q}_i W_i, \quad b_i^\top = z_0^\top \mathbf{F}^\top \mathbf{Q}_i W_{-i}. \end{aligned}$$

The game Jacobian is given by

$$J(x) = \begin{bmatrix} A_1 & B_1^\top \\ B_2^\top & A_2 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1 + W_1^\top \mathbf{Q}_1 W_1 & \left(\mathbf{R}_{1,2} + W_1^\top \mathbf{Q}_1 W_2\right) \\ \left(\mathbf{R}_{2,1} + W_2^\top \mathbf{Q}_2 W_1\right) & \mathbf{R}_2 + W_2^\top \mathbf{Q}_2 W_2 \end{bmatrix}$$

In a typical linear quadratic regulator problem it is assumed that $\mathbf{R}_i \succ 0$ and $Q_i \succeq 0$ in order for solutions to exist (there are conditions that weaken these assumptions), and hence $A_i \succ 0$. In this case A_i is non-degenerate for $i = 1, 2$, and hence a sufficient condition for the game Jacobian to be positive definite is checking that either Schur complement is positive definite.

The goal of a decision-maker here might be to design 'pricing mechanisms' to influence the equilibrium; e.g., they may optimize over the matrices $(R_i, R_{i,-i})$ (Coogan et al., 2013; Ratliff et al., 2012).

A.1.2 Cournot and Bertrand Competition

Both Cournot and Bertrand oligopoly models are monotone games under certain conditions on the cost parameters. In the Cournot model, firms choose quantities in non-cooperative competition, and the market determines the price of each good. On the other hand, in a Bertrand competition, firms set prices, and the market determines its demand for each type of good.

To see that these games are both strongly monotone, consider a setting with n firms.

Cournot Competition. Each firm supplies the market with a quantity $x_i \in [0, B_i]$ of some good or service where $B_i > 0$ is firm i's capacity for production. The market determines the price $P(x)$ for good. This is a decreasing function of the total supply to the market $\sum_{i=1}^n x_i$. For example, a common model is a linear model of the form $P(x) = r - q\sum_{i=1}^n x_i$ for some positive constants r, q. The i–th firm aims to maximize their utility which is given by

$$U_i(x) = x_i P(x) - c_i x_i \quad \text{where} \;\; c_i \text{ is the marginal production cost.}$$

Hence, the first term $x_i P(x)$ is the revenue generated from selling x_i goods in the market and $c_i x_i$ is the cost of production of said goods. Consider the game with costs $f_i(x) = -U_i(x)$ over strategy spaces $\mathcal{X}_i = [0, B_i]$. Then, the game is strongly monotone if there exists $\mu > 0$ such that

$$\begin{aligned}
\langle \omega(x) - \omega(x'), x - x' \rangle &= \left\langle \begin{bmatrix} -x_1 \nabla_1 P(x) - P(x) + c_1 \\ \vdots \\ -x_n \nabla_n P(x) - P(x) + c_n \end{bmatrix} - \begin{bmatrix} -x_1' \nabla_1 P(x') - P(x') + c_1 \\ \vdots \\ -x_n' \nabla_n P(x') - P(x') + c_n \end{bmatrix}, x - x' \right\rangle \\
&= \left\langle \begin{bmatrix} -x_1 \nabla_1 P(x) + x_1' \nabla_1 P(x') - P(x) + P(x') \\ \vdots \\ -x_n \nabla_n P(x) + x_n' \nabla_n P(x') - P(x) + P(x') \end{bmatrix}, x - x' \right\rangle \\
&\geq \mu \|x - x'\|^2.
\end{aligned}$$

Using the linear form of $P(x)$ as an example, we have that

$$\begin{aligned}
\langle \omega(x) - \omega(x'), x - x' \rangle &= \left\langle \begin{bmatrix} -x_1 \nabla_1 P(x) + x_1' \nabla_1 P(x') - P(x) + P(x') \\ \vdots \\ -x_n \nabla_n P(x) + x_n' \nabla_n P(x') - P(x) + P(x') \end{bmatrix}, x - x' \right\rangle \\
&= \left\langle \begin{bmatrix} q(x_1 - x_1') + q(\sum_i x_i - \sum_i x_i') \\ \vdots \\ q(x_n - x_n') + q(\sum_i x_i - \sum_i x_i') \end{bmatrix}, x - x' \right\rangle
\end{aligned}$$

Hence, after rearranging, we have that

$$\begin{aligned}
\langle \omega(x) - \omega(x'), x - x' \rangle &= q\|x - x'\|^2 + \left\langle \begin{bmatrix} q(\sum_i x_i - \sum_i x_i') \\ \vdots \\ q(\sum_i x_i - \sum_i x_i') \end{bmatrix}, \begin{bmatrix} x_1 - x_1' \\ \vdots \\ x_n - x_n' \end{bmatrix} \right\rangle \\
&= q \left\langle \begin{bmatrix} x_1 - x_1' \\ \vdots \\ x_n - x_n' \end{bmatrix}, \begin{bmatrix} x_1 - x_1' \\ \vdots \\ x_n - x_n' \end{bmatrix} \right\rangle + q \left\langle \begin{bmatrix} \sum_i x_i - \sum_i x_i' \\ \vdots \\ \sum_i x_i - \sum_i x_i' \end{bmatrix}, \begin{bmatrix} x_1 - x_1' \\ \vdots \\ x_n - x_n' \end{bmatrix} \right\rangle \\
&= q \sum_i (x_i - x_i')^2 + q \sum_i \left(\sum_j (x_j - x_j')(x_i - x_i') \right) \\
&= q\|x - x'\|^2 + q \left(\sum_i (x_i - x_i') \right)^2 \\
&\geq q\|x - x'\|^2
\end{aligned}$$

so that the game is strongly monotone with $\mu = q > 0$. In a Cournot competition, the market reaches an equilibrium where all firms choose a quantity that is their *best response* to their competitors' quantities. This turns out to be an inefficient equilibrium, in that the equilibrium price is above the price in perfect competition and therefore firms earn a profit. A third party (such as a government entity) may intervene in the market by modulating the price $P(x)$ or by taxing individual firms (thereby increasing the cost of production $c_i x_i$) in order to move the market to an efficient equilibrium.

Bertrand Competition. On the other hand, in a Bertrand competition, where prices are the strategic variable, firms are incentivized to set their price slightly lower than the competition. Since all firms are so incentivized, they repeatedly drop the price until the price reaches the price in perfect competition wherein firms do not earn a profit. A third party may intervene in this market to improve uncertainties related to forecasting demand. For example, often times demand depends on exogenous time varying quantity or signal, such as gross domestic product, for which an individual firm may not have a good (low variance) forecaster.

Indeed, again consider n firms, but now where the strategies $x_i \in [0, B_i]$ are the prices. The firms seek to maximize their revenue in this setting which is given by $R_i(x_i, x_{-i}, u) = x_i F_i(x_i, x_{-i}, u)$ where F_i is the marginal revenue function or demand curve given prices $x = (x_i, x_{-i})$. Here u is some exogenous signal as described above. Then, for a fixed u, we have that

$$\langle \omega(x) - \omega(x'), x - x' \rangle = \langle -x_i \nabla_i F_i(x, u) - F_i(x, u) - (-x_i' \nabla_i F_i(x', u) - F_i(x', u), x - x' \rangle$$

so that, just like with the Cournot competition, if the marginal revenue is an affine function with parameters (r, q) then the game is strongly monotone with $\mu = q$. The marginal revenue function, however, does not have to be linear for the game to be strongly monotone. Indeed a common form for the marginal revenue includes logarithmic terms (Bertsimas et al., 2015; Ratliff and Fiez, 2020). For instance consider marginal revenue function given by

$$M_i(x, u) = \log(x_i) + \theta_i^\top x + \xi_i + u_i,$$

where (θ_i, ξ_i) are parameters. Therefore we have that

$$\langle \omega(x) - \omega(x'), x - x' \rangle = \left\langle \begin{bmatrix} -(\log(x_1) + 2\theta_{1,1}x_1 + \theta_{1,-1}x_{-1}) + (\log(x_1') + 2\theta_{1,1}x_1' + \theta_{1,-1}x_{-1}') \\ \vdots \\ -(\log(x_n) + 2\theta_{1,1}x_1 + \theta_{1,-1}x_{-1}) + (\log(x_n') + 2\theta_{n,n}x_n' + \theta_{n,-n}x_{-n}') \end{bmatrix}, x - x' \right\rangle.$$

A sufficient condition for strong monotonicity is that the game Jacobian is positive definite. The game Jacobian is given by

$$\nabla \omega(x, u) = \begin{bmatrix} -2\theta_{1,1} - \frac{1}{x_1} & -\theta_{1,2} & \cdots & -\theta_{1,n} \\ -\theta_{2,1} & -2\theta_{2,2} - \frac{1}{x_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ -\theta_{n,1} & \cdots & -\theta_{n,(n-1)} & -2\theta_{n,n} - \frac{1}{x_n} \end{bmatrix}$$

Here for the game Jacobian to be positive definite the prices have to be strictly positive, and there are constraints on the parameters θ. One interesting question is if there is a natural mechanism that a third party could use to shape the game in order to ensure it is positive definite. For example, if u_i was a linear tariff such as $a_i x_i + b_i$ then the diagonal terms of the Jacobian would be $-2\theta_{i,i} - 2a_i - \frac{1}{x_i}$ and the third party could design a_i to ensure monotonicity.

A.2 Strongly Convex Potential Game

A game $\mathcal{G} = (f_1, \ldots, f_n)$ is called a potential game (Monderer and Shapley, 1996) if there exists a potential function $\Phi : \mathcal{X} \to \mathbb{R}$ such that

$$f_i(x_i, x_{-i}) - f_i(x_i', x_{-i}) = \Phi(x_i, x_{-i}) - \Phi(x_i', x_{-i}), \quad \forall\, i \in [n],\ \forall\, x \in \mathcal{X},\ x_i' \in \mathcal{X}_i.$$

If the potential function Φ is μ–strongly convex, it follows from convex analysis that the game is μ–strongly monotone (Rockafellar, 1970). The following is an example of such a game where there is a natural decision-maker influencing the outcomes.

Example: Power Control in Shared Wireless Channel Another interesting class which has a similar structure to the Kelly auction is power control for shared wireless channels (d'Oro et al., 2015; Duvocelle et al., 2023; Facchinei and Kanzow, 2007; Tse and Viswanath, 2005). Consider n wireless users that aim to transmit a set of packets to a common receiver over a set $\mathcal{S}$ of shared wireless channels (subcarriers). The aggregate received signal y_s over the $s \in \mathcal{S}$ subcarrier is

$$y_s = \sum_{i=1}^{n} h_{i,s}\xi_{i,s} + z_s$$

where $\xi_{i,s}$ is the transmitted signal of user i over the s-th subcarrier, $h_{i,s}$ is the corresponding channel coefficient, and z_s is the aggregate interference-plus-noise received from all sources not in $[n]$ and for which we have that $z_s \sim \mathcal{N}(0, \sigma_s^2)$ is a Gaussian random variable. The average transmit power of user i on subcarrier s is $x_{i,s} = \mathbb{E}|\xi_{i,s}|^2$ and each users total power $\boldsymbol{x}_i$ satisfies $\boldsymbol{x}_i = \sum_s x_{i,s} \leq P_i$ for some $P_i > 0$. Then the strategy space of user i is

$$\mathcal{X}_i = \left\{ x_i \in \mathbb{R}^{|\mathcal{S}|} \middle| \; x_{i,s} \geq 0 \quad \text{and} \quad \sum_{s \in \mathcal{S}} x_{i,s} \leq P_i \right\}.$$

Each users transmission rate is given by Shannon's formula:

$$R_i(x_i, x_{-i}) = \sum_{s \in \mathcal{S}} \log\left(1 + \gamma_{i,s}(x)\right) = \sum_{s \in \mathcal{S}} \left(\log\left(\sigma_s^2 + w_s(x)\right) - \log\left(\sigma_s^2 + \sum_{j \neq i} g_{j,s} x_{j,s}\right)\right),$$

where $w_s(x) = \sum_{i \in [n]} g_{i,s} x_{i,s}$ for each $s \in \mathcal{S}$ and such that $g_{i,s} = |h_{i,s}|^2$ is the channel gain of user i over subcarrier s. Additionally, the term

$$\gamma_{i,s}(x) = \frac{g_{i,s} x_{i,s}}{\sigma_s^2 + \sum_{j \neq i} g_{j,s} x_{j,s}}$$

is the signal-to-inference-and-noise ratio. The network operator (decision-maker) aims to design a pricing scheme for the channel so as to induce an efficient equilibrium. For example, in a cognitive radio scenario the users described above are *secondary users* that are free riding on the network and cause interference on the primary users and therefore the the network operator needs to ensure that the system's users meet the quality of service guarantees that they have already paid for—typically in the form of minimum rate requirements or maximum interference tolerance per subcarrier. How this is achieved is by designing a pricing mechanism that consists of a flat spectrum access price $\pi_0 : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}$ and a user specific price $\pi_i : \mathcal{X}_i \to \mathbb{R}$. Thus user i's utility is given by

$$U_i(x) = R_i(x) - (\pi_0(w) + \pi_i(x_i)) \quad \text{where} \quad w = (w_1, \ldots, w_s).$$

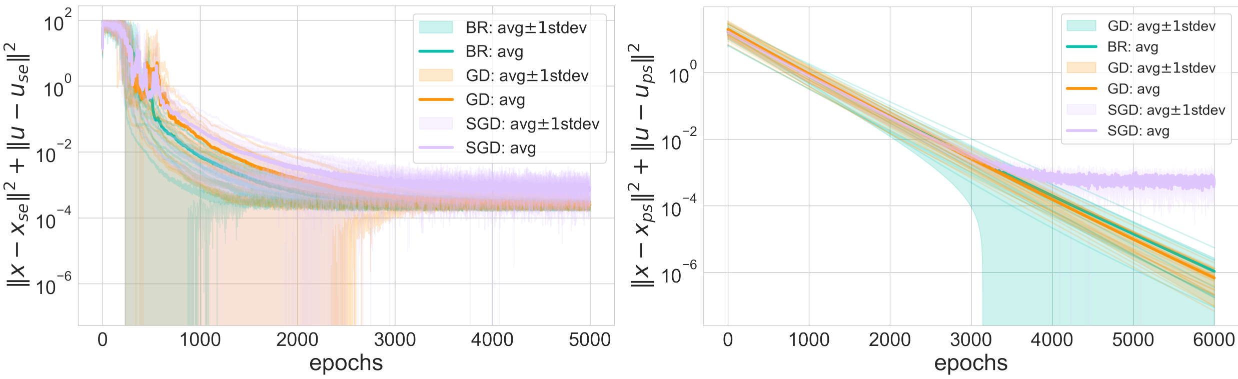
This game admits an exact potential function (d'Oro et al., 2015):

$$\Phi(x) = \sum_{s \in \mathcal{S}} \log(\sigma_s^2 + w_s) - \pi_0(w) - \sum_{i \in [n]} \pi_i(x_i).$$

To align with economic considerations on diminishing returns, it is common to assume that the pricing functions π_0 and each π_i is non-decreasing and convex in each of its arguments, and they are Lipschitz continuous. This ensures that $\Phi(x)$ is concave (though not necessarily strongly). Some regularization of the potential function would ensure its strongly concave but would induce a different set of Nash equilibrium than optimizing Φ. It is interesting to see how much regularization is introduced impacts the difference between the induced sets of equilibrium.

B Numerical Experiments

We briefly also describe our approach for our plot in Figure 1 in the following section.



(a) Strategic decision-maker

(b) Naïve decision-maker

Figure 2: Iteration complexity in quadratic games. The plots show the average over 20 random initializations and include the ± 1 standard deviation. Note that $(u_{\mathrm{se}}, x_{\mathrm{se}})$ denotes the Stackleberg equilibrium and $(u_{\mathrm{ps}}, x_{\mathrm{ps}})$ denotes the performatively stable Stackelberg equilibrium. Trajectories for three ρ–contracting algorithms are presented: exact best response with noisy observations (BR), deterministic gradient play with noisy observations (GD), and stochastic gradient play (SGD).

B.1 Interplay Between Sample Complexity, Performative Gap, & ε-Equilibrium

In Figure 1, we characterize the relationship between the expected tracking error, sample complexity, and the performative gap as parameterized by L_{eq}. In particular, we consider only values of L_{eq} where the conditions of Assumption 5, Theorem 2, and Theorem 4 are satisfied; in particular, games such that

$$L_{\mathsf{eq}} L_\ell < \alpha < L_\ell + L_{\mathsf{z}} L_{\mathsf{eq}} \quad \text{and} \quad \alpha - 2L_{\mathsf{eq}} L_\ell > 0.$$

These conditions ensure that there is a loss function for the decision-maker, that equilibrium exist, and that the performative gap is finite.

To help illustrate the tradeoff between the performatively stable Stackelberg equilibrium and the Stackelberg equilibrium, consider an ε-ball around the Stackelberg equilibrium u^*, defined formally as

$$B_\varepsilon(u^*) = \left\{ x \in \mathbb{R}^d \mid \|x - u^*\| \le \varepsilon \right\}$$

Suppose $u^{\mathsf{ps}} \in B_\varepsilon(u^*)$, which in turn implies that $\|u^{\mathsf{ps}} - u^*\| \le \varepsilon$; accordingly, denote the difference between the performative gap and ε as $\varepsilon' := \varepsilon - \|u^{\mathsf{ps}} - u^*\|$. If a decision-maker hopes to be ε-close to u^*, they have two options: follow the constants set in Corollary 4 to be within ε of u^* using the derivative-free method, *or* to use the repeated-gradient method and the constants set in Corollary 3 to be within ε' of u^{ps} and by construction within ε of u^*. Figure 1 reveals this tradeoff: while for many values of L_{eq} the sample complexity of the derivative-free method is significantly higher than that of the repeated-gradient method, as the performative gap grows, ε' shrinks, greatly increasing the sample complexity for the repeated-gradient method.

For instantiating the game used in Figure 1, we selected the following constants:

- $d = 2$, where d is the dimension of the decision-maker's action space $\mathcal{U}$ (see the top of Section 2).
- $R = 4$, where $R > \|x_0 - x^*(u_0)\|$ (see Theorem 4).
- $B = 1$, where B is the radius the ball containing $\mathcal{U}$ (see Assumption 5).
- $\alpha = 10$, where α is the strong-convexity constant for $\ell(\cdot, z)$ (see Assumption 3).
- $L_\ell = 0.8$, where L_ℓ is the Lipschitz constant for the map $(u, z) \mapsto \nabla_{u,z}\ell(u, z)$ (see Assumption 3).
- $\sigma^2 = 1$, where σ^2 bounds the variance of gradients g_t (see Assumption 4).

These constants may be changed in the provided code base.

B.2 Quadratic Ride-Sharing Game

We consider an example from ride-sharing markets. Demand signals may be used to create more efficient ride share markets without reducing individual revenue streams by enabling information-limited firms to recover latent demand. Using an analogous set up as in (Narang et al., 2022), we explore semi-synthetic competition between two ride-share platforms seeking to maximize their revenue given that the demand they experience is influenced by their own prices as well as their competitors. The data we use is from a prior Kaggle competition.[4]

Game Formulation. Each firm divides rides into \$5 price bins ranging from \$10 to \$30, and then chooses a additive surge on top of that price as described in (Narang et al., 2022). The social cost is given by

$$\mathcal{L}(u) = \mathop{\mathbb{E}}_{\xi\sim\mathcal{D}_0}[f_1^u(x_1, x_2) + f_2^u(x_1, x_2)]$$

where each firm i's cost is $f_i^u(x) = \mathbb{E}[-\frac{1}{2}z_i^\top x_i + \frac{\lambda_i}{2}\|x_i\|^2]$.

The action x_i is a vector of additive surge prices to each \$5 dollar bin, and the term $z_i^\top x_i$ represents the added revenue across bins achieved via surge pricing. We write z_i to be

$$z_i := \xi_i + A_{i,i}x_i + A_{i,-i}x_{-i} + u_i$$

In particular, the vector $A_i := \begin{bmatrix} A_{i,i} & A_{i,-i} \end{bmatrix}$ such that u_i acts as a demand signal informing the firms about latent demand. The regularization parameter λ_i serves to reduce the surge multiplier; that is, the firm does not want to inadvertently set the price too high.

B.2.1 Estimation of Price Elasticities

In many applications, a decision-maker may want to estimate the reactivity of agents. For example a local government may seek to estimate the price elasticity of agents—in this case ride-share companies—in a ride-share market so that they can then subsequently set taxes or subsidies on these agents or even the other side of the market (passengers).

In the context of the ride-share market example above, if a decision-maker aims to estimate each of the A_i's—i.e., the price elasticities of players—then they can run online least squares where in each round t they first query the environment and observe

$$z_{t,i} = \xi_{t,i} + A_i x_t \quad \text{where} \quad A_i = \begin{bmatrix} A_{i,i} & A_{i,-i} \end{bmatrix}.$$

Then, they perturb the prices with actions u_i for each player, and observe

$$q_{t,i} = A_i(x_t + u_{t,i}) + \xi'_{t,i}.$$

With these two queries, the decision maker updates their estimate of the price elasticities as follows:

$$\hat{A}_{t+1,i} = \hat{A}_{t,i} + \nu_t(q_{t,i} - z_{t,i} - \hat{A}_{t,i}u_{t,i})u_{t,i}^\top,$$

where ν_t is the step size.

In Narang et al. (2022), the authors show that if multiple firms are running a stochastic gradient method while simultaneously estimating their own price elasticities, then the joint strategy of the firms converges to the Nash equilibrium and the estimates of the price elasticities converge to the true values as long as the firms inject noise satisfying the following assumption.

Assumption 6. The sequence $u_t = (u_{t,1}, \ldots, u_{n,t}) \in \mathbb{R}^d$ is a zero-mean random vector that is independent of x_t, and independent of the previous random vectors $\{u_s | \, s < t\}$. Moreover, there exists constants $c_l, R > 0$ and $c_{u,i} > 0$ for each $i \in [n]$ such that for all $t \geq 0$ and $i \in [n]$ the random vector $v_i := u_{i,t}$ satisfies

$$0 \prec c_l \cdot I \preceq \mathbb{E}[v_i v_i^\top], \quad \mathbb{E}\|v_i\|^2 \leq c_{u,i}, \quad \text{and} \quad \mathbb{E}[\|v_i\|^2 v_i v_i^\top] \preceq R^2\mathbb{E}[v_i v_i^\top].$$

[4]Data is publicly available: https://www.kaggle.com/datasets/brllrb/uber-and-lyft-dataset-boston-ma

In our setting, it is an external third party that is injecting noise and they are decaying that noise over time with the goal of obtaining an approximate estimate of the price elasticities and then leaving the base system close to the nominal Nash equilibrium. The firms in this case are assumed to know their price elasticities A_i.[5] There is a tradeoff between how quickly the noise is decaying and the accuracy of the price elasticity estimates as well as where the agents actions end up relative to the nominal Nash equilibrium $x^\star$—i.e., the Nash equilibrium of the game $\mathcal{G}_u$ where $u = 0$.

Indeed, fixing a $\lambda \in (0, 1)$ and some horizon T, suppose that the decision-maker samples u_t from $\mathcal{N}(0, \sigma_t \cdot I_d)$ where $\sigma_{t+1} = \lambda \sigma_t$. Then, in Assumption 6 we have

$$c_l = \lambda^T \sigma_0, \quad c_{u,i} = d_i, \quad \text{and} \quad R^2 = 3 \max_{i \in [n]} d_i.$$

From Lemma 21 of Narang et al. (2022), we have that

$$\mathbb{E}\|\hat{A}_T - A\|_F^2 \leq \frac{\max\left\{\left(1 + \frac{2R^2}{\lambda^T \sigma_0}\right)\|\hat{A}_0 - A\|_F^2, \frac{8}{\lambda^{2T}\sigma_0^2}\sum_{i=1}^n \mathrm{Tr}(\Sigma_{0,i})c_{u,i}\right\}}{T + \frac{2R^2}{\lambda^T \sigma_0}}$$

where $\Sigma_{0,i} = \mathrm{diag}(\sigma_{0,1}, \ldots, \sigma_{0,n})$.

To bound the effect on the firms we analyze how injecting u_t impacts the convergence to the nominal Nash equilibrium. In this case, the firms are running stochastic gradient play with noise

$$\zeta_{t,i} = \xi_{i,t} + A_i u_{t,i}$$

where $\xi_{i,t} \sim \mathcal{D}_0$ and $u_{t,i} \sim \mathcal{N}(0, \sigma_{t,i} \cdot I_d)$. Hence, in the proof of the one step contraction for stochastic gradient play (cf. Lemma 5), we have that

$$P_1 \leq \frac{(\sigma_{\mathtt{a}} + \sigma_t)^2}{2\nu_1} + \frac{\nu_1 \mathbb{E}_t\|x_{t+1} - x_t\|^2}{2} = \frac{(\sigma_{\mathtt{a}} + \lambda^t \sigma_0)^2}{2\nu_1} + \frac{\nu_1 \mathbb{E}_t\|x_{t+1} - x_t\|^2}{2}$$

so that, after some algebra, we have that

$$\mathbb{E}_t\|x_{t+1} - x^*\|^2 \leq \frac{1}{1 + \gamma\mu}\|x_t - x^*\|^2 + \frac{2\gamma^2(\sigma_{\mathtt{a}} + \lambda^t \sigma_0)^2}{1 + \gamma\mu}.$$

We know that if the firms run stage-wise stochastic gradient play with some target accuracy $\varepsilon > 0$, then the agents obtain a Nash equilbirium in a total of T iterations where T is given in Corollary 5. Here T is fixed *a priori*, so in order to obtain an estimate for ε, consider that the total number of iterations satisfies

$$\begin{aligned}
T &= \sum_{k=0}^{K} T_k \\
&= \left\lceil \left(1 + \frac{2L_{\mathtt{a}}^2}{\mu^2}\right)\log\left(\frac{2R}{\varepsilon}\right)\right\rceil + \sum_{k=1}^{K}\left\lceil\left(1 + \frac{2^{k+1}L_{\mathtt{a}}^2}{\mu^2}\right)\log(4)\right\rceil \\
&= \left\lceil \left(1 + \frac{2L_{\mathtt{a}}^2}{\mu^2}\right)\log\left(\frac{2R}{\varepsilon}\right)\right\rceil + \frac{\left(-4\varepsilon L_{\mathtt{a}}^2 + \varepsilon\mu^2 + 8(\sigma_{\mathtt{a}} + \lambda^T\sigma_0)^2\right)\log(2) + \varepsilon\mu^2\log\left(\frac{(\sigma_{\mathtt{a}} + \lambda^T\sigma_0)^2}{\varepsilon L_{\mathtt{a}}^2}\right)}{\varepsilon\mu^2\log(2)}\log(4) \\
&= \left\lceil \left(1 + \frac{2L_{\mathtt{a}}^2}{\mu^2}\right)\log\left(\frac{2R}{\varepsilon}\right)\right\rceil + \left(1 - \frac{4L_{\mathtt{a}}^2}{\mu^2} + \frac{8(\sigma_{\mathtt{a}} + \lambda^T\sigma_0)^2}{\varepsilon\mu^2} + \frac{\log((\sigma_{\mathtt{a}} + \lambda^T\sigma_0)^2/(\varepsilon L_{\mathtt{a}}^2))}{\log(2)}\right) \\
&= \mathcal{O}\left(\frac{L_{\mathtt{a}}^2}{\mu^2}\log\left(\frac{2R}{\varepsilon}\right) + \frac{(\sigma_{\mathtt{a}} + \lambda^T\sigma_0)^2}{\mu^2\varepsilon}\right)
\end{aligned}$$

Hence, in terms of fixed T, the firms achieve an ε_T Nash equilibrium where

$$\varepsilon_T \asymp \frac{(\sigma_{\mathtt{a}} + \lambda^T\sigma_0)^2}{\mu^2 T}.$$

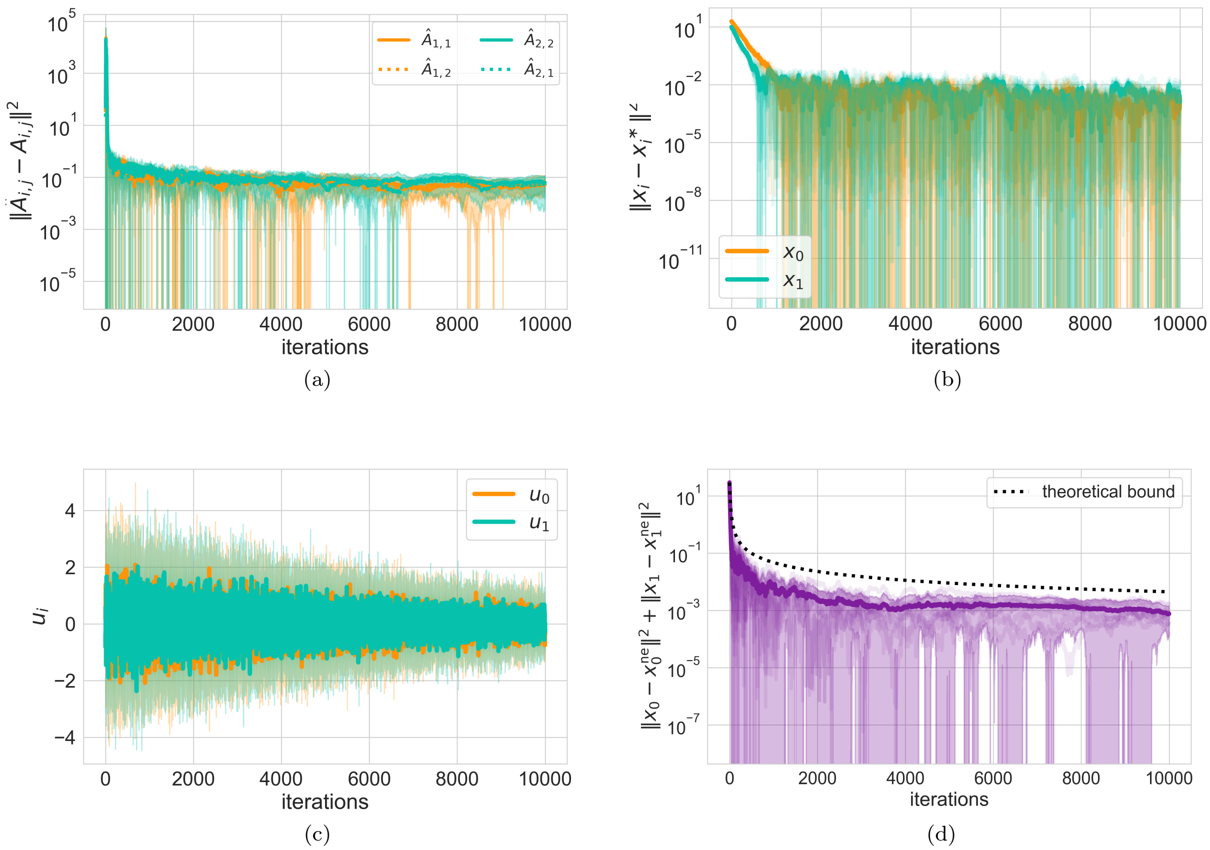


Figure 3: (a) Estimation error for the price elasticities $\hat{A}_{1,1}$, $\hat{A}_{1,2}$, $\hat{A}_{2,1}$, and $\hat{A}_{2,2}$. (b) Agent tracking error. (c) The deployed actions $u_{t,i}$. (d) The black dashed line is the theoretical bound on tracking error versus number of iterations, and the purple trajectories are the actual tracking error. For each of the plots, we run ten different random seeds and show the mean, the mean ± 1 standard deviation, and the actual trajectories using lower opacity. Note that although the magnitude of noise injected into the game by the decision-maker decays over time, the decision-maker still has sufficient information to estimate the price elasticities.

In Figure 3(a), we show that the oblivious decision-maker is able to effectively estimate the price elasticities of each of the agents. Indeed, through repeatedly perturbing the learning agents via a decaying random sequence $\{u_t\}$ (cf. Figure 3(c))—which we may interpret here as the decision-maker exploring the space of subsidies/taxes—the decision-maker learns the A_i's. This process of perturbation makes sure that the system is persistently excited—analogous to adaptive control—and therefore the estimation problem is well-posed. Since the perturbations are decaying, the decision-maker can learn the A_i's while the agents converge to the Nash equilibrium of the unperturbed game as seen in Figure 3(b). This likewise aligns with our theoretical understanding of their behavior under drift induced by the decision-maker as shown in Figure 3(d). Essentially here the decision-maker is decaying the drift thereby eventually making the system enter the low-drift-to-noise regime as identified in the theoretical results.

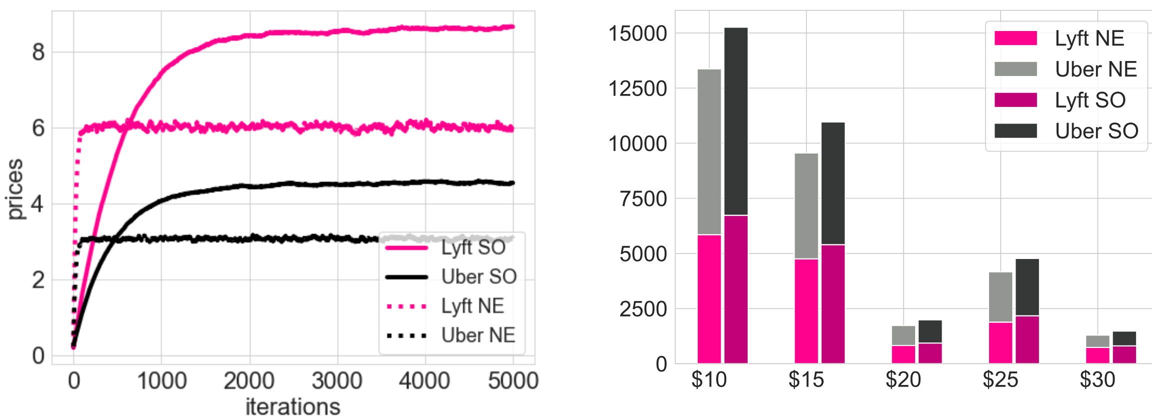
[5] Another interesting example would be having the firms be simultaneously estimating A_i and looking at the combined effects of injecting noise by the decision-maker and the firms. If multiple entities are injecting noise it could be the case that the combined effect reduces the time for convergence or makes it worse depending on the injected noise distributions.

B.2.2 Socially Optimal Demand Signal Provisioning

Recall that the social cost is $\mathcal{L}(u) = \mathbb{E}_{\xi\sim\mathcal{D}_0}[f^1(x) + f^2(x)]$. Define the socially optimal intervention to be

$$u^{\mathsf{so}} := \underset{u\in\mathcal{U}}{\operatorname{argmin}}\ \mathcal{L}(u).$$

In this numerical example, we explore the effect of the decision-maker intervening with the socially optimal demand signal versus no intervention. To this end, we compute u^{so} using a symbolic solver (i.e., Mathematica) and then simulate stochastic gradient play on the player objectives $f_i^{u^{\mathsf{so}}}(x)$.



(a) (b)

Figure 4: (a) Surge prices in the \$10 price bin given no intervention—i.e., $u = 0$—and the socially optimal intervention—i.e. the optimal demand signal $u^{\mathsf{so}} = \operatorname{argmin}_u \mathcal{L}(u)$. With no intervention the agents converge to the Nash equilibrium (NE) of their nominal game; with the socially optimal intervention, the agents converge to the socially optimal (SO) equilibrium. (b) Welfare for each firm in each of the price bins under the Nash equilibrium and socially optimal equilibrium.

Figure 4 illustrates the players' behavior when there is no intervention by the decision-maker (i.e., $u = 0$) and when the decision-maker intervenes with the socially optimal intervention (i.e., $u = u^{\mathsf{so}}$). In particular, Figure 4(a) shows that stochastic gradient play converges to the Nash equilibrium under $u = 0$ and to the social optimum under $u = u^{\mathsf{so}}$. We plot this for a single location and price bin, both of which can be changed in the code base. The simulations show that when provided with the optimal demand signal, the firms are induced to increase their prices which indicates that user are willing to pay more for the service and competition between services actually drives prices down. An interesting question would be exploring a social cost that includes the cost to users, incorporating elements such as cost of alternative means of transportation such as public transit; unfortunately this data set does not include such information and would need to be augmented so we leave that to future work.

Figure 4(b) shows that the welfare for each firm is higher in all the price bins under the social optimum, though more marginally as the price increases. This demonstrates that a decision-maker who is able to provide optimal and informative demand signals may be able to improve the welfare of the ridesharing marketplace, even under competition by the firms, since both the prices are lower (thereby increasing the demand and cost to passengers) and the revenue is higher for all players. It is also interesting to observe that the smaller player in the market—namely Lyft which has less demand in the data set—has a larger marginal gain than the larger player (Uber) in the market.

Algorithm 2 Geometric Decay Schedule

1: **Input** $y_0 \in \mathbb{R}^d$, $C, D > 0$, $\delta_0 \in (0,1)$, estimate $\Delta \geq h(y_0)$, accuracy $\epsilon > 0$, algorithm $\mathcal{A}(y, \delta, T)$ satisfying (9);
2: **Initialize:** Set $y_0 = \mathcal{A}(y_0, \delta_0, T_0)$ with $T_0 = \frac{1}{\psi(\delta_0)} \cdot \log\left(\frac{2C\Delta}{\epsilon}\right)$;
3: Set $K = \left\lceil 1 + \log_2\left(\frac{D\delta_0}{\epsilon}\right)\right\rceil$.
4: **for** $k = 1, \ldots, K$ **do**
5: Set $y_k = \mathcal{A}(y_{k-1}, \delta_k, T_k)$ with $\delta_k = 2^{-k}\delta_0$, $T_k = \left\lceil \frac{1}{\psi(\delta_k)} \cdot \log(4C)\right\rceil$.
6: **end for**
7: **Return** y_K.

C Technical Lemmas

We need the following standard technical lemma for convergence of sequences and high-probability guarantees.

Lemma 1. Consider a sequence $w_t \geq 0$ for $t \geq 1$ and constants $t_0 \geq 0$, $a > 0$ satisfying

$$a_{t+1} \leq \left(1 - \frac{2}{t+t_0}\right) a_t + \frac{c}{(t+t_0)^2} \tag{7}$$

Then the following estimate holds:

$$a_t \leq \frac{\max\{(1+t_0)a_1, c\}}{t+t_0} \quad \forall\, t \geq 1. \tag{8}$$

We also restate the following Lemma, adapted from (Drusvyatskiy and Xiao, 2023).

Lemma 2 (Lemma B.2, (Drusvyatskiy and Xiao, 2023))**.** Suppose we have a stochastic algorithm $\mathcal{A}(y_0, \delta, T)$ such that as long as $\delta < \delta_0$, the method generates a point satisfying

$$\mathbb{E}[h(Y_T)] \leq C(1-\psi(\delta))^T h(y_0) + D\delta, \tag{9}$$

where h is a non-negative function, $C, D > 0$, and $\delta_0 \in (0,1)$ are constants specific to the algorithm, and ψ is a function mapping $[0, \delta_0)$ into $(0,1)$. The point y returned by Algorithm 2 satisfies $\mathbb{E}[h(y_K)] \leq \epsilon$ with the efficiency estimate

$$\sum_{k=0}^{K} T_k = \left\lceil \frac{1}{\psi(\delta_0)} \cdot \log\left(\frac{2C\Delta}{\epsilon}\right)\right\rceil + \sum_{k=1}^{K} \left\lceil \frac{\log(4C)}{\psi(2^{-k}\delta_0)}\right\rceil$$

Note that this does not immediately apply to our general case: h is generally a fixed non-negative function, such as $h(y) = \|y - \bar{x}\|^2$ the distance to some fixed equilibrium $\bar{x}$. However, in our case, our target equilibrium is changing depending on the action u_t from the decision-maker. A short corollary, however, gives us the desired result.

Lemma 3. Suppose we have a stochastic algorithm $\mathcal{A}(y_0, \delta, T)$ such that as long as $\delta < \delta_0$, the method generates a point satisfying

$$\mathbb{E}\,\|y_T - y_T^*\|^2 \leq C(1-\psi(\delta))^T \|y_0 - y_0^*\|^2 + D\delta$$

for C, D, δ_0, and ψ as defined above. Then the point y returned by Algorithm 2 satisfies $\mathbb{E}\,\|y_K - y_K^*\| \leq \epsilon$ with the efficiency estimate

$$\sum_{k=0}^{K} T_k = \left\lceil \frac{1}{\psi(\delta_0)} \cdot \log\left(\frac{2C\Delta}{\epsilon}\right)\right\rceil + \sum_{k=1}^{K} \left\lceil \frac{\log(4C)}{\psi(2^{-k}\delta_0)}\right\rceil$$

Proof. For posterity, we include a short proof. Suppose that our target accuracy is $2D\delta$. Then note that it would be sufficient to run our algorithm for T_0 iterations such that

$$C(1-\psi(\delta))^{T_0}\|y_0 - y_0^*\|^2 \leq D\delta.$$

Note that since $\psi(\delta) \in (0,1)$, we have that $-\log(1-\psi(\delta)) > \psi(\delta)$, so

$$\frac{1}{\psi(\delta)} \log\left(\frac{C\|y_0 - y_0^*\|^2}{D\delta}\right) \leq T_0.$$

By the concavity of log, it is in fact sufficient to chose

$$\frac{1}{\psi(\delta)} \log\left(\frac{C\|y_0 - y_0^*\|^2}{\epsilon}\right) \leq T_0.$$

Then we proceed just as that of (Drusvyatskiy and Xiao, 2023, B.2). □

D Regularity of the Equilibrium Response

Recall that $\omega_u(x) := (\nabla_1 f_1^u, \ldots, \nabla_n f_n^u)$. Strong metric regularity allows for Lipschitz continuity of solutions to $\omega_u(x) \in N_{\mathcal{X}}(x)$ to be Lipschitz continuous. The following proposition is a formal statement of the discussion in Section 2.

Proposition 5 (Inner Problem Regularity: Polyhedral Constraints)**.** Under Assumption 1.i–iii, suppose that, for any fixed $u \in \mathcal{U}$, the Jacobian of $\omega_u(x)$ with respect to x is non-degenerate and the Jacobian with respect to u has finite operator norm—i.e., $\det(D_x\omega_u(x)) \neq 0$ and $\|D_u\omega_u(x)\|_{\mathsf{op}} < \infty$. Then, Assumption 1.iv is satisfied with $\kappa := \frac{1}{\mu}\sup_{(u,x)\in\mathcal{X}\times\mathcal{Y}} \|D_u\omega_u(x)\|_{\mathsf{op}}$.

This proposition follows precisely from Dontchev et al. (2009, Chapter 2.F, Chapter 3); indeed, the strong metric regularity parameter in this case is equivalent to the Lipschitz continuity parameter of the implicit function.

To give some intuition, we can consider the case where $\mathcal{X}$ is the whole Euclidean space $\mathbb{R}^m$. In this section alone, we define $\omega(u,x) := \omega_u(x)$ for the purpose of clarity on the derivatives herein. In this case, by the fact that the joint strategy space $\mathcal{X}$ is unconstrained, for any fixed $u \in \mathcal{U}$, assuming $\det(D_x\omega(u,x)) \neq 0$, the Nash equilibrium $x^*(u)$ is defined as an implicit function (cf. Abraham et al. (2012)) that solves $\omega(u, x^*(u)) = 0$. By the implicit function theorem, the derivative $Dx^*(u)$ is given by

$$Dx^*(u) = -D_x\omega(u, x^*(u))^{-1} D_u\omega(u, x^*(u)).$$

We have the following lemma which provides sufficient conditions for $x^*(u)$ to be Lipschitz by assuming suitable bounds on $\|D_u\omega(u,x)\|_{\mathsf{op}}$ and $\|D_x\omega(u,x)\|_{\mathsf{op}}$.

Lemma 4. Suppose that $\|D_x\omega(u,x)\|_{\mathsf{op}} \geq \mu_1$ and $\|D_u\omega(u,x)\|_{\mathsf{op}} \leq \mu_2$ for all $x \in \mathcal{X}$ and set $L_{\mathsf{eq}} := \frac{\mu_2}{\mu_1}$. Then $x^*(u)$ is L_{eq}-Lipschitz.

Proof. We realize that by the mean value theorem,

$$\begin{aligned}
\|x^*(u) - x^*(u')\| &\leq \left\| \left(\int_0^1 Dx^*((1-\lambda)u + \lambda u')\lambda \right)(u - u') \right\| \\
&\leq \sup_{\lambda\in[0,1]} \|Dx^*((1-\lambda)u + \lambda u')\|_{\mathsf{op}} \|u - u'\|
\end{aligned}$$

By the assumption, we have $\|D_x\omega(u, x^*(u))\| \geq \mu_1$ and $\|D_u\omega(u, x^*(u))\| \leq \mu_2$ for all $u \in \mathcal{U}$, then we have that

$$\sup_{\lambda\in[0,1]} \|Dx^*((1-\lambda)x + \lambda x')\| \leq \frac{\mu_2}{\mu_1} := L_{\mathsf{eq}},$$

which concludes the proof. □

Again, in the polyhedral constraint case, the analysis above is almost identical; see, e.g., Dontchev et al. (2009).

E Strong Monotonicity of Agent Game

Recall that Assumption 1.i. Here we explore the strength of this assumption. In particular, this assumption requires that the decision maker is only deploying u's such that the agents game is strongly monotone. A sufficient condition for this game to be strongly monotone for a given u is that the game Jacobian is positive definite:

$$J_u(x) := \begin{bmatrix} \nabla_1^2 f_1^u & \nabla_{12} f_1^u & \cdots & \nabla_{1n} f_n^u \\ \nabla_{21} f_2^u & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \nabla_{(n-1)n} f_{n-1}^u \\ \nabla_{n1} f_n^u & \cdots & \nabla_{n(n-1)} f_n^u & \nabla_n^2 f_n^u \end{bmatrix} \succ 0.$$

It is instructive to see what this implies via an example. Consider a revenue maximization game amongst the players

$$f_i^u(x_i, x_{-i}) = (A_{ii} x_i + A_{i,-i} x_{-i} + \xi_i + u_i)^\top x_i + \frac{\lambda_i}{2} \|x_i\|^2$$

here u_i is some demand signal to correct for the implicit bias agent i has about the demand.

For simplicity let's consider the two player case:

$$\omega_u(x) = (\nabla_1 f_1^u, \nabla_2 f_2^u) = ((2A_{11} + \lambda_1 I)x_1 + \xi_1 + u_1 + A_{12} x_2, (2A_{22} + \lambda_2 I)x_2 + \xi_2 + u_2 + A_{21} x_1)$$

$$J_u(x) = \begin{bmatrix} 2A_{11} + \lambda_1 I & A_{12} \\ A_{21} & 2A_{22} + \lambda_2 I \end{bmatrix}$$

Then,

$$\frac{1}{2}(J_u(x) + J_u^\top(x)) = \begin{bmatrix} 2A_{11} + \lambda_1 I & A_{12} + A_{21}^\top \\ A_{12}^\top + A_{21} & 2A_{22} + \lambda_2 I \end{bmatrix}$$

Then as long as the game when $u = 0$ is strongly monotone, so is any induced game.

Prior work by Ratliff and Fiez (2020) on adaptive incentive design with simultaneous utility estimation incorporates this condition as a constraint on the optimization problem (for choosing the next u_t). One direction for future work is specifiying the space of $\mathcal{U}$ via a similar radial basis function method as in Ratliff and Fiez (2020), and then characterizing the additional constraint to be added to the epoch-based algorithms we propose herein. If this expansion results in a closed, convex set $\mathcal{U}$, then our results will apply and the decision-maker will only be choosing actions u that retain strong monotonicity of the agents' game.

F Contracting Agent Updates

In this section, we show that several natural learning dynamics are ρ-contracting for some $\rho \in [0, 1)$. The following is a modified version of Assumption 1 where we remove the decision-maker for simplicity.

Assumption 7. The following hold:

1. The game $\mathcal{G} := (f_1, \ldots, f_n)$ is a C^1-smooth convex game and μ–strongly monotone;
2. The mappings $x_i \mapsto \nabla_i f_i(x_i, x_{-i})$ are L_i–Lipschitz continuous;
3. The game $\mathcal{G}$ is κ–strongly metrically regular.

In this section, set $\omega(x) := (\nabla_1 f_1(x), \ldots, \nabla_n f_n(x))$, and recall that we have set $L_{\mathtt{a}} := \max_{i \in [n]} L_i$.

F.1 Stochastic Gradient Play and Asynchronous Gradient Play

Consider first players updating according to the stochastic gradient method given by

$$x_{t+1} = \operatorname*{proj}_{\mathcal{X}}(x_t - \gamma \widehat{\omega}(x_t)), \tag{10}$$

where $\mathbb{E}[\widehat{\omega}(x_t)] = \omega(x_t)$. This is *stochastic gradient play.*

Assumption 8. Suppose that there exists a constant $\sigma_{\mathtt{a}} > 0$ satisfying

$$\mathbb{E}[\|\widehat{\omega}(x_t) - \omega(x_t)\|^2] \leq \sigma_{\mathtt{a}}^2.$$

Given the above assumption on the variance of the estimator $\widehat{\omega}$ of the vector of individual gradients, we have the following lemma showing that stochastic gradient play is ρ–contracting.

Lemma 5. Under Assumptions 7 and 8, suppose that players update according to stochastic gradient play with $\gamma \leq \frac{\mu}{2L_{\mathtt{a}}^2}$. Then, the dynamics satisfy

$$\mathbb{E}\,\|x_{t+1} - x^*\|^2 \leq \frac{1}{1+\mu\gamma}\mathbb{E}\|x_t - x^*\|^2 + \frac{2\gamma^2\sigma_{\mathtt{a}}^2}{1+\gamma\mu},$$

so that (10) is ρ-contracting with $\rho^2 = \frac{1}{1+\mu\gamma}$ and $c = \sqrt{2}\gamma$.

Proof. Observe that $x \mapsto \frac{1}{2}\|x_t - \gamma\widehat{\omega}(x_t) - x\|^2$ is a 1-strongly convex function over $\mathcal{X}$. Hence we deduce that

$$\begin{aligned}
\frac{1}{2}\|x_{t+1} - x^*\|^2 &\leq \frac{1}{2}\|x_t - \gamma\widehat{\omega}(x_t) - x^*\|^2 - \frac{1}{2}\|x_t - \gamma\widehat{\omega}(x_t) - x_{t+1}\|^2 \\
&\leq \frac{1}{2}\|x_t - x^*\|^2 - \gamma\langle\widehat{\omega}(x_t), x_{t+1} - x^*\rangle - \frac{1}{2}\|x_{t+1} - x_t\|^2 \\
&= \frac{1}{2}\|x_t - x^*\|^2 - \gamma\langle\widehat{\omega}(x_t), x_t - x^*\rangle - \frac{1}{2}\|x_{t+1} - x_t\|^2 - \gamma\langle\widehat{\omega}(x_t), x_{t+1} - x_t\rangle.
\end{aligned}$$

Taking expectations, we have that

$$\begin{aligned}
\frac{1}{2}\mathbb{E}_t\|x_{t+1} - x^*\|^2 &\leq \frac{1}{2}\|x_t - x^*\|^2 - \gamma\langle\mathbb{E}_t\widehat{\omega}(x_t), x_t - x^*\rangle - \frac{1}{2}\mathbb{E}_t\|x_{t+1} - x_t\|^2 - \gamma\mathbb{E}_t\langle\widehat{\omega}(x_t), x_{t+1} - x_t\rangle \\
&\leq \frac{1}{2}\|x_t - x^*\|^2 - \gamma\langle\omega(x_t), x_t - x^*\rangle - \frac{1}{2}\mathbb{E}_t\|x_{t+1} - x^*\|^2 - \gamma\mathbb{E}_t\langle\widehat{\omega}(x_t), x_{t+1} - x_t\rangle \\
&= \frac{1}{2}\|x_t - x^*\|^2 - \gamma\mathbb{E}_t\langle\omega(x_{t+1}), x_{t+1} - x^*\rangle - \frac{1}{2}\mathbb{E}_t\|x_{t+1} - x_t\|^2 \\
&\quad + \gamma\underbrace{\mathbb{E}_t\langle\widehat{\omega}(x_t) - \omega(x_t), x_t - x_{t+1}\rangle}_{=:P_1} + \gamma\underbrace{\mathbb{E}_t\langle\omega(x_t) - \omega(x_{t+1}), x^* - x_{t+1}\rangle}_{=:P_2}.
\end{aligned}$$

Since the game is μ–strongly monotone, we have that

$$\langle\omega(x_{t+1}), x_{t+1} - x^*\rangle \geq \langle\omega(x_{t+1}) - \omega(x^*), x_{t+1} - x^*\rangle \geq \mu\|x_{t+1} - x^*\|^2.$$

This in turn implies that

$$\frac{1+2\gamma\mu}{2}\|x_{t+1} - x^*\|^2 \leq \frac{1}{2}\|x_t - x^*\|^2 - \frac{1}{2}\mathbb{E}_t\|x_{t+1} - x_t\|^2 + \gamma(P_1 + P_2).$$

Employing Young's inequality, we upper bound P_1 as follows:

$$P_1 \leq \frac{\sigma_{\mathtt{a}}^2}{2\nu_1} + \frac{\nu_1\mathbb{E}_t\|x_{t+1} - x_t\|^2}{2}.$$

Applying Young's inequality, we bound P_2 as follows:

$$\begin{aligned}
P_2 &\leq \frac{\mathbb{E}_t\|\omega(x_t) - \omega(x_{t+1})\|^2}{2\nu_2} + \frac{\nu_2\mathbb{E}_t\|x_{t+1} - x^*\|^2}{2}, \\
&\leq \frac{L_{\mathtt{a}}^2\mathbb{E}_t\|x_t - x_{t+1}\|^2}{2\nu_2} + \frac{\nu_2\mathbb{E}_t\|x_{t+1} - x^*\|^2}{2},
\end{aligned}$$

so that

$$\frac{1+2\gamma\mu - \gamma\nu_2}{2}\mathbb{E}_t\|x_{t+1} - x^*\|^2 \leq \frac{1}{2}\|x_t - x^*\|^2 + \frac{\sigma_{\mathtt{a}}^2}{2\nu_1} - \frac{(1 - \gamma L_{\mathtt{a}}^2\nu_2^{-1} - \gamma\nu_1)}{2}\mathbb{E}_t\|x_{t+1} - x_t\|^2.$$

Setting $\nu_2 = \mu$ and $\nu_1 = \gamma^{-1} - L_{\mathtt{a}}^2/\mu$, we have that the last term on the right hand side is zero, and since $\gamma \le \frac{\mu}{2L_{\mathtt{a}}^2}$ we have that $\nu_1 \ge \frac{1}{2\gamma}$; indeed, $-\frac{1}{2\eta} \le -\frac{2L_{\mathtt{a}}^2}{\mu}$ so that $\nu_1 = \frac{1}{\eta} - \frac{2L_{\mathtt{a}}^2}{\mu} \ge \frac{1}{\eta} - \frac{1}{2\eta} = \frac{1}{2\eta}$. Therefore

$$\mathbb{E}_t\|x_{t+1} - x^*\|^2 \le \frac{1}{1+\gamma\mu}\|x_t - x^*\|^2 + \frac{2\gamma^2\sigma_{\mathtt{a}}^2}{1+\mu\gamma},$$

which concludes the proof. $\square$

The following corollary demonstrates how agents could use a stage-based algorithm to decrease the bias in their tracking error estimate and achieve a target accuracy.

Corollary 5 (Stage-wise Stochastic Gradient Play)**.** Consider some target accuracy $\varepsilon > 0$ and suppose we have a constant $R \ge \|x_0 - x^*\|^2$. Define $\psi(\gamma) := 1 - \frac{1}{1+\mu\gamma}$, $C := 1$, and $D := \frac{2\sigma_{\mathtt{a}}^2}{\mu}$. Set $\gamma_0 := \frac{\mu}{2L_{\mathtt{a}}^2}$. Then running Algorithm 2 with stochastic gradient play as $\mathcal{A}$, guarantees that $\mathbb{E}\|x_t - x^*\|^2 \le \varepsilon$ after

$$T = \sum_{k=0}^{K} T_k = \left\lceil \left(1 + \frac{2L_{\mathtt{a}}^2}{\mu^2}\right)\log\left(\frac{2R}{\varepsilon}\right)\right\rceil + \sum_{k=1}^{K}\left\lceil \left(1 + \frac{2^{k+1}L_{\mathtt{a}}^2}{\mu^2}\right)\log(4)\right\rceil$$

total iterations where $K := \left\lceil 1 + \log_2\left(\frac{\sigma_{\mathtt{a}}^2}{L_{\mathtt{a}}^2\varepsilon}\right)\right\rceil$.

Proof. Note that from Lemma 5, we can iterate further to get that

$$\begin{aligned}
\mathbb{E}_t\|x_t - x^*\|^2 &\le \frac{1}{1+\mu\gamma}\|x_{t-1} - x^*\|^2 + \frac{2}{1+\mu\gamma}\gamma^2\sigma_{\mathtt{a}}^2 \\
&\le \frac{1}{1+\mu\gamma}\left(\frac{1}{1+\mu\gamma}\|x_{t-2} - x^*\|^2 + \frac{2}{1+\mu\gamma}\gamma^2\sigma_{\mathtt{a}}^2\right) + \frac{2}{1+\mu\gamma}\gamma^2\sigma_{\mathtt{a}}^2 \\
&\le \left(\frac{1}{1+\mu\gamma}\right)^t\|x_0 - x^*\|^2 + 2\sigma_{\mathtt{a}}^2\gamma^2\sum_{s=1}^{t}\frac{1}{(1+\mu\gamma)^s} \\
&\le \left(\frac{1}{1+\mu\gamma}\right)^t\|x_0 - x^*\|^2 + \gamma\frac{2\sigma_{\mathtt{a}}^2}{\mu} \\
&\le \left(1 - \left(1 - \frac{1}{1+\mu\gamma}\right)\right)^t\mathbb{E}\|x_0 - x^*\|^2 + \gamma\frac{2\sigma_{\mathtt{a}}^2}{\mu}.
\end{aligned}$$

Letting $\psi(\gamma) = 1 - \frac{1}{1+\mu\gamma}$, $C = 1$, and $D = \frac{2\sigma_{\mathtt{a}}^2}{\mu}$, invoking Corollary 3 and running Algorithm 2 with the specified parameters gives us the desired result. $\square$

Asynchronous Updates. In practice, it may not be the case that the agents observe data or actions synchronously, and as a result they may not have the requisite information to update their action in every time step. A natural model to capture asynchronous updates is one in which agent i receives sufficient information to update its decision y_i with probability p_i. For instance, this means that

$$x_{i,t+1} = \begin{cases} \operatorname{proj}_{\mathcal{X}_i}\left(x_{i,t} - \gamma\nabla_i f_i(x_{i,t}, x_{-i.t})\right), & \text{w.p. } p_i \\ x_{i,t}, & \text{w.p. } (1-p_i) \end{cases} \tag{11}$$

Let $p_{\mathtt{max}} := \max_{i\in[n]} p_i$ and $p_{\mathtt{min}} := \min_{i\in[n]} p_i$. Then as described in (Narang et al., 2022), this can be dealt with using techniques from preconditioning in optimization—see, e.g., (Chasnov et al., 2020b; Huo and Huang, 2017; Lian et al., 2015; Recht et al., 2011; Zhou et al., 2018) and references therein.

The analysis in Lemma 5 does not change much; the primary difference is that the Lipschitz constant $L_{\mathtt{a}}$ is rescaled by $p_{\mathtt{max}}$ and the strong monotonicity constant μ is rescaled by $p_{\mathtt{min}}$. The reason this works out is that we can simply perform the exact same analysis using a modified inner product as has been performed in prior literature—i.e., we simply perform the analysis in the inner product $[x, y] = \langle P^{-1}x, y\rangle$ where $P = \operatorname{diag}(p_1, \ldots, p_n)$.

F.2 Momentum Updates: Strongly Convex-Strongly Concave Zero-Sum Games

Consider a strongly convex, strongly concave zero sum game $(f, -f)$ where player one seeks to minimize $f(x_1, x_2)$ with respect to x_1 and player two seeks to maximize f with respect to x_2. It is known that such games are strongly monotone. Momentum based updates such as optimistic gradient descent-ascent (OGDA) and negative momentum are ρ contracting for such games. This family of updates is given by

$$x_{t+1} = (1+\beta)x_t - \beta x_{t-1} - \gamma((1+\alpha)\omega(x_t) - \alpha\omega(x_{t-1})), \tag{12}$$

where α is the extrapolation parameter, β is the momentum parameter, and

$$\omega(x) = \begin{bmatrix} \nabla_1 f(x_1, x_2) \\ -\nabla_2 f(x_1, x_2) \end{bmatrix}.$$

For example, standard gradient descent-ascent is equivalent to setting $(\alpha, \beta) = (0, 0)$. OGDA is given by $(\alpha, \beta) = (1, 0)$ and negative momentum is given by $(\alpha, \beta) = (0, \beta)$ for some $\beta < 1$.

Let $\kappa := L_{\mathsf{a}}/\mu$. Gradient decent ascent is a commonly studied update and has been shown to be ρ-contracting with $\rho = O(1 - \kappa^{-2})$ (Ryu and Boyd, 2016). Mokhtari et al. (2020) study both OGDA and proximal point methods for this class of games. They show that OGDA is ρ-contracting with $\rho = O(1-\kappa^{-1})$, and that proximal point methods are also ρ-contracting with $\rho = 1/(1+\gamma\mu)$ and $c = 0$ in both cases. Zhang and Wang (2021) show that the negative momentum based update is ρ-contracting for strongly convex, strongly concave zero sum games, which are known to be strongly monotone. Specifically they say that negative momentum is suboptimal, but nonetheless, still ρ-contracting with $\rho = 1 - \Theta(\kappa^{-1.5})$.

F.3 Best Response Dynamics

Now, we show that the best response dynamics converge linearly to the Nash equilibrium of the game. This result is commonly known and the proof is analogous to (Narang et al., 2022, Theorem 1), with one exception where we obtain a tighter bound on the regime where linear convergence is guaranteed. Nevertheless, we include it for convenience. Define

$$\mathsf{BR}(x) := \{x' \in \mathcal{X} : \ x'_i \text{ is a best response to } \ x'_{-i} \ \forall i \in [n]\}.$$

That is, unrolling notation, given a current decision vector x_t, the updated decision vector x_{t+1} is such that

$$x_{i,t+1} = \operatorname*{argmin}_{x_i \in \mathcal{X}_i} f_i(x_i, x_{-i,t}) \quad \forall\ i \in [n]. \tag{13}$$

Lemma 6. Under Assumption 7, set $\rho := \frac{L_{\mathsf{a}}\sqrt{n-1}}{\mu}$ and suppose that we are in the regime where $\rho < 1$, and that players update according to (13). Then, the game admits a unique Nash equilibrium $x^\star \in \mathcal{X}$ and the best response process converges linearly:

$$\|x_{t+1} - x^\star\| \le \rho\|x_t - x^\star\| \quad \forall\ t \ge 0.$$

Proof. Since the game is μ strongly monotone, we have that

$$\sum_{i=1}^{n} \langle \nabla_i f_i(u) - \nabla_i f_i(u'), u_i - u'_i \rangle \ge \mu\|u - u'\|^2. \tag{14}$$

We will show the map $\mathsf{BR}(\cdot)$ is Lipschitz continuous with parameter ρ. To this end, consider a point $w \in \mathcal{X}$ and set $x := \mathsf{BR}(w)$. For each $i \in [n]$, first order optimality conditions for x_i guarantee that

$$\langle \nabla_i f_i(x_i, w_{-i}), x_i - x'_i \rangle \le 0 \quad \forall\ x'_i \in \mathcal{X}_i.$$

Strong monotonicity implies that, for any $x_i, x'_i \in \mathcal{X}_i$, we have that

$$\langle \nabla_i f_i(x_i, w_{-i}) - \nabla_i f_i(x'_i, w_{-i}), x_i - x'_i \rangle \ge \mu\|x_i - x'_i\|^2 \quad \text{for each } i \in [n].$$

Indeed this follows from (14) bx letting $u = (x_i, w_{-i})$ and $u' = (x_i', w_{-i})$ for each $i \in [n]$. Hence

$$\mu\|x_i - x_i'\|^2 \leq \langle \nabla_i f_i(x_i, w_{-i}) - \nabla_i f_i(x_i', w_{-i}), x_i - x_i' \rangle \leq \langle -\nabla_i f_i(x_i', w_{-i}), x_i - x_i' \rangle.$$

Since this holds for any x_i' we can replace x_i' with $x_i^\star$ to get that

$$\mu \sum_{i=1}^n \|x_i - x_i^\star\|^2 \leq \sum_{i=1}^n \langle \nabla_i f_i(x_i^\star, w_{-i}), x_i^\star - x_i \rangle$$

Since $x^\star$ is a Nash equilibrium, we have that

$$\langle \omega(x^\star), x - x^\star \rangle \leq 0 \quad \forall\, x \in \mathcal{X}.$$

Hence, we deduce that

$$\begin{aligned}
\mu\|x - x^\star\|^2 &\leq \sum_{i=1}^n \langle \nabla_i f_i(x_i^\star, w_{-i}), x_i^\star - x_i \rangle, \\
&\leq \langle \omega(x_i^\star, w_{-i}) - \omega(x^\star), x^\star - x \rangle, \\
&\leq \|\omega(x_i^\star, w_{-i}) - \omega(x^\star)\| \|x^\star - x\|, \\
&\leq L_{\mathsf{a}} \|x^\star - x\| \sum_{i=1}^n \|w_{-i} - x_{-i}^\star\|, \\
&= L_{\mathsf{a}} \langle u_1, u_2 \rangle,
\end{aligned}$$

where $u_1 = (\|x_1^\star - x_1\|, \ldots, \|x_n^\star - x_n\|)$ and $u_2 = (\|w_{-1} - x_{-1}^\star\|, \ldots, \|w_{-n} - x_{-n}^\star\|)$. Letting $\zeta = \|w - x^\star\|$, we have that $\zeta^2 = \|w_{-i} - x_{-i}^\star\|^2 + \|w_i - x_i^\star\|^2$ for each $i \in [n]$. Observe that

$$\begin{aligned}
\|u_2\| &= \left(\sum_{i=1}^n \|w_{-i} - x_{-i}^\star\|^2 \right)^{1/2} \\
&= \left(\sum_{i=1}^n \zeta^2 - \|w_i - x_i^\star\|^2 \right)^{1/2} \\
&= |\zeta| \left(n - \frac{1}{\zeta^2} \|w_i - x_i^\star\|^2 \right)^{1/2} \\
&= |\zeta| \sqrt{n-1}.
\end{aligned}$$

Therefore, we deduce that

$$\mu\|x - x^\star\|^2 \leq L_{\mathsf{a}} \langle u_1, u_2 \rangle \leq L_{\mathsf{a}} \|u_1\| \|u_2\| = L_{\mathsf{a}} \sqrt{n-1} \|w - x^\star\| \|x^\star - x\|$$

so that by dividing through we have

$$\|x - x^\star\| \leq \frac{L_{\mathsf{a}} \sqrt{n-1}}{\mu} \|w - x^\star\|.$$

Since this holds for any w and corresponding $y := \mathsf{BR}(w)$ we have that

$$\|x_{t+1} - x^\star\| \leq \rho \cdot \|x_t - x^\star\| \quad \text{where} \quad \rho := \frac{L_{\mathsf{a}} \sqrt{n-1}}{\mu},$$

as claimed. The rest of the result follows immediately from the Banach fixed point theorem. □

G Proofs for Oblivious Decision-Maker Setting

In this section we prove the bounds on the equilibrium tracking error and the dynamic regret of the agents given an oblivious decision-maker who is deploying a sequence of actions $\{x_s\}_{s=1}^t$ that are of bounded variation and sufficiently contracting.

G.1 Worst-Case Expected Equilibrium Tracking Error.

The first natural question in this setting relates to bounding the time to track the time-varying equilibrium in expectation given the drift $\Delta_t = \|x_t^\star - x_{t+1}^\star\|$ where $x_t^\star \in \mathtt{Eq}(\gamma_{u_t})$ where $\gamma_{u_t} := (f_1^{u_t}, \dots, f_n^{u_t})$ That is, given that the decision-maker is obliviously deploying sequence $\{u_s\}_{s=0}^t$, *how long does it take for* $\|x_t - x_t^\star\|^2$ *to be less than some error tolerance* ε *and how to we optimize that error?*

Assumption 9. There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $(\mathcal{F}_t)_{t\geq 0}$ such that $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Iterates and corresponding Nash equilibrium, $x_t, x_t^\star : \Omega \to \mathbb{R}^m$, are $\mathcal{F}_t$-measurable.

Proposition 6 (Formal Statement of Proposition 1)**.** Suppose that Assumption 9 holds, that agents employ a ρ-contracting update (Definition 1), and we are in the regime $\rho \in [0, 1)$. Then, the expected equilibrium tracking error satisfies

$$\mathbb{E}\|x_t - x_t^\star\|^2 \lesssim \left(1 - \frac{(1-\rho^2)}{2}\right)^t \|x_0 - x_0^\star\|^2 + \frac{(c\sigma_{\mathtt{a}})^2}{1-\rho^2} + \left(\frac{\Delta_{\mathtt{a}}}{1-\rho^2}\right)^2$$

where $\Delta_{\mathtt{a}} := \max\{\|x_{t+1}^\star - x_t^\star\|\}$.

Proof. We have that

$$\mathbb{E}_t\|x_{t+1} - x_t^\star\|^2 \leq \rho^2\|x_t - x_t^\star\| + \rho^2(c\sigma_{\mathtt{a}})^2$$

Now observe that

$$\begin{aligned}
\mathbb{E}_t\|x_{t+1} - x_{t+1}^\star\|^2 &= \|x_{t+1} - x_t^\star\|^2 + \|x_t^\star - x_{t+1}^\star\|^2 + 2\mathbb{E}_t\langle x_{t+1} - x_t^\star, x_t^\star - x_{t+1}^\star\rangle \\
&\leq \mathbb{E}_t\|x_{t+1} - x_t^\star\|^2 + \|x_t^\star - x_{t+1}^\star\|^2 + 2\mathbb{E}_t\|x_{t+1} - x_t^\star\|\|x_t^\star - x_{t+1}^\star\| \\
&\leq (1+\lambda)\,\mathbb{E}_t\|x_{t+1} - x_t^\star\|^2 + \left(1+\lambda^{-1}\right)\|x_t^\star - x_{t+1}^\star\|^2 \\
&\leq (1+\lambda)\left(\rho^2\|x_t - x_t^\star\|^2 + \rho^2(c\sigma_{\mathtt{a}})^2\right) + \left(1+\lambda^{-1}\right)\|x_t^\star - x_{t+1}^\star\|^2,
\end{aligned}$$

where the second to last inequality holds since $\|a\|\|b\| \leq \lambda\|a\|^2 + \lambda^{-1}\|b\|^2$ for any $\lambda > 0$. For algebraic convenience, let $\rho^2 = (1-\tau)$ for some variable $\tau > 0$. Setting $\lambda = \frac{\tau}{2}$, we have that

$$\begin{aligned}
\mathbb{E}_t\|x_{t+1} - x_{t+1}^\star\|^2 &\leq \left(1+\frac{\tau}{2}\right)\left((1-\tau)\|x_t - x_t^\star\|^2 + (1-\tau)(c\sigma_{\mathtt{a}})^2\right) + \left(1+\frac{2}{\tau}\right)\|x_t^\star - x_{t+1}^\star\|^2, \\
&\leq \left(1-\frac{\tau}{2}\right)\|x_t - x_t^\star\|^2 + \left(1-\frac{\tau}{2}\right)(c\sigma_{\mathtt{a}})^2 + \left(1+\frac{2}{\tau}\right)\Delta_{\mathtt{a}}^2
\end{aligned}$$

where $\Delta_{\mathtt{a}}^2 := \max\{\|x_t^\star - x_{t+1}^\star\|^2\}$. Iterating this expression, we have that

$$\begin{aligned}
\mathbb{E}_t\|x_{t+1} - x_{t+1}^\star\|^2 &\leq \left(1-\frac{\tau}{2}\right)\left(\left(1-\frac{\tau}{2}\right)\|x_{t-1} - x_{t-1}^\star\|^2 + \left(1-\frac{\tau}{2}\right)(c\sigma_{\mathtt{a}})^2 + \left(1+\frac{2}{\tau}\right)\Delta_{\mathtt{a}}^2\right) \\
&\quad + \left(1-\frac{\tau}{2}\right)(c\sigma_{\mathtt{a}})^2 + \left(1+\frac{2}{\tau}\right)\Delta_{\mathtt{a}}^2 \\
&\leq \left(1-\frac{\tau}{2}\right)^{t+1}\|x_0 - x_0^\star\|^2 + \sum_{k=1}^{t+1}(c\sigma_{\mathtt{a}})^2\left(1-\frac{\tau}{2}\right)^k + \Delta_{\mathtt{a}}^2\left(1+\frac{2}{\tau}\right)\sum_{k=0}^{t}\left(1-\frac{\tau}{2}\right)^k \\
&\leq \left(1-\frac{\tau}{2}\right)^{t+1}\|x_0 - x_0^\star\|^2 + \frac{2(c\sigma_{\mathtt{a}})^2}{\tau} + \Delta_{\mathtt{a}}^2\left(1+\frac{2}{\tau}\right)\frac{2 - 2(1-\frac{\tau}{2})^t + (1-\frac{\tau}{2})^t\tau}{\tau} \\
&\leq \left(1-\frac{\tau}{2}\right)^{t+1}\|x_0 - x_0^\star\|^2 + \frac{2(c\sigma_{\mathtt{a}})^2}{\tau} + \Delta_{\mathtt{a}}^2\left(1+\frac{2}{\tau}\right)\frac{2}{\tau}.
\end{aligned}$$

Observe that $(1+\frac{2}{w})\frac{2}{w} \leq \frac{2(2+w)}{w^2} \leq \frac{6}{w^2}$ for $w \in (0,1]$. In the regime where $\rho \leq 1$ we have that $\tau = 1-\rho^2 \leq 1$. Hence, we have that

$$\begin{aligned}
\mathbb{E}_t\|x_{t+1} - x_{t+1}^\star\|^2 &\leq \left(1-\frac{\tau}{2}\right)^{t+1}\|x_0 - x_0^\star\|^2 + \frac{2(c\sigma_{\mathtt{a}})^2}{\tau} + \Delta_{\mathtt{a}}^2\left(1+\frac{2}{\tau}\right)\frac{2}{\tau} \\
&\leq \left(1-\frac{\tau}{2}\right)^{t+1}\|x_0 - x_0^\star\|^2 + \frac{2(c\sigma_{\mathtt{a}})^2}{\tau} + 6\left(\frac{\Delta_{\mathtt{a}}}{\tau}\right)^2.
\end{aligned}$$

Algorithm 3 Projected Stochastic Gradient Play

1: **Input:** Step-size $\gamma \le \frac{\mu}{2L_{\mathsf{a}}^2}$; initial condition x_0; decision-maker input sequence $\{u_s\}_{k=0}^t$
2: **for** $k = 1, \dots, t-1$ **do**
3: **for** $i \in [n]$ **do**
4: Set $x_{i,k+1} = \underset{\mathcal{X}_i}{\mathrm{proj}}\,(x_{i,k} - \gamma\widehat{\omega}_{i,k}) \quad \forall\, i \in [n] \quad$ where $\widehat{\omega}_{i,k} := \widehat{\nabla_i} f_i^{u_k}(x_k)$
5: **end for**
6: **end for**

Since $\tau := (1-\rho^2)$, we have that

$$\mathbb{E}_t\|x_{t+1} - x_{t+1}^\star\|^2 \le \left(1 - \frac{(1-\rho^2)}{2}\right)^{t+1}\|x_0 - x_0^\star\|^2 + \frac{2(c\sigma_{\mathsf{a}})^2}{1-\rho^2} + 6\left(\frac{\Delta_{\mathsf{a}}}{1-\rho^2}\right)^2.$$

This concludes the proof. □

Let us specialize to the stochastic gradient play setting. Set $\omega_{i,t} := \nabla_i f_i^{u_t}(x_t)$, $\zeta_{i,t} := \omega_{i,t} - \widehat{\omega}_{i,t}$ and $\zeta_t := (\zeta_{1,t}, \dots, \zeta_{n,t})$. Define also $\omega_t := (\omega_{1,t}, \dots, \omega_{n,t})$. Here ζ_t is the noise of the vector of individual gradients.

Assumption 10. Suppose $\exists \Delta_{\mathsf{a}}, \sigma_{\mathsf{a}} > 0$ such that (a) the drift $\Delta_{\mathsf{a},t} := \|x_{t+1}^\star - x_t^\star\|$ is such that $\mathbb{E}\,\Delta_{\mathsf{a},t}^2 \le \Delta_{\mathsf{a}}^2$ for all t, (b) the gradient noise ζ_t satisfies $\mathbb{E}\,\|\zeta_t\|^2 \le \sigma_{\mathsf{a}}^2$, and (c) the gradient noise $\zeta_t : \Omega \to \mathbb{R}^d$ is $\mathcal{F}_{t+1}$-measurable with $\mathbb{E}[\zeta_t|\mathcal{F}_t] = 0$.

Recall that $\Delta := \max\{\|u_{t+1} - u_t\|\}$ and $\Delta_{\mathsf{a}} \le L_{\mathsf{eq}}\Delta$.

Corollary 6 (Formal Statement of Corollary 1)**.** Under the assumptions of Proposition 1 and Assumption 10, suppose that the agents are running stochastic gradient play (Algorithm 3) with $\gamma \le \mu/(2L_{\mathsf{a}}^2)$. Then $\rho^2 = \frac{1}{1+\gamma\mu}$ and $c = \sqrt{2}\gamma$ so that

$$\mathbb{E}_t\|x_{t+1} - x_{t+1}^\star\|^2 \le \left(1 - \frac{\mu\gamma}{4}\right)^{t+1}\|x_0 - x_0^\star\|^2 + \frac{8\gamma\sigma_{\mathsf{a}}^2}{\mu} + 24\left(\frac{L_{\mathsf{eq}}\Delta}{\gamma\mu}\right)^2.$$

Given Lemma 5, the proof of the above corollary follows an identical argument to Proposition 1.

The preceding corollary shows that in order to obtain last iterate convergence guarantees there is a clear tradeoff between the stepsize and the drift-to-noise ratio. Using Corollary 1, we define the *asymptotic tracking error* of Algorithm 3 and the optimal step-size as follows:

$$\varepsilon_\star := \min_{\gamma\in(0,\mu/(2L_{\mathsf{a}}^2)]}\left\{\frac{8\gamma\sigma_{\mathsf{a}}^2}{\mu} + 24\left(\frac{L_{\mathsf{eq}}\Delta}{\mu\gamma}\right)^2\right\} \quad \text{and} \quad \gamma_\star := \min\left\{\frac{\mu}{2L_{\mathsf{a}}^2}, \left(\frac{6L_{\mathsf{eq}}^2\Delta^2}{\mu\sigma_{\mathsf{a}}^2}\right)^{1/3}\right\}.$$

so that the high and low regimes are determined by

$$\frac{\mu}{2L_{\mathsf{a}}^2} = \left(\frac{6L_{\mathsf{eq}}^2\Delta^2}{\mu\sigma_{\mathsf{a}}^2}\right)^{1/3} \iff \frac{\mu^4}{6\cdot 2^3 L_{\mathsf{a}}^6} = \frac{L_{\mathsf{eq}}^2\Delta^2}{\sigma_{\mathsf{a}}^2} \iff \left(\frac{\mu^4}{3\cdot 2^4 L_{\mathsf{a}}^6 L_{\mathsf{eq}}^2}\right)^{1/2} = \frac{\mu^2}{4\sqrt{3}\cdot L_{\mathsf{a}}^3 L_{\mathsf{eq}}} = \frac{\Delta}{\sigma_{\mathsf{a}}}.$$

Therefore the high drift-to-noise regime is $\frac{\Delta}{\sigma_{\mathsf{a}}} > \frac{\mu^2}{4\sqrt{3}\cdot L_{\mathsf{a}}^3 L_{\mathsf{eq}}}$ and otherwise we are in the low drift-to-noise regime. Plugging $\gamma_\star$ into $\varepsilon_\star$, we have that

$$\varepsilon_\star \asymp \begin{cases} \frac{L_{\mathsf{eq}}^2 L_{\mathsf{a}}^4\Delta^2}{\mu^4} + \frac{\sigma_{\mathsf{a}}^2}{L_{\mathsf{a}}^2} & \text{if } \frac{\Delta}{\sigma_{\mathsf{a}}} \ge \frac{\mu^2}{4\sqrt{3}\cdot L_{\mathsf{eq}} L_{\mathsf{a}}^3}, \\ \left(\frac{L_{\mathsf{eq}}\Delta\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} & \text{otherwise.} \end{cases}$$

Precisely

$$\varepsilon_\star = \begin{cases} 96\frac{L_{\mathsf{eq}}^2 L_{\mathsf{a}}^4\Delta^2}{\mu^4} + 4\frac{\sigma_{\mathsf{a}}^2}{L_{\mathsf{a}}^2} & \text{if } \frac{\Delta}{\sigma_{\mathsf{a}}} \ge \frac{\mu^2}{4\sqrt{3}\cdot L_{\mathsf{eq}} L_{\mathsf{a}}^3}, \\ 12\cdot 6^{1/3}\cdot\left(\frac{L_{\mathsf{eq}}\Delta\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} & \text{otherwise.} \end{cases}$$

High drift-to-noise regime. When $\frac{\Delta}{\sigma_{\mathsf{a}}} \geq \frac{\mu^2}{4\sqrt{3}\cdot L_{\mathsf{eq}} L_{\mathsf{a}}^3}$, we are in the *high drift-to-noise regime.* In this case, then running Algorithm 3 with $\gamma_\star \asymp \mu/(2L_{\mathsf{a}}^2)$ will result in a point x_t such that

$$\mathbb{E}\,\|x_t - x_t^\star\|^2 \lesssim \varepsilon_\star \quad \text{in} \quad t \lesssim \frac{L_{\mathsf{a}}^2}{\mu^2} \log\left(\frac{\|x_0 - x_0^\star\|^2}{\varepsilon}\right) \quad \text{time steps.}$$

This case is less interesting since the decision-maker is deploying a sequence $\{u_k\}_{k=0}^t$ such that the drift (i.e., change in the equilibrium corresponding to the induced sequence of games) is higher than the stochastic noise in the game. Here, the agents then must use a learning rate γ that is as large as the deterministic setting and therefore, achieve a expected tracking error within a constant factor of $\frac{L_{\mathsf{eq}}^2 L_{\mathsf{a}}^4 \Delta^2}{\mu^4} + \frac{\sigma_{\mathsf{a}}^2}{L_{\mathsf{a}}^2}$.

Low drift-to-noise regime. The more interesting regime is when the deployed sequence is resulting in low drift relative to the noise at the lower level—i.e. when $\frac{\Delta}{\sigma_{\mathsf{a}}} < \frac{\mu^2}{4\sqrt{3}\cdot L_{\mathsf{eq}} L_{\mathsf{a}}^2}$. In this case, it is possible that the agents can choose a step-size such that the tracking error is within a constant factor of $\varepsilon_\star$. Indeed, with $\gamma_\star \asymp \left(\frac{6L_{\mathsf{eq}}^2\Delta^2}{\mu\sigma_{\mathsf{a}}^2}\right)^{1/3}$, its straightforward to show that stochastic gradient play (Algorithm 3) finds a point $x_t \in \mathcal{X}$ such that

$$\mathbb{E}\,\|x_t - x_t^\star\|^2 \lesssim \varepsilon_\star \quad \text{in} \quad t \lesssim \frac{\sigma_{\mathsf{a}}^2}{\mu^2\varepsilon_\star} \log\left(\frac{\|x_0 - x_0^\star\|^2}{\varepsilon_\star}\right) \quad \text{time steps.}$$

The following proposition (formal statement of Proposition 2 from the main body) shows that there is an algorithm (a super algorithm that consumes stochastic gradient play) that proceeds in stages to obtain a stronger convergence guarantee.

Proposition 7 (Formal Statement of Proposition 2: Expected Tracking Error for Induced Time-Varying Game.)**.** Suppose that Assumptions 1, 9, and 10 hold and that the decision-maker deploys a sequence $\{u_s\}_{s=0}^t$ satisfying $\frac{\Delta}{\sigma_{\mathsf{a}}} < \frac{\mu^2}{4\sqrt{3}L_{\mathsf{eq}}L_{\mathsf{a}}^3}$. Set constants $\gamma_\star := \left(6L_{\mathsf{eq}}^2\Delta^2/(\mu\sigma_{\mathsf{a}}^2)\right)^{1/3}$ and $\varepsilon_\star := (L_{\mathsf{eq}}\Delta\sigma_{\mathsf{a}}^2/\mu^2)^{2/3}$. Suppose that there is a constant R available such that $R \geq \|x_0 - x_0^\star\|^2$. Further, set constants $K = 1 + \left\lceil \log_2\left(\frac{\mu}{L_{\mathsf{a}}^2}\cdot\left(\frac{\sigma_{\mathsf{a}}^2\mu}{\Delta_{\mathsf{a}}^2}\right)^{1/3}\right)\right\rceil$ and

$$\gamma_0 = \frac{\mu}{2L_{\mathsf{a}}^2}, \quad T_0 = \left\lceil \frac{8L_{\mathsf{a}}^2}{\mu^2}\log\left(\frac{L_{\mathsf{a}}^2 R}{\sigma_{\mathsf{a}}^2}\right)^+\right\rceil, \quad \gamma_k = \frac{\gamma_{k-1} - \gamma_\star}{2}, \quad T_k = \left\lceil \frac{4\log(4)}{\mu\gamma_k}\right\rceil,$$

for all $k \geq 1$. Consider running stochastic gradient play (Algorithm 3) in $k = 0, \ldots, K-1$ stages. Then $T = T_0 + \cdots + T_{K-1}$ satisfies

$$T \lesssim \frac{L_{\mathsf{a}}^2}{\mu^2}\log\left(\frac{L_{\mathsf{a}}^2 R}{\sigma_{\mathsf{a}}^2}\right)^+ + \frac{\sigma_{\mathsf{a}}^2}{\mu^2\varepsilon_\star} \leq \frac{L_{\mathsf{a}}^2}{\mu^2}\log\left(\frac{R}{\varepsilon_\star}\right)^+ + \frac{\sigma_{\mathsf{a}}^2}{\mu^2\varepsilon_\star},$$

and the expected tracking error satisfies $\mathbb{E}\,\|x_K - x_K^\star\|^2 \lesssim \varepsilon_\star$.

In the above corollary $(\cdot)^+ := \max\{(\cdot), 0\}$. We use this operator since some of the logarithmic terms can be negative depending on the size of constants.

Proof. Set $t_0 := 0$ and for each stage index k, let $t_k := \sum_{s=0}^{k-1} T_s$ be the total cumulative time up to stage k. Define $x_k^\star$ be the Nash equilibrium of the induced game $\gamma_{u_{t_k}}$, and set

$$\epsilon_k := \frac{8}{\mu}\left(\gamma_k\sigma_{\mathsf{a}}^2 + 3\frac{(L_{\mathsf{eq}}\Delta)^2}{\mu\gamma_\star^2}\right).$$

Recall that $\gamma_k \geq \gamma_\star$. Corollary 1 implies that

$$\begin{aligned}\mathbb{E}\,\|x_{k+1} - x_{k+1}^\star\|^2 &\leq \left(1 - \frac{\mu\gamma_k}{4}\right)^{T_k}\mathbb{E}\,\|x_k - x_k^\star\|^2 + \frac{8}{\mu}\left(\gamma_k\sigma_{\mathsf{a}}^2 + 3\frac{(L_{\mathsf{eq}}\Delta)^2}{\mu\gamma_k^2}\right)\\ &\leq e^{-\frac{\mu\gamma_k}{4}T_k}\,\mathbb{E}\,\|x_k - x_k^\star\|^2 + \epsilon_k.\end{aligned}$$

We claim that $\mathbb{E}\|x_k - x_k^\star\|^2 \leq 2\epsilon_{k-1}$ for all $k \geq 1$. The argument proceeds by induction. The base case holds since

$$T_0 = \left\lceil \frac{8L_{\mathsf{a}}^2}{\mu^2} \log\left(\frac{L_{\mathsf{a}}^2 R}{\sigma_{\mathsf{a}}^2}\right)^+ \right\rceil$$

implies that

$$\mathbb{E}\|x_1 - x_1^\star\|^2 \leq e^{-\frac{\mu\gamma_0}{4}T_0}\|x_0 - x_0^\star\|^2 + \epsilon_0 \leq \exp\left(-\frac{\mu^2}{4\cdot 2L_{\mathsf{a}}^2}\frac{8L_{\mathsf{a}}^2}{\mu^2}\log\left(\frac{L_{\mathsf{a}}^2R^2}{\sigma_{\mathsf{a}}^2}\right)\right)\|x_0 - x_0^\star\|^2 + \epsilon_0 \leq \frac{\sigma_{\mathsf{a}}^2}{L_{\mathsf{a}}^2} + \epsilon_0 \leq 2\epsilon_0.$$

Suppose that the claim holds for some $k \geq 1$—i.e., $\mathbb{E}\|x_k - x_k^\star\|^2 \leq 2\epsilon_{k-1}$ for some fixed k. Then, we have that $e^{-\mu\gamma_k T_k} = e^{-\mu\gamma_k\left\lceil\frac{4\log(4)}{\mu\gamma_k}\right\rceil} \leq \frac{1}{4}$. Further, its easy to deduce that $\frac{1}{4} \leq \frac{\epsilon_k}{2\epsilon_{k-1}}$. Hence, putting these facts together, we have that

$$\begin{aligned}\mathbb{E}\|x_{k+1} - x_{k+1}^\star\|^2 &\leq e^{-\mu\gamma_k T_k}\,\mathbb{E}\|x_k - x_k^\star\|^2 + \epsilon_k \leq \frac{1}{4}\,\mathbb{E}\|x_k - x_k^\star\|^2 + \epsilon_k \\ &\leq \frac{\epsilon_k}{2\epsilon_{k-1}}\,\mathbb{E}\|x_k - x_k^\star\|^2 + \epsilon_k \leq 2\epsilon_k,\end{aligned}$$

by the induction hypothesis $\mathbb{E}\|x_k - x_k^\star\|^2 \leq 2\epsilon_{k-1}$. In particular, this implies that $\mathbb{E}\|x_K - \bar{x}_K\|^2 \leq 2\epsilon_{K-1}$.

Now, we need to show that the claimed efficiency estimate holds. That is, we need to show that $\epsilon_{K-1} \asymp \varepsilon_\star$. Observe that for some constants c that we will set later, the following is true:

$$\begin{aligned}\epsilon_{k-1} - c\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} &= \frac{8}{\mu}\left(\gamma_{k-1}\sigma_{\mathsf{a}}^2 + 3\frac{\Delta_{\mathsf{a}}^2}{\mu\gamma_\star^2}\right) - c\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} \\ &= \frac{8}{\mu}\left(\gamma_{k-1}\sigma_{\mathsf{a}}^2 + 3\left(\frac{\Delta_{\mathsf{a}}^2}{\mu\sigma_{\mathsf{a}}^2}\right)^{1/3}\frac{\sigma_{\mathsf{a}}^2}{6^{2/3}}\right) - c\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} \\ &= \frac{8\sigma_{\mathsf{a}}^2}{\mu}\left(\gamma_{k-1} + 3\left(\frac{\Delta_{\mathsf{a}}^2}{6^2\sigma_{\mathsf{a}}^2\mu}\right)^{1/3}\right) - c\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} \\ &= \frac{8\sigma_{\mathsf{a}}^2}{\mu}\left(\gamma_{k-1} + 3\left(\frac{\Delta_{\mathsf{a}}^2}{6^2\sigma_{\mathsf{a}}^2\mu}\right)^{1/3} - \frac{c}{8}\left(\frac{\Delta_{\mathsf{a}}^2}{\mu\sigma_{\mathsf{a}}^2}\right)^{1/3}\right).\end{aligned}$$

Thus by setting $c := 12\cdot 6^{1/3}$ we have that

$$\epsilon_{k-1} - 12\cdot 6^{1/3}\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} = \frac{8\sigma_{\mathsf{a}}^2}{\mu}\left(\gamma_{K-1} - \gamma_\star\right) = \frac{8\sigma_{\mathsf{a}}^2}{\mu}\cdot\frac{\gamma_0 - \gamma_\star}{2^{K-1}} \leq 4\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} = \varepsilon_\star.$$

Indeed, this inequality holds since

$$\begin{aligned}\frac{8\sigma_{\mathsf{a}}^2}{\mu}\cdot\frac{\gamma_0 - \gamma_\star}{2^{K-1}} &= \frac{8\sigma_{\mathsf{a}}^2}{\mu}\cdot\frac{\frac{\mu}{2L_{\mathsf{a}}^2} - \left(6L_{\mathsf{eq}}^2\Delta^2/(\mu\sigma_{\mathsf{a}}^2)\right)^{1/3}}{\left(\frac{\mu}{L_{\mathsf{a}}^2}\cdot\left(\frac{\sigma_{\mathsf{a}}^2\mu}{\Delta_{\mathsf{a}}^2}\right)^{1/3}\right)} \\ &= \frac{8\sigma_{\mathsf{a}}^2L_{\mathsf{a}}^2\Delta_{\mathsf{a}}^{2/3}}{\mu^2\sigma_{\mathsf{a}}^{2/3}\mu^{1/3}}\left(\frac{\mu}{2L_{\mathsf{a}}^2} - \frac{6^{1/3}\Delta_{\mathsf{a}}^{2/3}}{\mu^{1/3}\sigma_{\mathsf{a}}^{2/3}}\right) \\ &= \left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3}\frac{8L_{\mathsf{a}}^2}{\mu}\left(\frac{\mu}{2L_{\mathsf{a}}^2} - \frac{6^{1/3}\Delta_{\mathsf{a}}^{2/3}}{\mu^{1/3}\sigma_{\mathsf{a}}^{2/3}}\right) \\ &= \left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3}\left(4 - \frac{6^{1/3}\cdot 8\cdot L_{\mathsf{a}}^2\Delta_{\mathsf{a}}^{2/3}}{\mu^{4/3}\sigma_{\mathsf{a}}^{2/3}}\right) \\ &\leq 4\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} \\ &\leq \varepsilon_\star,\end{aligned}$$

where we used the fact that

$$K = 1 + \left\lceil \log_2 \left(\frac{\mu}{L_{\mathtt{a}}^2} \cdot \left(\frac{\sigma_{\mathtt{a}}^2 \mu}{\Delta_{\mathtt{a}}^2} \right)^{1/3} \right) \right\rceil.$$

Therefore, we have that

$$\mathbb{E} \left\| x_K - x_K^\star \right\|^2 \leq 2(1 + 12 \cdot 6^{1/3}) \left(\frac{\Delta_{\mathtt{a}} \sigma_{\mathtt{a}}^2}{\mu^2} \right)^{2/3} \asymp \varepsilon_\star.$$

What remains is to show that the total time T satisfies the claimed bound. Recall that we set $K = 1 + \left\lceil \log_2 \left(\frac{\mu}{L_{\mathtt{a}}^2} \cdot \left(\frac{\sigma_{\mathtt{a}}^2 \mu}{\Delta_{\mathtt{a}}^2} \right)^{1/3} \right) \right\rceil$. Observe that

$$T \lesssim \frac{L_{\mathtt{a}}^2}{\mu^2} \log \left(\frac{L_{\mathtt{a}}^2 R}{\sigma_{\mathtt{a}}^2} \right)^+ + \frac{1}{\mu} \sum_{k=1}^{K-1} \frac{1}{\gamma_k}.$$

We need to show that the sum on the left is asymptotically proportional to $\frac{\sigma_{\mathtt{a}}^2}{\mu \varepsilon_\star}$. To this end, observe that

$$\sum_{k=1}^{K-1} \frac{1}{\gamma_k} \leq \frac{2L_{\mathtt{a}}^2}{\mu} \sum_{k=1}^{K-1} 2^k \leq \frac{2L_{\mathtt{a}}^2}{\mu} \cdot 2^K = \frac{2 \cdot 2L_{\mathtt{a}}^2}{\mu} 2^{K-1}.$$

Using the definition of K, we have that

$$2^{K-1} = 2^{\log_2 \left(\frac{\mu}{L_{\mathtt{a}}^2} \cdot \left(\frac{\sigma_{\mathtt{a}}^2 \mu}{\Delta_{\mathtt{a}}^2} \right)^{1/3} \right)} = \frac{\mu}{L_{\mathtt{a}}^2} \cdot \left(\frac{\sigma_{\mathtt{a}}^2 \mu}{\Delta_{\mathtt{a}}^2} \right)^{1/3},$$

Hence, we deduce that

$$\sum_{k=1}^{K-1} \frac{1}{\gamma_k} \leq \frac{2 \cdot 2L_{\mathtt{a}}^2}{\mu} 2^{K-1} \leq \frac{2 \cdot 2L_{\mathtt{a}}^2}{\mu} \cdot \frac{\mu}{L_{\mathtt{a}}^2} \cdot \left(\frac{\sigma_{\mathtt{a}}^2 \mu}{\Delta_{\mathtt{a}}^2} \right)^{1/3} = 4 \left(\frac{\sigma_{\mathtt{a}}^2 \mu}{\Delta_{\mathtt{a}}^2} \right)^{1/3} = \frac{4\sigma_{\mathtt{a}}^2}{\mu} \cdot \left(\frac{\Delta_{\mathtt{a}} \sigma_{\mathtt{a}}^2}{\mu^2} \right)^{-2/3} \asymp \frac{\sigma_{\mathtt{a}}^2}{\mu \varepsilon_\star},$$

as claimed. This completes the proof. □

G.2 Beyond Worst Case Expected Tracking Error

The preceding results provide a "worst-case" bound in the sense that $\Delta = \max\{\|u_{t+1} - u_t\|^2\}$ is the largest difference in the decision maker's actions. Here, we want to understand what happens when we make "reasonable" assumptions on the behavior of $\Delta_t := \|u_{t+1} - u_t\|^2$. For instance, one reasonable assumption is that the decision-maker is employing some stochastic gradient method with a convergence guarantee of the form $\mathbb{E} \|u_t - u^*\|^2 \leq \mathcal{O}((t+1)^{-2a})$. Here u^* might be a locally optimal point for $\mathcal{L}$ or $\operatorname{argmin}_{u \in \mathcal{U}} \mathcal{L}(u)$ given that players are playing a Nash $x^*(u) \in \mathtt{Eq}(\mathcal{G}_u)$ or even some other solution concept—e.g., in Section 4.1 we introduce the notion of a performatively stable Stackelberg equilibrium. Note that $\mathbb{E} \|u_{t+1} - u_t\|^2 \leq \mathcal{O}((t+1)^{-2a})$ means there exists a constant $c_{\mathtt{d}} > 0$ such that $\mathbb{E} \|u_{t+1} - u_t\|^2 \leq \frac{c_{\mathtt{d}}}{(t+1)^{2a}}$.

Proposition 8. Suppose that Assumptions 1, 9, and 10 hold and that the decision-maker deploys a sequence of actions such that $\mathbb{E} \|u_{t+1} - u_t\|^2 \leq \frac{c_{\mathtt{d}}}{(t+1)^{-2a}}$ for some $a \in (0, 1/2]$ and absolute constant $c_{\mathtt{d}} > 0$. Set $\gamma_t = \frac{8}{\mu(t+t_0)^b}$ for some $b \in (0, 1]$ and integer $t_0 \geq 1$ and consider agents running stochastic gradient play with time varying stepsize γ_t. Then, the iterates satisfy

$$\mathbb{E}\|x_t - x_t^\star\|^2 \leq \max\{(1 + t_0)\|x_0 - x_0^\star\|^2, c_{\mathtt{a}}\} \cdot \begin{cases} (t + t_0)^{-b}, & \text{if } \ b > \frac{2}{3}a, \\ (t + t_0)^{b/2 - a}, & \text{otherwise}, \end{cases}$$

where $c_{\mathtt{a}} := \frac{8\sigma_{\mathtt{a}}^2}{\mu^2} + \frac{5L_{\mathtt{eq}}^2 c_{\mathtt{d}}^2}{8}$.

Proof. Let $\Delta_{\mathtt{a},t} := \|x^\star_{t+1} - x^\star_t\|$. We know from the proof of Lemma 5, that stochastic gradient play is ρ-contracting. Moreover, for a fixed u_t which induces $x^\star_t \in \mathtt{Eq}(\mathcal{G}_{u_t})$, we have that

$$\mathbb{E}_t\|x_{t+1} - x^\star_t\|^2 \leq \frac{1}{1+\gamma\mu}\|x_t - x^\star_t\|^2 + \frac{2\gamma^2\sigma_{\mathtt{a}}^2}{1+\gamma\mu}.$$

Now observe that

$$\begin{aligned}
\|x_{t+1} - x^\star_{t+1}\|^2 &= \|x_{t+1} - x^\star_t\|^2 + \|x^\star_t - x^\star_{t+1}\|^2 + 2\langle x_{t+1} - x^\star_t, x^\star_t - x^\star_{t+1}\rangle \\
&\leq \|x_{t+1} - x^\star_t\|^2 + \|x^\star_t - x^\star_{t+1}\|^2 + 2\|x_{t+1} - x^\star_t\|\|x^\star_t - x^\star_{t+1}\| \\
&\leq \left(1 + \frac{\mu\gamma}{4}\right)\|x_{t+1} - x^\star_t\|^2 + \left(1 + \frac{4}{\mu\gamma}\right)\|x^\star_t - x^\star_{t+1}\|^2
\end{aligned}$$

where the last inequality follows from Young's inequality. Since $1 - \frac{\gamma\mu}{1+\gamma\mu} \leq 1 - \frac{\mu\gamma}{2}$, we have that

$$\begin{aligned}
\mathbb{E}_t\|x_{t+1} - x^\star_{t+1}\|^2 &\leq \left(1 + \frac{\gamma\mu}{4}\right)\left(\left(1 - \frac{\gamma\mu}{2}\right)\|x_t - x^\star_t\|^2 + 2\gamma^2\sigma_{\mathtt{a}}^2\left(1 - \frac{\gamma\mu}{2}\right)\right) \\
&\quad + \left(1 + \frac{4}{\gamma\mu}\right)\|x^\star_t - x^\star_{t+1}\|^2 \\
&\leq \left(1 - \frac{\gamma\mu}{4}\right)\|x_t - x^\star_t\|^2 + 2\gamma^2\sigma_{\mathtt{a}}^2\left(1 - \frac{\gamma\mu}{4}\right) + \left(1 + \frac{4}{\gamma\mu}\right)\Delta_{\mathtt{a},t}^2 \\
&\leq \left(1 - \frac{\gamma\mu}{4}\right)\|x_t - x^\star_t\| + 2\gamma^2\sigma_{\mathtt{a}}^2 + \frac{5}{\mu\gamma}L_{\mathtt{eq}}^2\|u_{t+1} - u_t\|^2.
\end{aligned}$$

The agents are engaging in stochastic gradient play given the induced sequence of games $\mathcal{G}_{u_t}$ with $\gamma_t = \frac{8}{\mu(t+t_0)^b}$. Hence, plugging γ_t in to the above bound, we have that

$$\begin{aligned}
\mathbb{E}_t\|x_{t+1} - x^\star_{t+1}\|^2 &\leq \left(1 - \frac{2}{(t+t_0)^b}\right)\|x_t - x^\star_t\|^2 + \frac{8\sigma_{\mathtt{a}}^2}{\mu^2(t+t_0)^{2b}} + \frac{5(t+t_0)^b}{8}L_{\mathtt{eq}}^2\|u_{t+1} - u_t\|^2 \\
&\leq \left(1 - \frac{2}{(t+t_0)^b}\right)\|x_t - x^\star_t\|^2 + \frac{8\sigma_{\mathtt{a}}^2}{\mu^2(t+t_0)^{2b}} + \frac{5(t+t_0)^b}{8}L_{\mathtt{eq}}^2\frac{c_{\mathtt{d}}^2}{(t+t_0)^{2a}} \\
&\leq \left(1 - \frac{2}{(t+t_0)^b}\right)\|x_t - x^\star_t\|^2 + \frac{8\sigma_{\mathtt{a}}^2}{\mu^2(t+t_0)^{2b}} + \frac{5}{8}L_{\mathtt{eq}}^2\frac{c_{\mathtt{d}}^2}{(t+t_0)^{2a-b}}.
\end{aligned}$$

Define $D_t := \mathbb{E}\|x_t - x^\star_t\|$. Then there are two cases to analyze.

Case 1: If $b > \frac{2}{3}a$, then the above bound reduces to

$$\mathbb{E}_t\|x_{t+1} - x^\star_{t+1}\|^2 \leq \left(1 - \frac{2}{(t+t_0)^b}\right)\|x_t - x^\star_t\|^2 + \frac{c_{\mathtt{a}}}{(t+t_0)^{2b}}$$

Then we claim that

$$D_t \leq \frac{\max\{(1+t_0)D_0, c_{\mathtt{a}}\}}{(t+t_0)^b}.$$

Indeed, it clearly holds for $t = 0$. Hence we may use induction to conclude the argument. Suppose it holds for some fixed $t \geq 1$. Then, we have that

$$\begin{aligned}
D_{t+1} &\leq \left(1 - \frac{2}{(t+t_0)^b}\right)\frac{c_{\mathtt{a}}}{(t+t_0)^b} + \frac{c_{\mathtt{a}}}{(t+t_0)^{2b}} \\
&\leq c_{\mathtt{a}}\left(\frac{1}{(t+t_0)^b} - \frac{2}{(t+t_0)^{2b}}\right) + \frac{c_{\mathtt{a}}}{(t+t_0)^{2b}} \\
&\leq c_{\mathtt{a}}\left(\frac{1}{(t+t_0)^b} - \frac{1}{(t+t_0)^{2b}}\right) \\
&\leq \frac{c_{\mathtt{a}}}{(t+1+t_0)^b},
\end{aligned}$$

where the last inequality holds since $\frac{1}{(t+t_0)^b} - \frac{1}{(t+t_0)^{2b}} \leq \left(\frac{1}{(t+t_0)} - \frac{1}{(t+t_0)^2}\right)^b \leq \frac{1}{(t+1+t_0)^b}$ for any $t \geq 1$ and $b \in (0, 1]$. This verifies the claim.

Case 2: Suppose now that $b \leq \frac{2}{3}a$. Then the bound on D_{t+1} reduces to

$$D_{t+1} \leq \left(1 - \frac{2}{(t+t_0)^b}\right) D_t + \frac{c_{\mathsf{a}}}{(t+t_0)^{2a-b}} \leq \left(1 - \frac{2}{(t+t_0)^{a-b/2}}\right) D_t + \frac{c_{\mathsf{a}}}{(t+t_0)^{2a-b}},$$

where the last inequality holds since $b \leq \frac{2}{3}a$. Using a completely analogous argument to case 1, we have that

$$D_t \leq \frac{\max\{(1+t_0)D_1, c_{\mathsf{a}}\}}{(t+t_0)^{a-b/2}}.$$

Therefore, putting the two cases together, concludes the proof. □

As noted in the main, this proposition shows that if the decision-maker is employing a reasonably well-behaved sequence of actions (i.e., that is stabilizing at a sufficient rate), then the agents can utilize time varying stepsizes to control the drift and obtain a expected tracking error bound that is decaying in time. The rate of decay however highly depends on the behavior of the decision maker's sequence. For instance, if $a = 1/2$, then choosing $b = 1$ leads to a rate of $\mathcal{O}(1/t)$. Here, $a = 1/2$ is not just a reasonable rate for a stochastic gradient method for the decision-maker as we will see in Section 4, but likely the best we could hope for. However, if the agents choose a much slower rate such as $b < 1/3$, then even with $a = 1/2$ the tracking error decays at a rate of $\mathcal{O}(t^{b/2-1/2})$ so that, somewhat counter intuitively, the rate is much slower as $b \to \frac{1}{3}$. This is because the rate of the decision-maker dominates.

G.3 High-Probability Guarantees

The above results are characterized in terms of the *expected* tracking error; accordingly, characterizing the guarantees of the algorithm are only meaningful if it is run multiple times. Instead, if our algorithm were deployed in real-time with *irreversible* drift, we would like high-probability efficiency results to characterize the performance of our algorithm if it were executed only once. Here, we present high-probability guarantees for the tracking error. We require the following tail assumptions on the equilibrium drift and gradient noise.

Assumption 11 (Sub-Gaussian drift and noise)**.** There exist constants $\Delta_{\mathsf{a}}, \sigma_{\mathsf{a}} > 0$ such that the following two conditions hold for all $t \geq 0$:

(a) The drift $\Delta_{\mathsf{a},t}^2$ is sub-exponential conditioned on $\mathcal{F}_t$ with parameter Δ_{a}^2:

$$\mathbb{E}[\exp(\lambda\Delta_{\mathsf{a},t}^2)|\, \mathcal{F}_t] \leq \exp(\lambda\Delta_{\mathsf{a}}^2) \quad \text{for all} \quad 0 \leq \lambda \leq \Delta_{\mathsf{a}}^2$$

(b) The gradient noise ξ_t is norm sub-Gaussian conditioned on $\mathcal{F}_t$ with parameter $\sigma_a/2$:

$$\mathbb{P}(\|\xi_t\| \geq \zeta|\, \mathcal{F}_t) \leq 2\exp(-2\zeta^2/\sigma_{\mathsf{a}}^2) \quad \text{for all} \quad \zeta > 0.$$

Note that Assumption 11 implies Assumption 10 under the with the same constants Δ_{a}, σ_{a}. We need the following (albeit simplified) proposition from (Cutler et al., 2023), which is an extension of Claim D.1 from (Harvey et al., 2019).

Proposition 3. Consider a scalar stochastic process $\{V_t, D_t, X_t\}$ on a probability space with filtration $\mathcal{H}_t$ such that V_t is nonnegative and $\mathcal{H}_t$-measurable, and satisfies

$$V_{t+1} \leq \alpha_t V_t + X_t$$

for some deterministic constant $\alpha_t \in (-\infty, 1]$. Suppose that the moment generating functions of X_t conditioned on $\mathcal{H}_t$ satisfies

$$\mathbb{E}[\exp(\lambda X_t)|\, \mathcal{H}_t] \leq \exp(\lambda\nu_t) \quad \forall\ 0 \leq \lambda \leq 1/\nu_t,$$

for some constants $\sigma_t, \nu_t > 0$. Then the inequality

$$\mathbb{E}[\exp(\lambda V_{t+1})] \leq \exp(\lambda \cdot \nu_t)\mathbb{E}\left[\exp\left(\lambda \alpha_t V_t\right)\right],$$

holds for all $0 \leq \lambda \leq \frac{1}{2\nu_t}$.

Proof. For any index t and scalar $\lambda \geq 0$, the tower rule of expectations implies that

$$\mathbb{E}[\exp(\lambda V_{t+1})] \leq \mathbb{E}[\exp(\lambda(\alpha_t V_t + X_t)] = \mathbb{E}\left[\exp(\lambda \alpha_t V_t)\, \mathbb{E}[\exp(\lambda X_t)|\ \mathcal{H}_t]\right].$$

By assumption, we have that $\mathbb{E}[\exp(\lambda X_t)|\ \mathcal{H}_t] \leq \exp(\lambda \nu_t)$ for $0 \leq \lambda \leq \frac{1}{2\nu_t}$. Thus, we have that

$$\mathbb{E}[\exp(\lambda V_{t+1})] \leq \mathbb{E}\left[\exp(\lambda \alpha_t V_t)\, \mathbb{E}[\exp(\lambda X_t)|\ \mathcal{H}_t]\right] \leq \exp(\lambda v_t)\, \mathbb{E}[\exp(\lambda \alpha_t)],$$

which completes the proof. □

Given this proposition, we have the following high probability bounds.

Theorem 1 (High probability tracking error.)**.** Suppose that Assumptions 1, 9, and 11 hold and that the decision-maker deploys a sequence $\{u_s\}_{s=0}^t$ satisfying $\frac{\Delta}{\sigma_{\mathtt{a}}} < \frac{\mu^2}{4\sqrt{3}L_{\mathtt{eq}}L_{\mathtt{a}}^3}$ so that the agents are in the low drift-to-noise regime. Let $\{x_t\}$ be the iterates produced by the agents running stochastic gradient play (Algorithm 3) with $\gamma \leq \frac{\mu}{2L_{\mathtt{a}}^2}$. Then there exists an absolute constant $c > 0$ such that for any specified $t \in \mathbb{N}$ and $\delta \in (0, 1)$, the following estimate holds with probability at least $1 - \delta$:

$$\|x_t - x_t^\star\|^2 \leq \left(1 - \frac{\mu\gamma}{4}\right)^t \|x_0 - x_0^\star\|^2 + c\left(\frac{\sigma_{\mathtt{a}}^2\gamma}{\mu} + \left(\frac{\Delta_{\mathtt{a}}}{\mu\gamma}\right)^2\right)\log\left(\frac{e}{\delta}\right). \tag{5}$$

Proof. By Young's inequality, we have that

$$\begin{aligned}
\|x_{t+1} - x_{t+1}^\star\|^2 &= \|x_{t+1} - x_t^\star\|^2 + \|x_t^\star - x_{t+1}^\star\|^2 + 2\langle x_{t+1} - x_t^\star, x_t^\star - x_{t+1}^\star\rangle \\
&\leq \|x_{t+1} - x_t^\star\|^2 + \|x_t^\star - x_{t+1}^\star\|^2 + 2\|x_{t+1} - x_t^\star\|\|x_t^\star - x_{t+1}^\star\| \\
&\leq (1+\lambda)\, \|x_{t+1} - x_t^\star\|^2 + \left(1 + \lambda^{-1}\right) \|x_t^\star - x_{t+1}^\star\|^2
\end{aligned}$$

for some λ. Observe that $x \mapsto \frac{1}{2}\|x_t - \gamma\widehat{\omega}(x_t) - x\|^2$ is a 1-strongly convex function over $\mathcal{X}$. Hence we deduce that

$$\begin{aligned}
\frac{1}{2}\|x_{t+1} - x_t^\star\|^2 &\leq \frac{1}{2}\|x_t - \gamma\widehat{\omega}(x_t) - x_t^\star\|^2 - \frac{1}{2}\|x_t - \gamma\widehat{\omega}(x_t) - x_{t+1}\|^2 \\
&\leq \frac{1}{2}\|x_t - x_t^\star\|^2 - \gamma\langle\widehat{\omega}(x_t), x_{t+1} - x_t^\star\rangle - \frac{1}{2}\|x_{t+1} - x_t\|^2 \\
&= \frac{1}{2}\|x_t - x_t^\star\|^2 - \gamma\langle\widehat{\omega}(x_t), x_t - x_t^\star\rangle - \frac{1}{2}\|x_{t+1} - x_t\|^2 - \gamma\langle\widehat{\omega}(x_t), x_{t+1} - x_t\rangle.
\end{aligned}$$

Taking expectations, we have that

$$\begin{aligned}
\frac{1}{2}\|x_{t+1} - x_t^\star\|^2 &\leq \frac{1}{2}\|x_t - x^*\|^2 - \gamma\langle\widehat{\omega}(x_t), x_t - x_t^\star\rangle - \frac{1}{2}\|x_{t+1} - x_t\|^2 - \gamma\langle\widehat{\omega}(x_t), x_{t+1} - x_t\rangle \\
&\leq \frac{1}{2}\|x_t - x_t^\star\|^2 - \gamma\langle\omega(x_t), x_t - x_t^\star\rangle - \frac{1}{2}\|x_{t+1} - x^*\|^2 - \gamma\langle\widehat{\omega}(x_t), x_{t+1} - x_t\rangle \\
&= \frac{1}{2}\|x_t - x_t^\star\|^2 - \gamma\langle\omega(x_{t+1}), x_{t+1} - x_t^\star\rangle - \frac{1}{2}\|x_{t+1} - x_t\|^2 \\
&\quad + \gamma\underbrace{\langle\widehat{\omega}(x_t) - \omega(x_t), x_t - x_{t+1}\rangle}_{=:P_1} + \gamma\underbrace{\langle\omega(x_t) - \omega(x_{t+1}), x_t^\star - x_{t+1}\rangle}_{=:P_2}.
\end{aligned}$$

Since each induced game $\mathcal{G}_u$ is is μ–strongly monotone, we have that

$$\langle\omega(x_{t+1}), x_{t+1} - x_t^\star\rangle \geq \langle\omega(x_{t+1}) - \omega(x_t^\star), x_{t+1} - x_t^\star\rangle \geq \mu\|x_{t+1} - x_t^\star\|^2.$$

This in turn implies that

$$\frac{1+2\gamma\mu}{2}\|x_{t+1}-x_t^\star\|^2 \le \frac{1}{2}\|x_t-x^*\|^2 - \frac{1}{2}\|x_{t+1}-x_t\|^2 + \gamma(P_1+P_2).$$

Employing Young's inequality, we upper bound P_1 as follows:

$$P_1 \le \frac{\|\xi_t\|^2}{2\nu_1} + \frac{\nu_1\|x_{t+1}-x_t\|^2}{2}.$$

Applying Young's inequality, we bound P_2 as follows:

$$\begin{aligned} P_2 &\le \frac{\|\omega(x_t)-\omega(x_{t+1})\|^2}{2\nu_2} + \frac{\nu_2\|x_{t+1}-x_t^\star\|^2}{2}, \\ &\le \frac{L_{\mathsf{a}}^2\|x_t-x_{t+1}\|^2}{2\nu_2} + \frac{\nu_2\|x_{t+1}-x_t^\star\|^2}{2}, \end{aligned}$$

so that

$$\frac{1+2\gamma\mu-\gamma\nu_2}{2}\|x_{t+1}-x_t^\star\|^2 \le \frac{1}{2}\|x_t-x_t^\star\|^2 + \frac{\|\xi_t\|^2}{2\nu_1} - \frac{(1-\gamma L_{\mathsf{a}}^2\nu_2^{-1}-\gamma\nu_1)}{2}\|x_{t+1}-x_t\|^2.$$

Setting $\nu_2=\mu$ and $\nu_1=\gamma^{-1}-L_{\mathsf{a}}^2/\mu$, we have that the last term on the right hand side is zero, and since $\gamma \le \frac{\mu}{2L_{\mathsf{a}}^2}$ we have that $\nu_1 \ge \frac{1}{2\gamma}$; indeed, $-\frac{1}{2\eta} \le -\frac{2L_{\mathsf{a}}^2}{\mu}$ so that $\nu_1 = \frac{1}{\eta} - \frac{2L_{\mathsf{a}}^2}{\mu} \ge \frac{1}{\eta} - \frac{1}{2\eta} = \frac{1}{2\eta}$. Therefore

$$\|x_{t+1}-x_t^\star\|^2 \le \frac{1}{1+\gamma\mu}\|x_t-x^*\|^2 + \frac{2\gamma^2\|\xi_t\|^2}{1+\mu\gamma}.$$

Thus, we have that

$$\begin{aligned} \|x_{t+1}-x_{t+1}^\star\|^2 &= \|x_{t+1}-x_t^\star\|^2 + \|x_t^\star-x_{t+1}^\star\|^2 + 2\langle x_{t+1}-x_t^\star, x_t^\star-x_{t+1}^\star\rangle \\ &\le \|x_{t+1}-x_t^\star\|^2 + \|x_t^\star-x_{t+1}^\star\|^2 + 2\|x_{t+1}-x_t^\star\|\|x_t^\star-x_{t+1}^\star\| \\ &\le (1+\lambda)\left(\frac{1}{1+\gamma\mu}\|x_t-x_t^\star\|^2 + \frac{2\gamma^2\|\xi_t\|^2}{1+\mu\gamma}\right) + \left(1+\lambda^{-1}\right)\|x_t^\star-x_{t+1}^\star\|^2 \\ &\le (1+\lambda)\left(\left(1-\frac{\mu\gamma}{2}\right)\|x_t-x_t^\star\|^2 + 2\left(1-\frac{\mu\gamma}{2}\right)\gamma^2\|\xi_t\|^2\right) + \left(1+\lambda^{-1}\right)\|x_t^\star-x_{t+1}^\star\|^2 \\ &\le \left(1-\frac{\mu\gamma}{4}\right)\|x_t-x_t^\star\|^2 + 2\left(1-\frac{\mu\gamma}{4}\right)\gamma^2\|\xi_t\|^2 + \left(1+\frac{4}{\mu\gamma}\right)\Delta_{\mathsf{a},t}^2, \end{aligned}$$

where we have set $\lambda = \frac{\mu\gamma}{4}$. Bounding the last two terms, we have that

$$\|x_{t+1}-x_{t+1}^\star\|^2 \le \left(1-\frac{\mu\gamma}{4}\right)\|x_t-x_t^\star\|^2 + 2\gamma^2\|\xi_t\|^2 + \frac{5}{\mu\gamma}\Delta_{\mathsf{a},t}^2. \tag{15}$$

Under Assumption 11, there exists an absolute constant $c \ge 1$ such that $\|\xi_t\|^2$ is sub-exponential conditioned on $\mathcal{F}_t$ with parameter $c\sigma_{\mathsf{a}}^2$ and ξ_t is mean-zero sub-Gaussian conditioned on $\mathcal{F}_t$ with parameter $c\sigma_{\mathsf{a}}$ for all t (Jin et al., 2019, Lemma 3). Thus $\Delta_{\mathsf{a},t}^2$ is sub-exponential conditioned on $\mathcal{F}_t$ with parameter Δ_{a}^2 by Assumption 11. Given (15), we apply Proposition 3 with parameters $V_t = \|x_t-x_t^\star\|^2$, $D_t=0$, $X_t = 2\gamma^2\|\xi_t\|^2 + \frac{5}{\mu\gamma}\Delta_{\mathsf{a},t}^2$, $\alpha_t = 1-\frac{\mu\gamma}{4}$, $\kappa_t=0$, and $\nu_t = 2\gamma^2c\sigma_{\mathsf{a}}^2 + 5\Delta_{\mathsf{a}}^2/(\mu\gamma)$. This yields the estimate

$$\mathbb{E}[\exp(\lambda\|x_{t+1}-x_{t+1}^\star\|^2)] \le \exp\left(\lambda\left(2\gamma^2c\sigma_{\mathsf{a}}^2 + \frac{5\Delta_{\mathsf{a}}^2}{\mu\gamma}\right)\right)\mathbb{E}\left[\exp\left(\lambda\left(1-\frac{\mu\gamma}{8}\right)\|x_t-x_t^\star\|^2\right)\right], \tag{16}$$

for all

$$0 \le \lambda \le \frac{1}{2(2\gamma^2c\sigma_{\mathsf{a}}^2 + 5\Delta_{\mathsf{a}}^2/(\mu\gamma))}. \tag{17}$$

Now, since $\gamma_t \equiv \gamma$ and Assumption 11 holds, iterating (16), we have that

$$\begin{aligned}
\mathbb{E}\left[\exp(\lambda\|x_t - x_t^\star\|^2)\right] &\le \exp(\lambda\nu)\exp\left(\lambda\left(\left(1-\frac{\mu\gamma}{4}\right)^t\|x_0 - x_0^\star\|^2 + \left(2\gamma^2\sigma_{\mathsf{a}}^2 + \frac{5\Delta_{\mathsf{a}}^2}{\mu\gamma}\right)\sum_{s=1}^{t-1}\left(1-\frac{\mu\gamma}{4}\right)^s\right)\right. \\
&\le \exp\left(\lambda\left(\left(1-\frac{\mu\gamma}{4}\right)^t\|x_0 - x_0^\star\|^2 + \left(2\gamma^2\sigma_{\mathsf{a}}^2 + \frac{5\Delta_{\mathsf{a}}^2}{\mu\gamma}\right)\sum_{s=0}^{t-1}\left(1-\frac{\mu\gamma}{4}\right)^s\right)\right. \\
&\le \exp\left(\lambda\left(\left(1-\frac{\mu\gamma}{4}\right)^t\|x_0 - x_0^\star\|^2 + \left(\frac{8\gamma\sigma_{\mathsf{a}}^2}{\mu} + \frac{20\Delta_{\mathsf{a}}^2}{(\mu\gamma)^2}\right)\right)\right.
\end{aligned}$$

for all λ satisfying (17). Let $\nu := \frac{32(c\sigma_{\mathsf{a}})^2\gamma}{\mu} + 20\left(\frac{\Delta_{\mathsf{a}}}{\mu\gamma}\right)^2$. Recall that $c \ge 1$ and $\mu\gamma \le 1$ so that

$$\left(\frac{8\gamma\sigma_{\mathsf{a}}^2}{\mu} + \frac{20\Delta_{\mathsf{a}}^2}{(\mu\gamma)^2}\right) \le \nu$$

and

$$\frac{1}{\nu} = \frac{\mu}{32\gamma(c\sigma_{\mathsf{a}})^2 + 20\Delta_{\mathsf{a}}^2/(\mu\gamma^2)} \le \min\left\{\frac{\mu}{32\cdot c^2\gamma\sigma_{\mathsf{a}}^2}, \frac{1}{2(2\gamma^2 c\sigma_{\mathsf{a}}^2 + 5\Delta_{\mathsf{a}}^2/(\mu\gamma))}\right\}.$$

Hence, we have that

$$\mathbb{E}\left[\exp\left(\lambda\left(\|x_t - x_t^\star\|^2 - \left(1-\frac{\mu\gamma}{4}\right)^t\|x_0 - x_0^\star\|^2\right)\right)\right] \le \exp(\lambda\nu) \quad \forall\ \ 0 \le \lambda \le \frac{1}{\nu}.$$

Rewriting this expression, we have that

$$\frac{\mathbb{E}\left[\exp\left(\lambda\left(\|x_t - x_t^\star\|^2 - (1-\frac{\mu\gamma}{4})^t\|x_0 - x_0^\star\|^2\right)\right)\right]}{\exp(\lambda\nu)} \le 1.$$

Applying Markov's inequality, we have that

$$\begin{aligned}
&\Pr\left(\exp\left(\lambda\left(\|x_t - x_t^\star\|^2 - \left(1-\frac{\mu\gamma}{4}\right)^t\|x_0 - x_0^\star\|^2\right)\right) \ge \frac{\exp(\lambda\nu)}{\delta}\right) \\
&\le \frac{\mathbb{E}\left[\exp\left(\lambda\left(\|x_t - x_t^\star\|^2 - (1-\frac{\mu\gamma}{4})^t\|x_0 - x_0^\star\|^2\right)\right)\right]}{\exp(\lambda\nu)/\delta} \le \delta
\end{aligned}$$

Therefore, setting $\lambda = \frac{1}{\nu}$, with probability $1-\delta$, we have that

$$\|x_t - x_t^\star\|^2 \le \left(1-\frac{\mu\gamma}{4}\right)^t\|x_0 - x_0^\star\|^2 + \left(\frac{32(c\sigma_{\mathsf{a}})^2\gamma}{\mu} + 20\left(\frac{\Delta_{\mathsf{a}}}{\mu\gamma}\right)^2\right)\log\left(\frac{e}{\delta}\right),$$

as claimed. □

The above theorem can be translated to a time-to-track high probability result.

Corollary 2 (Time to track with high probability.)**.** Suppose that the assumptions of Theorem 1 hold so that we are in the low drift-to-noise regime. Let $\{x_t\}$ be the iterates produced by the agents running stochastic gradient play (Algorithm 3) with $\gamma \le \frac{\mu}{2L_{\mathsf{a}}^2}$. Moreover suppose that we have Moreover, suppose that there is a constant R available such that $R \ge \|x_0 - x_0^\star\|^2$. Further, set constants

$$K = 1 + \left\lceil \log_2\left(\frac{\mu}{L_{\mathsf{a}}^2}\cdot\left(\frac{\sigma_{\mathsf{a}}^2\mu}{\Delta_{\mathsf{a}}^2}\right)^{1/3}\right)\right\rceil$$

and

$$\gamma_0 = \frac{\mu}{2L_{\mathsf{a}}^2}, \quad T_0 = \left\lceil \frac{8L_{\mathsf{a}}^2}{\mu^2} \log\left(\frac{L_{\mathsf{a}}^2 R}{\sigma_{\mathsf{a}}^2}\right)^+ \right\rceil, \quad \gamma_k = \frac{\gamma_{k-1}+\gamma_\star}{2}, \quad T_k = \left\lceil \frac{4\log(12)}{\mu\gamma_k} \right\rceil, \quad \text{for all } \; k \geq 1.$$

Consider running stochastic gradient play (Algorithm 3) in $k = 0, \ldots, K-1$ stages. Then $T = T_0 + \cdots + T_{K-1}$ satisfies

$$T \lesssim \frac{L_{\mathsf{a}}^2}{\mu^2} \log\left(\frac{L_{\mathsf{a}}^2 R}{\sigma_{\mathsf{a}}^2}\right)^+ + \frac{\sigma_{\mathsf{a}}^2}{\mu^2 \varepsilon_\star} \leq \frac{L_{\mathsf{a}}^2}{\mu^2} \log\left(\frac{R}{\varepsilon_\star}\right)^+ + \frac{\sigma_{\mathsf{a}}^2}{\mu^2\varepsilon_\star},$$

and for any given $\delta \in (0,1)$, the expected tracking error satisfies $\mathbb{E}\,\|x_K - x_K^\star\|^2 \lesssim \varepsilon_\star \log\left(\frac{e}{\delta}\right)$ with probability at least $1-\delta$.

Proof. Set $t_0 := 0$. For each k, let $t_k := T_0 + \cdots + T_{k-1}$, let $x_t^\star$ be the Nash equilibrium of the game $\mathcal{G}_{u_t}$, and set

$$E_k := c\left(\frac{\gamma_k \sigma_{\mathsf{a}}^2}{\mu} + \left(\frac{\Delta_{\mathsf{a}}}{\mu\gamma_\star}\right)^2\right)$$

where $c \geq 1$ is an absolute constant satisfying the bound (5). Since $\gamma_k \geq \gamma_\star$, Theorem 1 implies that for any specified index k and $\delta \in (0,1)$, the following estimate holds with probability at least $1-\delta$:

$$\begin{aligned}
\|x_{k+1} - x_{k+1}^\star\|^2 &\leq \left(1 - \frac{\mu\gamma_k}{4}\right)^{T_k} \|x_0 - x_0^\star\|^2 + c\left(\frac{\sigma_{\mathsf{a}}^2\gamma_k}{\mu} + \left(\frac{\Delta_{\mathsf{a}}}{\mu\gamma_k}\right)^2\right)\log\left(\frac{e}{\delta}\right), \\
&\leq e^{-\mu\gamma_k T_k/4}\|x_k - x_k^\star\|^2 + E_k \log\left(\frac{e}{\delta}\right).
\end{aligned}$$

We claim that an induction-based argument yields the following: for each $k \geq 1$, the estimate $\|x_k - x_k^*\|^2 \leq AE_{k-1}\log(e/\delta)$ holds with probability at least $1-\delta$ for all $\delta \in (0,1)$. To see the base case, observe that

$$e^{-\mu\gamma_0 T_0/4}\|x_0 - x_0^\star\|^2 \leq \exp\left(-\frac{\mu^2}{4\cdot 2L_{\mathsf{a}}^2}\frac{8L_{\mathsf{a}}^2}{\mu^2}\log\left(\frac{L_{\mathsf{a}}^2 R}{\sigma_{\mathsf{a}}^2}\right)\right)\|x_0 - x_0^\star\|^2 = \frac{\sigma_{\mathsf{a}}^2}{L_{\mathsf{a}}^2 R}\|x_0 - x_0^\star\|^2 \leq \frac{\sigma_{\mathsf{a}}^2}{L_{\mathsf{a}}^2}$$

and

$$E_0 = c\left(\frac{\gamma_0\sigma_{\mathsf{a}}^2}{\mu} + \left(\frac{\Delta_{\mathsf{a}}}{\mu\gamma_\star}\right)^2\right) \geq c\left(\frac{\sigma_{\mathsf{a}}^2}{2L_{\mathsf{a}}^2} + \left(\frac{\Delta_{\mathsf{a}}}{\mu\gamma_0}\right)^2\right) \geq c\left(\frac{3\sigma_{\mathsf{a}}^2}{4L_{\mathsf{a}}^2}\right) \geq \left(\frac{3\sigma_{\mathsf{a}}^2}{4L_{\mathsf{a}}^2}\right)$$

since $\gamma_k \geq \gamma_\star$ and $c \geq 1$, and we are in the regime where $\frac{\Delta_{\mathsf{a}}}{\sigma_{\mathsf{a}}} < \frac{\mu^2}{4L_{\mathsf{a}}^3}$. Therefore we have that

$$e^{-\mu\gamma_0 T_0/4}\|x_0 - x_0^\star\|^2 \leq \frac{\sigma_{\mathsf{a}}^2}{L_{\mathsf{a}}^2} \leq \frac{4}{3}E_0.$$

Hence

$$\|x_1 - x_1^\star\|^2 \leq e^{-\mu\gamma_0 T_0/4}\|x_0 - x_0^\star\|^2 + E_0\log\left(\frac{e}{\delta}\right) \leq \frac{7}{3}E_0\log\left(\frac{e}{\delta}\right) \leq 3E_0\log\left(\frac{e}{\delta}\right)$$

since $\log(e/\delta) \geq 1$, and where we take the bound in the last inequality to simplify constants.

Now, suppose the claim holds for some index $k \geq 1$ and let $\delta \in (0,1)$; then $\|x_k - x_k^\star\|^2 \leq 3E_{k-1}\log(2e/\delta)$ with probability at least $1-\delta/2$. Since

$$e^{-\mu\gamma_k T_k/4} \leq \exp\left(-\mu\gamma_k\left\lceil\frac{4\log(12)}{\mu\gamma_k}\right\rceil\cdot\frac{1}{4}\right) \leq \frac{1}{12},$$

we also have that

$$\begin{aligned}
\|x_{k+1} - x_{k+1}\|^2 &\leq e^{-\mu\gamma_k T_k/4}\|x_k - x_k^\star\|^2 + E_k\log\left(\frac{2e}{\delta}\right), \\
&\leq \frac{1}{12}\|x_k - x_k^\star\|^2 + E_k\log\left(\frac{2e}{\delta}\right) \\
&\leq \frac{E_k}{6E_{k-1}} + E_k\log\left(\frac{2e}{\delta}\right), \\
&\leq \frac{3}{6}E_k\log\left(\frac{2e}{\delta}\right) + E_k\log\left(\frac{2e}{\delta}\right)
\end{aligned}$$

with probability at least $1-\delta/2$. Taking a union bound, we have that

$$\|x_{k+1}-x_{k+1}\|^2 \le \frac{3}{2}E_k \log\left(\frac{2e}{\delta}\right) \le 3E_k \log\left(\frac{e}{\delta}\right),$$

with probability at least $1-\delta$. This completes the inductive proof. Hence, fixing $\delta \in (0,1)$, we have that $\|x_K - x_K^\star\|^2 \le 3E_{K-1}\log(e/\delta)$ with probability at least $1-\delta$.

Observe that for some constants C that we will set later, the following is true: Recall that we are in the regime where $\frac{\Delta}{\sigma_{\mathsf{a}}} < \frac{\mu^2}{4\sqrt{3}L_{\mathrm{eq}}L_{\mathsf{a}}^3}$ and we have set constants $\gamma_\star := \left(6\Delta_{\mathsf{a}}^2/(\mu\sigma_{\mathsf{a}}^2)\right)^{1/3}$ and $\varepsilon_\star := (\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2/\mu^2)^{2/3}$.

$$\begin{aligned}
\frac{2}{c}E_{k-1} - C\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} &= \frac{2}{\mu}\left(\gamma_{k-1}\sigma_{\mathsf{a}}^2 + \frac{\Delta_{\mathsf{a}}^2}{\mu\gamma_\star^2}\right) - C\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} \\
&= \frac{2}{\mu}\left(\gamma_{k-1}\sigma_{\mathsf{a}}^2 + \left(\frac{\Delta_{\mathsf{a}}^2}{\mu\sigma_{\mathsf{a}}^2}\right)^{1/3}\frac{\sigma_{\mathsf{a}}^2}{6^{2/3}}\right) - C\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} \\
&= \frac{2\sigma_{\mathsf{a}}^2}{\mu}\left(\gamma_{k-1} + \left(\frac{\Delta_{\mathsf{a}}^2}{6^2\sigma_{\mathsf{a}}^2\mu}\right)^{1/3}\right) - C\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} \\
&= \frac{2\sigma_{\mathsf{a}}^2}{\mu}\left(\gamma_{k-1} + \left(\frac{\Delta_{\mathsf{a}}^2}{6^2\sigma_{\mathsf{a}}^2\mu}\right)^{1/3} - \frac{C}{2}\left(\frac{\Delta_{\mathsf{a}}^2}{\mu\sigma_{\mathsf{a}}^2}\right)^{1/3}\right).
\end{aligned}$$

Thus by setting $C := 7\cdot\left(\frac{2}{9}\right)^{1/3}$ we have that

$$\frac{2}{c}E_{K-1} - 7\cdot\left(\frac{2}{9}\right)^{1/3}\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} = \frac{2\sigma_{\mathsf{a}}^2}{\mu}\left(\gamma_{K-1}-\gamma_\star\right) = \frac{2\sigma_{\mathsf{a}}^2}{\mu}\cdot\frac{\gamma_0-\gamma_\star}{2^{K-1}} \le \left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3} = \varepsilon_\star.$$

Therefore, we have that

$$E_{K-1} \le \frac{c}{2}\left(1+7\cdot\left(\frac{2}{9}\right)^{1/3}\right)\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3}\log\left(\frac{e}{\delta}\right)$$

so that

$$\mathbb{E}\,\|x_K - x_K^\star\|^2 \le 3E_{K-1}\log\left(\frac{e}{\delta}\right) \le \frac{3\cdot c}{2}\left(1+7\cdot\left(\frac{2}{9}\right)^{1/3}\right)\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{2/3}\log\left(\frac{e}{\delta}\right) \asymp \varepsilon_\star\log\left(\frac{e}{\delta}\right).$$

What remains is to show that the total time T satisfies the claimed bound. Recall that we set $K = 1+\left\lceil \log_2\left(\frac{\mu}{L_{\mathsf{a}}^2}\cdot\left(\frac{\sigma_{\mathsf{a}}^2\mu}{\Delta_{\mathsf{a}}^2}\right)^{1/3}\right)\right\rceil$. Observe that

$$T \lesssim \frac{L_{\mathsf{a}}^2}{\mu^2}\log\left(\frac{L_{\mathsf{a}}^2 R}{\sigma_{\mathsf{a}}^2}\right)^+ + \frac{1}{\mu}\sum_{k=1}^{K-1}\frac{1}{\gamma_k}.$$

We need to show that the sum on the left is asymptotically proportional to $\frac{\sigma_{\mathsf{a}}^2}{\mu\varepsilon_\star}$. To this end, observe that

$$\sum_{k=1}^{K-1}\frac{1}{\gamma_k} \le \frac{2L_{\mathsf{a}}^2}{\mu}\sum_{k=1}^{K-1}2^k \le \frac{2L_{\mathsf{a}}^2}{\mu}\cdot 2^K = \frac{2\cdot 2L_{\mathsf{a}}^2}{\mu}2^{K-1}.$$

Using the definition of K, we have that

$$2^{K-1} = 2^{\log_2\left(\frac{\mu}{L_{\mathsf{a}}^2}\cdot\left(\frac{\sigma_{\mathsf{a}}^2\mu}{\Delta_{\mathsf{a}}^2}\right)^{1/3}\right)} = \frac{\mu}{L_{\mathsf{a}}^2}\cdot\left(\frac{\sigma_{\mathsf{a}}^2\mu}{\Delta_{\mathsf{a}}^2}\right)^{1/3},$$

Hence, we deduce that

$$\sum_{k=1}^{K-1}\frac{1}{\gamma_k} \le \frac{2\cdot 2L_{\mathsf{a}}^2}{\mu}2^{K-1} \le \frac{2\cdot 2L_{\mathsf{a}}^2}{\mu}\cdot\frac{\mu}{L_{\mathsf{a}}^2}\cdot\left(\frac{\sigma_{\mathsf{a}}^2\mu}{\Delta_{\mathsf{a}}^2}\right)^{1/3} = 4\left(\frac{\sigma_{\mathsf{a}}^2\mu}{\Delta_{\mathsf{a}}^2}\right)^{1/3} = \frac{4\sigma_{\mathsf{a}}^2}{\mu}\cdot\left(\frac{\Delta_{\mathsf{a}}\sigma_{\mathsf{a}}^2}{\mu^2}\right)^{-2/3} \asymp \frac{\sigma_{\mathsf{a}}^2}{\mu\varepsilon_\star},$$

as claimed. This completes the proof. $\square$

Algorithm 4 Epoch-Based Algorithm Framework for Stochastic Stackelberg Games

Inputdecision maker algorithm $\texttt{Alg}_{\texttt{dm}}$, horizon T, stepsize schedule $\{\eta_t\}$, initial parameter $u_1 \in \mathcal{U}$, and query radius $\delta > 0$ (if $\texttt{Alg}_{\texttt{dm}} = \texttt{DFM}$)
for $t = 1, \dots, T$ **do**
  if $\texttt{Alg}_{\texttt{dm}} = \texttt{DFM}$ **then** ▷ i.e., derivative-free method
    Sample $v_t \sim \mathbb{S}^d$ uniformly at random
    Set $\tilde{u}_t = u_t + \delta v_t$ and $\widetilde{\mathcal{U}} = (1-\delta)\mathcal{U}$
  else if $\texttt{Alg}_{\texttt{dm}} = \texttt{RGM}$ **then** ▷ i.e., repeated gradient method
    Set $\tilde{u}_t = u_t$ and $\widetilde{\mathcal{U}} = \mathcal{U}$
  end if
  for $k = 1, \dots, \tau_t$ **do**
    Query agents with $\tilde{u}_t$ ▷ i.e., agents update with any ρ contracting method
  end for
  Decision-maker observes $x_t^{\tau_t}(\tilde{u}_t)$
  if $\texttt{Alg}_{\texttt{dm}} = \texttt{DFM}$ **then**
    Set $\widehat{g}_t = \frac{d}{\delta}\ell(\tilde{u}_t, x_t^{\tau_t}(\tilde{u}_t) + \xi)v_t$
  else if $\texttt{Alg}_{\texttt{dm}} = \texttt{RGM}$ **then**
    Set $\widehat{g}_t = \nabla_u \ell(\tilde{u}_t, x_t^{\tau_t}(\tilde{u}_t) + \xi)$
  end if
  Update $u_{t+1} = \operatorname{proj}_{\widetilde{\mathcal{U}}}(u_t - \eta_t \widehat{g}_t)$
end for

H Proofs for the Naïve Decision-Maker Setting

In this appendix section, we provide the formal statements for the results in Section 4.1 and the proofs. Algorithm 4 provides a more detailed version of the brief algorithm (Algorithm 1) given in the main body.

H.1 Existence of Performatively Stable Stackelberg Equilibrium

Recall that performatively stable Stackelberg equilibrium are precisely the fixed points of the map $\texttt{Stack}(u') := \{u \in \mathcal{U} :\ u \text{ is optimal for } \mathbb{E}_{\xi\sim\mathcal{D}_o}\, \ell(u, x^*(u') + \xi) \text{ and } x^*(u') \in \texttt{Eq}(\mathcal{G}_{u'})\}$.

Theorem 2 (Existence & Uniqueness of Performatively Stable Stackelberg Equilibrium)**.** Under Assumptions 1 and 3 and when $1 < \alpha/(L_{\texttt{eq}} L_\ell)$, there exists a unique performatively stable Stackelberg equilibrium.

Proof. We show that $\texttt{Stack}(\cdot)$ is Lipschitz continuous with parameter λ. Since the induced game $\mathcal{G}_u$ is μ strongly monotone for any $u \in \mathcal{U}$ by assumption, there is a unique induced Nash equilibrium $x^*(u) \in \texttt{Eq}(\mathcal{G}_u)$ for each u. Consider two points u and u' and set $w := \texttt{Stack}(u)$ and $w' := \texttt{Stack}(u')$. First order optimality conditions for w and w' guarantee

$$\langle g_u(w), w - w' \rangle \le 0 \quad \text{and} \quad \langle g_{u'}(w'), w' - w \rangle \le 0,$$

where $g_v(v') = \mathbb{E}_{\xi\sim\mathcal{D}_o} \nabla_u \ell(v', x^*(v) + \xi)$. Since the loss $\ell(\cdot, x)$ is α-strongly convex for any $x \in \mathcal{X}$, we have that

$$\begin{aligned}
\alpha \cdot \|w - w'\|^2 &\le \langle g_u(w) - g_u(w'), w - w' \rangle \\
&\le \langle g_{u'}(w') - g_u(w'), w - w' \rangle \\
&\le \|g_{u'}(w') - g_u(w')\| \cdot \|w - w'\| \\
&= \left\| \mathop{\mathbb{E}}_{\xi\sim\mathcal{D}_o} \nabla_u \ell(w', x^*(u') + \xi) - \mathop{\mathbb{E}}_{\xi\sim\mathcal{D}_o} \nabla_u \ell(w', x^*(u) + \xi) \right\| \cdot \|w - w'\| \\
&\le L_\ell L_{\texttt{eq}} \|u' - u\| \cdot \|w - w'\|.
\end{aligned}$$

Dividing through by $\|w - w'\|$ guarantees

$$\|w - w'\| = \|\texttt{Stack}(u) - \texttt{Stack}(u')\| \le \frac{L_\ell L_{\texttt{eq}}}{\alpha}\|u - u'\|.$$

In the regime where $L_{\texttt{eq}} < \frac{\alpha}{L_\ell}$, then $\lambda \in [0,1)$ so that that $\texttt{Stack}(\cdot)$ is indeed a contraction as claimed. The result follows immediately from the Banach fixed point theorem. □

H.2 Characterization of Performative Gap

Below we prove the claimed bound on the performative gap.

Proof of Proposition 4. First observe that

$$\|u^* - u^{\texttt{ps}}\| + \|x^*(u^*) - x^*(u^{\texttt{ps}})\| \le (1 + L_{\texttt{eq}})\|u^* - u^{\texttt{ps}}\|. \tag{18}$$

Next, since ℓ is α–strongly convex in u, we have that

$$\begin{aligned}
\alpha\|u^* - u^{\texttt{ps}}\|^2 &\le \langle G_{u^{\texttt{ps}}}(u^*) - G_{u^{\texttt{ps}}}(u^{\texttt{ps}}), u^* - u^{\texttt{ps}}\rangle \\
&\le \langle G_{u^{\texttt{ps}}}(u^*), u^* - u^{\texttt{ps}}\rangle \\
&\le \langle G_{u^{\texttt{ps}}}(u^*) - G_{u^*}(u^*) + G_{u^*}(u^*), u^* - u^{\texttt{ps}}\rangle \\
&\le \langle G_{u^*}(u^*), u^* - u^{\texttt{ps}}\rangle + \|\nabla_u \ell(u^*, x^*(u^{\texttt{ps}})) - \nabla_u \ell(u^*, x^*(u^*))\|\|u^* - u^{\texttt{ps}}\| \\
&\le \langle G_{u^*}(u^*), u^* - u^{\texttt{ps}}\rangle + L_\ell L_{\texttt{eq}}\|u^* - u^{\texttt{ps}}\|.
\end{aligned}$$

Now, since u^* is optimal for $\min_u \mathcal{L}(u)$, we have that

$$0 \in G_{u^*}(u^*) + \mathop{\mathbb{E}}_{\xi\sim\mathcal{D}_o} \nabla_x \ell(u^*, x^*(u^*) + \xi) \cdot \nabla x^*(u^*) + N_{\mathcal{U}}(u^*).$$

By Lipschitz conitnuity of ℓ in z and of $x^*(\cdot)$, we have that $\|\frac{d}{dv}\mathbb{E}_{z\sim\mathcal{D}(v)} \ell(u^*, z)|_{v=u^*}\| \le L_{\texttt{z}} L_{\texttt{eq}}$. Therefore combining the above results we have that

$$\|u^* - u^{\texttt{ps}}\| \le \frac{L_{\texttt{z}} L_{\texttt{eq}}}{\alpha - L_\ell L_{\texttt{eq}}}$$

so that

$$\|u^* - u^{\texttt{ps}}\| + \|x^*(u^*) - x^*(u^{\texttt{ps}})\| \le (1 + L_{\texttt{eq}})\frac{L_{\texttt{z}} L_{\texttt{eq}}}{\alpha - L_\ell L_{\texttt{eq}}},$$

as claimed. □

H.2.1 Technical Lemmas

For both the naïve and strategic settings, we will need the following technical lemma on the behavior of the stochastic agents play. Recall that in each epoch t the agents initialize their algorithm at $x_t^0 := x_{t-1}^\tau$ and that, by an abuse of notation, $x_0 = x_0^0$.

Recall that when the agents' algorithms are deterministic, Definition 1 reduces to $\|x_t^{k+1} - x^*(u_t)\|^2 \le \rho^2\|x_t^k - x^*(u_t)\|^2$, so that

$$\|x_t^{k+1} - x^*(u_t)\| \le \rho\|x_t^k - x^*(u_t)\|.$$

The following lemma will be used in the proof of Theorem 3 when the agents are deterministic.

Lemma 7 (Deterministic Agent Contraction)**.** Suppose that the decision-maker is running Algorithm 4 with $\texttt{Alg} := \texttt{RGM}$ using step-size η and under the assumption that $\sup_{(u,x)\in\mathcal{U}\times\mathcal{X}} \|\nabla_u \ell(u, x)\| \le L_{\texttt{u}}$. Further, suppose agents use a ρ-contracting update (Definition 1) with $\rho \in [0, 1)$ and $\sigma_{\texttt{a}} = 0$. Under Assumption 1, the following bound holds:

$$\|x_t^\tau(u_t) - x^*(u_t)\| \le \rho^\tau \left(\rho^{t-1}\|x_0 - x^*(u_0)\| + \frac{\eta L_{\texttt{eq}} L_{\texttt{u}}}{1 - \rho}\right).$$

Proof. Given Definition 1, we have that

$$\|x_t^\tau(u_t) - x^*(u_t)\| \le \rho \|x_t^{\tau-1}(u_t) - x^*(u_t)\|.$$

Iterating this expression we have that

$$\|x_t^\tau(u_t) - x^*(u_t)\| \le \rho^\tau \|x_{t-1}^\tau(u_{t-1}) - x^*(u_t)\|$$

Adding and subtracting appropriate terms we have that

$$\begin{aligned}\|x_t^\tau(u_t) - x^*(u_t)\| &\le \rho^\tau \|x_{t-1}^\tau(u_{t-1}) - x^*(u_{t-1}) + x^*(u_{t-1}) - x^*(u_t)\| \\ &\le \rho^\tau \|x_{t-1}^\tau(u_{t-1}) - x^*(u_{t-1})\| + \rho^\tau L_{\mathsf{eq}} \|u_{t-1} - u_t\|\end{aligned}$$

Continuing in this fashion we have that

$$\begin{aligned}\|x_t^\tau(u_t) - x^*(u_t)\| &\le \rho^\tau \|x_{t-1}^\tau(u_{t-1}) - x^*(u_{t-1})\| + \rho^\tau L_{\mathsf{eq}} \|u_{t-1} - u_t\| \\ &\le \rho^\tau(\rho^\tau \|x_{t-2}^\tau(u_{t-2}) - x^*(u_{t-2})\| + \rho^\tau L_{\mathsf{eq}} \|u_{t-2} - u_{t-1}\|) + \rho^\tau L_{\mathsf{eq}} \|u_{t-1} - u_t\| \\ &\le \rho^\tau \rho^{t-1} \|x_0 - x^*(u_0)\| + L_{\mathsf{eq}} \rho^\tau \sum_{s=1}^{t} \rho^s \|u_{t-s} - u_{t-s-1}\| \\ &\le \rho^\tau \rho^{t-1} \|x_0 - x^*(u_0)\| + L_{\mathsf{eq}} L_{\mathsf{u}} \eta \frac{\rho^\tau}{1-\rho}\end{aligned}$$

where in the second to last inequality we use the fact that $\rho^\tau \le \rho$ for any $\tau \ge 1$, and in the last inequality we use the fact that $u_t = u_{t-1} - \eta \nabla_u \ell(u_{t-1}, z_{t-1})$ and $\sup_{(u,x)\in\mathcal{U}\times\mathcal{X}} \|\nabla_u \ell(u,x)\| \le L_{\mathsf{u}}$. ☐

H.2.2 Naïve Decision-Maker: Deterministic Agents with Stochastic Observations

Given the preceding technical lemma, we know prove Theorem 3. Let's us restate it more formally.

Theorem 5 (Formal Statement of Theorem 3)**.** Suppose that Assumptions 1, 3, and 4 hold, that

$$\sup_{(u,x)\in\mathcal{U}\times\mathcal{X}} \mathbb{E}[\|\nabla_u \ell(u, x+\xi)\|] \le L_{\mathsf{u}},$$

and that we have available a constant $R > \|x_0 - x^*(u_0)\|$. Further, suppose the decision-maker runs Algorithm 4 with $\texttt{Alg} := \texttt{RGM}$ using step-size $\eta \le \frac{\alpha}{4L_\ell^2(1+L_{\mathsf{eq}}^2)}$, and the agents employ a ρ–contracting algorithm $\mathcal{A}$ with $\rho \in [0,1)$. Consider the following two cases:

- **Case 1:** The agents employ deterministic algorithms (i.e., $\sigma_{\mathsf{a}} = 0$) and the decision-maker receives a noisy observation $\mathcal{A}(x_{t-1}, u_t) + \xi$ in each round where ξ is zero mean and finite variance. In this case, set the epoch length such that $\tau \ge \log\left(\frac{2L_\ell^2}{\alpha\eta\sigma^2}(\rho^{t-1}R + \frac{\eta L_{\mathsf{eq}} L_{\mathsf{u}}}{1-\rho})^2\right)\frac{1}{\log(1/\rho^2)}$;
- **Case 2:** The agents employ stochastic gradient play with $\sigma_{\mathsf{a}} \in (0,\infty)$ run stage-wise via Algorithm 2. In this case, set the epoch length to

$$\tau = \sum_{k=0}^{K} T_k = \left\lceil \left(1 + \frac{2L_{\mathsf{a}}^2}{\mu^2}\right) \log\left(\frac{2R}{\epsilon_\tau}\right) \right\rceil + \sum_{k=1}^{K} \left\lceil \left(1 + \frac{2^{k+1}L_{\mathsf{a}}^2}{\mu^2}\right) \log(4) \right\rceil, \tag{19}$$

  and tolerance $\epsilon_\tau = \eta^2\sigma^2$ where $K = \left\lceil 1 + \log_2\left(\frac{\sigma_{\mathsf{a}}^2}{\epsilon_\tau L_{\mathsf{a}}^2}\right)\right\rceil$.

Then the following estimate holds:

$$\mathbb{E}_t \|u_{t+1} - u^{\mathsf{ps}}\|^2 \le \left(1 - \frac{\alpha\eta}{2}\right)^t \|u_0 - u^{\mathsf{ps}}\|^2 + \frac{4\eta\sigma^2}{\alpha}.$$

Recall from Corollary 5, that if the agents run stochastic gradient play in stages then we are able to characterize precisely the number of iterations required to hit a particular specified error tolerance. This is where the epoch length in (24) is derived.

Proof of Theorem 3. The proofs for the two cases start out similarly. Once they deviate, we will break the proof into cases. Define the following objects:

$$\begin{aligned} g_t &:= \nabla_u \ell(u_t, x_t^\tau(u_t) + \xi), \text{ where } \xi \sim \mathcal{D}_o; \\ G_t(u_t) &:= \mathop{\mathbb{E}}_{\xi\sim\mathcal{D}_o} \nabla_u \ell(u_t, x_t^\tau(u_t) + \xi); \\ G_{\mathtt{ps}}(u_t) &:= \mathop{\mathbb{E}}_{\xi\sim\mathcal{D}_o} \nabla_u \ell(u_t, x^*(u_t) + \xi). \end{aligned}$$

Also note that $\mathbb{E}_t[g_t] = G_t(u_t)$—i.e., the gradient estimate g_t is an unbiased estimate of the time varying expected gradient G_t—and

$$u^{\mathtt{ps}} = \operatorname*{argmin}_{u\in\mathcal{U}} \mathop{\mathbb{E}}_{z\sim\mathcal{D}(u^{\mathtt{ps}})} \ell(u,z) \quad \text{so that} \quad \langle G_{\mathtt{ps}}(u^{\mathtt{ps}}), u - u^{\mathtt{ps}}\rangle \geq 0 \ \ \forall\, u \in \mathcal{U}.$$

Fix two constants $\nu_1, \nu_2 > 0$ to be specified later. Noting that u_{t+1} is the minimizer of the 1-strongly convex function $u \mapsto \frac{1}{2}\|u_t - \eta g_t - u\|^2$ over $\mathcal{U}$, we deduce that

$$\frac{1}{2}\|u_{t+1} - u^{\mathtt{ps}}\|^2 \leq \frac{1}{2}\|u_t - \eta g_t - u^{\mathtt{ps}}\|^2 - \frac{1}{2}\|u_t - \eta_t g_t - u_{t+1}\|^2.$$

Expanding the squares on the right hand side and combining terms yields

$$\begin{aligned} \frac{1}{2}\|u_{t+1} - u^{\mathtt{ps}}\|^2 &\leq \frac{1}{2}\|u_t - u^{\mathtt{ps}}\|^2 - \eta_t\langle g_t, u_{t+1} - u^{\mathtt{ps}}\rangle - \frac{1}{2}\|u_{t+1} - u_t\|^2 \\ &= \frac{1}{2}\|u_t - u^{\mathtt{ps}}\|^2 - \eta\langle g_t, u_t - u^{\mathtt{ps}}\rangle - \frac{1}{2}\|u_{t+1} - u_t\|^2 - \eta\langle g_t, u_{t+1} - u_t\rangle. \end{aligned}$$

Using the fact that $\mathbb{E}_t[g_t] = G_t(u_t)$, we successively compute

$$\begin{aligned} \frac{1}{2}\mathbb{E}_t\|u_{t+1} - u^{\mathtt{ps}}\|^2 &\leq \frac{1}{2}\|u_t - u^{\mathtt{ps}}\|^2 - \eta\langle \mathbb{E}_t g_t, u_t - u^{\mathtt{ps}}\rangle - \frac{1}{2}\mathbb{E}_t\|u_{t+1} - u_t\|^2 - \eta\mathbb{E}_t\langle g_t, u_{t+1} - u_t\rangle, \\ &\leq \frac{1}{2}\|u_t - u^{\mathtt{ps}}\|^2 - \eta\langle G_t(u_t), u_t - u^{\mathtt{ps}}\rangle - \frac{1}{2}\mathbb{E}_t\|u_{t+1} - u_t\|^2 - \eta\mathbb{E}_t\langle g_t, u_{t+1} - u_t\rangle, \\ &= \frac{1}{2}\|u_t - u^{\mathtt{ps}}\|^2 - \eta\mathbb{E}_t\langle G_{\mathtt{ps}}(u_{t+1}), u_{t+1} - u^{\mathtt{ps}}\rangle - \frac{1}{2}\mathbb{E}_t\|u_{t+1} - u_t\|^2 \\ &\quad + \eta\underbrace{\mathbb{E}_t\langle g_t - G_t(u_t), u_t - u_{t+1}\rangle}_{P_1} + \eta_t\underbrace{\mathbb{E}_t\langle G_t(u_t) - G_{\mathtt{ps}}(u_{t+1}), u^{\mathtt{ps}} - u_{t+1}\rangle}_{P_2}. \end{aligned}$$

Recall that for any z, the loss $\ell(u,z)$ is α–strongly convex in u so that $\langle G_{\mathtt{ps}}(u_{t+1}), u_{t+1} - u^{\mathtt{ps}}\rangle \geq \alpha\|u_{t+1} - u^{\mathtt{ps}}\|^2$. Therefore we have that

$$\frac{1+2\eta\alpha}{2}\mathbb{E}_t\|u_{t+1} - u^{\mathtt{ps}}\|^2 \leq \frac{1}{2}\|u_t - u^{\mathtt{ps}}\|^2 - \frac{1}{2}\mathbb{E}_t\|u_{t+1} - u_t\|^2 + \eta(P_1 + P_2). \tag{20}$$

Applying Young's inequality to P_1, we have that

$$P_1 \leq \frac{\mathbb{E}_t\|g_t - G_t(u_t)\|^2}{2\nu_1} + \frac{\nu_1\mathbb{E}_t\|u_{t+1} - u_t\|^2}{2} \leq \frac{\sigma^2}{2\nu_1} + \frac{\nu_1\mathbb{E}_t\|u_{t+1} - u_t\|^2}{2}. \tag{21}$$

Then we have for P_2, the upper bound

$$\begin{aligned} P_2 &\leq \frac{\mathbb{E}_t\|G_t(u_t) - G_{\mathtt{ps}}(u_{t+1})\|^2}{2\nu_2} + \frac{\nu_2\mathbb{E}_t\|u^{\mathtt{ps}} - u_{t+1}\|^2}{2} \\ &\leq \frac{2\mathbb{E}_t\|G_t(u_t) - G_{\mathtt{ps}}(u_t)\|^2 + 2\mathbb{E}_t\|G_{\mathtt{ps}}(u_t) - G_{\mathtt{ps}}(u_{t+1})\|^2}{2\nu_2} + \frac{\nu_2\mathbb{E}_t\|u^{\mathtt{ps}} - u_{t+1}\|^2}{2} \\ &\leq \frac{2\mathbb{E}_t\|G_t(u_t) - G_{\mathtt{ps}}(u_t)\|^2 + 2L_\ell^2(1 + L_{\mathtt{eq}}^2)\|u_t - u_{t+1}\|^2}{2\nu_2} + \frac{\nu_2\mathbb{E}_t\|u^{\mathtt{ps}} - u_{t+1}\|^2}{2}. \end{aligned}$$

Next observe that

$$\begin{aligned} \mathbb{E}_t\|G_t(u_t) - G_{\mathtt{ps}}(u_t)\|^2 &= \mathbb{E}_t\|\mathbb{E}_\xi\nabla_u\ell(u_t, \mathcal{A}(x_{t-1}, u_t) + \xi) - \mathbb{E}_\xi\nabla_u\ell(u_t, x^*(u_t) + \xi)\|^2 \\ &\leq L_\ell^2\mathbb{E}_t\|\mathcal{A}(x_{t-1}, u_t) - x^*(u_t)\|^2. \end{aligned} \tag{22}$$

Now we break the proof into the two cases.

Proof for Case 1. Given our assumption on the stochastic contractive dynamics of the followers, by Lemma 7, we have that

$$\begin{aligned}\mathbb{E}_t\|G_t(u_t) - G_{\mathtt{ps}}(u_t)\|^2 &\le L_\ell \mathbb{E}_t\|\mathcal{A}(x_{t-1}, u_t) - x^*(u_t)\|^2 \\ &\le L_\ell^2 \rho^{2\cdot\tau}\left(\rho^{t-1}\|x_0 - x^*(u_0)\| + \frac{\eta L_{\mathtt{eq}} L_{\mathtt{u}}}{1-\rho}\right)^2 \\ &\le L_\ell^2 \rho^{2\cdot\tau}\underbrace{\left(\rho\|x_0 - x^*(u_0)\| + \frac{\eta L_{\mathtt{eq}} L_{\mathtt{u}}}{1-\rho}\right)^2}_{:=C^2}\end{aligned}$$

Therefore

$$P_2 \le \frac{L_\ell^2}{\nu_2}\left(2\rho^{2\tau}C^2\right) + \frac{2L_\ell^2(1+L_{\mathtt{eq}}^2)\|u_t - u_{t+1}\|^2}{2\nu_2} + \frac{\nu_2 \mathbb{E}_t\|u^{\mathtt{ps}} - u_{t+1}\|^2}{2}.$$

Coming back to the bound in (20), we have that

$$\begin{aligned}\frac{1+2\eta\alpha}{2}\mathbb{E}_t\|u_{t+1} - u^{\mathtt{ps}}\|^2 &\le \frac{1}{2}\|u_t - u^{\mathtt{ps}}\|^2 - \frac{1}{2}\mathbb{E}_t\|u_{t+1} - u_t\|^2 + \eta\left(\frac{\sigma^2}{2\nu_1} + \frac{\nu_1 \mathbb{E}_t\|u_{t+1} - u_t\|^2}{2}\right) \\ &\quad + \eta\left(\frac{L_\ell^2}{\nu_2}\left(2\rho^{2\tau}C^2\right) + \frac{L_\ell^2(1+L_{\mathtt{eq}}^2)\,\mathbb{E}_t\|u_t - u_{t+1}\|^2}{\nu_2} + \frac{\nu_2 \mathbb{E}_t\|u^{\mathtt{ps}} - u_{t+1}\|^2}{2}\right),\end{aligned}$$

so that

$$\begin{aligned}\frac{1+2\eta\alpha - \eta\nu_2}{2}\mathbb{E}_t\|u_{t+1} - u^{\mathtt{ps}}\|^2 &\le \frac{1}{2}\|u_t - u^{\mathtt{ps}}\|^2 + \frac{\eta\sigma^2}{2\nu_1} + \eta\frac{2L_\ell^2\rho^{2\tau}C^2}{\nu_2} \\ &\quad - \frac{1 - 2L_\ell^2(1+L_{\mathtt{eq}}^2)\eta\nu_2^{-1} - \eta\nu_1}{2}\mathbb{E}_t\|u_{t+1} - u_t\|^2.\end{aligned}$$

Letting $\nu_1 = \eta^{-1} - \frac{2L_\ell^2(1+L_{\mathtt{eq}}^2)}{\alpha}$ and $\nu_2 = \alpha$ ensures that the last term on the right is zero. By our assumption that $\eta \le \frac{\alpha}{4L_\ell^2(1+L_{\mathtt{eq}}^2)}$ we have that $\frac{1}{\eta} \ge \frac{4L_\ell^2(1+L_{\mathtt{eq}}^2)}{\alpha}$ so that $\nu_1 \ge \frac{1}{2\eta}$; indeed,

$$\nu_1 = \eta^{-1} - \frac{2L_\ell^2(1+L_{\mathtt{eq}}^2)}{\alpha} \ge \frac{4L_\ell^2(1+L_{\mathtt{eq}}^2)}{\alpha} - \frac{2L_\ell^2(1+L_{\mathtt{eq}}^2)}{\alpha} = \frac{2L_\ell^2(1+L_{\mathtt{eq}}^2)}{\alpha} = \frac{1}{2\eta}.$$

Hence we have that

$$\frac{1+\eta\alpha}{2}\mathbb{E}_t\|u_{t+1} - u^{\mathtt{ps}}\|^2 \le \frac{1}{2}\|u_t - u^{\mathtt{ps}}\|^2 + \eta^2\sigma^2 + \eta\frac{2L_\ell^2\rho^{2\tau_t}C_t^2}{\alpha}$$

Now, choose τ as stated in the theorem to ensure that $\eta\rho^{2\tau}\frac{2L_\ell^2C^2}{\alpha} \le \eta^2\sigma^2$. Indeed, this inequality is equivalent to

$$\tau\log\rho^2 \le \log\left(\frac{\alpha\eta\sigma^2}{2L_\ell^2C^2}\right) \iff \tau \ge \log\left(\frac{2L_\ell^2C^2}{\alpha\eta\sigma^2}\right)\frac{1}{\log(1/\rho^2)},$$

which is precisely the stated lower bound on τ. Hence, we have that

$$\mathbb{E}_t\|u_{t+1} - u^{\mathtt{ps}}\|^2 \le \frac{1}{1+\eta\alpha}\|u_t - u^{\mathtt{ps}}\|^2 + \frac{4}{1+\eta\alpha}\eta^2\sigma^2.$$

Recursively iterating the above expression, we have that

$$\begin{aligned}\mathbb{E}_t\|u_{t+1} - u^{\mathtt{ps}}\|^2 &\le \frac{1}{1+\eta\alpha}\left(\frac{1}{1+\eta\alpha}(\|u_{t-1} - u^{\mathtt{ps}}\|^2 + \frac{4}{1+\eta\alpha}\eta^2\sigma^2\right) + \frac{4}{1+\eta\alpha}\eta^2\sigma^2 \\ &\le \left(\frac{1}{1+\eta\alpha}\right)^t\|u_0 - u^{\mathtt{ps}}\|^2 + 4\eta^2\sigma^2\sum_{s=1}^{t}\left(\frac{1}{1+\eta\alpha}\right)^t \\ &\le \left(\frac{1}{1+\eta\alpha}\right)^t\|u_0 - u^{\mathtt{ps}}\|^2 + 4\eta^2\sigma^2\frac{1}{\eta\alpha}\end{aligned}$$

Given the choice of $\eta \leq \frac{\alpha}{4L_\ell^2(1+L_{\text{eq}}^2)}$, we have that

$$\mathbb{E}_t\|u_{t+1} - u^{\text{ps}}\|^2 \leq \left(1 - \frac{\alpha\eta}{2}\right)^t \|u_0 - u^{\text{ps}}\|^2 + \frac{4\eta\sigma^2}{\alpha}.$$

This completes the proof for **Case 1**.

Proof for Case 2. Now we switch to **Case 2.** The proof starts from the bound in (22). Recall that by the assumption that agents are running stage-based stochastic gradient play (i.e., Algorithm 2 with $\mathcal{A}$ as stochastic gradient play). Hence, we have that $\mathbb{E}_t\|x_t^\tau(u_t) - x^*(u_t)\|^2 \leq \epsilon_\tau$ where

$$\tau = \sum_{k=0}^{K} T_k = \left\lceil \left(1 + \frac{2L_{\mathsf{a}}^2}{\mu^2}\right) \log\left(\frac{2R}{\epsilon_\tau}\right)\right\rceil + \sum_{k=1}^{K} \left\lceil \left(1 + \frac{2^{k+1}L_{\mathsf{a}}^2}{\mu^2}\right) \log(4)\right\rceil,$$

and $K = \left\lceil 1 + \log_2\left(\frac{\sigma_{\mathsf{a}}^2}{L_{\mathsf{a}}^2\epsilon_\tau}\right)\right\rceil$. Moreover, the decision-maker has set $\epsilon_\tau = \eta^2\sigma^2$. Therefore, we deduce that

$$P_2 \leq \frac{L_\ell^2\epsilon_\tau}{\nu_2} + \frac{2L_\ell^2\|u_t - u_{t+1}\|^2}{2\nu_2} + \frac{\nu_2\mathbb{E}_t\|x^{\text{ps}} - u_{t+1}\|^2}{2}.$$

Coming back to the bound in (20), we have that

$$\begin{aligned}\frac{1+2\eta\alpha}{2}\mathbb{E}_t\|u_{t+1} - x^{\text{ps}}\|^2 \leq \frac{1}{2}\|u_t - x^{\text{ps}}\|^2 - \frac{1}{2}\mathbb{E}_t\|u_{t+1} - u_t\|^2 + \eta\left(\frac{\sigma^2}{2\nu_1} + \frac{\nu_1\mathbb{E}_t\|u_{t+1} - u_t\|^2}{2}\right)\\ + \eta\left(\frac{L_\ell^2\epsilon_\tau}{\nu_2} + \frac{2L_\ell^2\|u_t - u_{t+1}\|^2}{2\nu_2} + \frac{\nu_2\mathbb{E}_t\|x^{\text{ps}} - u_{t+1}\|^2}{2}\right)\end{aligned}$$

so that

$$\frac{1+2\eta\alpha - \eta\nu_2}{2}\mathbb{E}_t\|u_{t+1} - x^{\text{ps}}\|^2 \leq \frac{1}{2}\|u_t - x^{\text{ps}}\|^2 + \frac{\eta\sigma^2}{2\nu_1} + \eta\frac{L_\ell^2\epsilon_\tau}{\nu_2} - \frac{1 - 2L_\ell^2\eta\nu_2^{-1} - \eta\nu_1}{2}\mathbb{E}_t\|u_{t+1} - u_t\|^2$$

Letting $\nu_1 = \eta^{-1} - \frac{2L_\ell^2}{\alpha}$ and $\nu_2 = \alpha$ ensures that the last term on the right is zero. By our assumption that $\eta \leq \frac{\alpha}{4L_\ell^2}$ we have that $\frac{1}{\eta} \geq \frac{4L_\ell^2}{\alpha}$ so that $\nu_1 \geq \frac{1}{2\eta}$; indeed,

$$\nu_1 = \eta^{-1} - \frac{2L_\ell^2}{\alpha} \geq \frac{4L_\ell^2}{\alpha} - \frac{2L_\ell^2}{\alpha} = \frac{2L_\ell^2}{\alpha} = \frac{1}{2\eta}.$$

Hence we have that

$$\frac{1+\eta\alpha}{2}\mathbb{E}_t\|u_{t+1} - x^{\text{ps}}\|^2 \leq \frac{1}{2}\|u_t - x^{\text{ps}}\|^2 + \eta^2\sigma^2 + \frac{\epsilon_\tau}{2}$$

Then since $\epsilon_\tau = \eta^2\sigma^2$, we have that

$$\mathbb{E}_t\|u_{t+1} - x^{\text{ps}}\|^2 \leq \frac{1}{1+\eta\alpha}\|u_t - x^{\text{ps}}\|^2 + \frac{4}{1+\eta\alpha}\eta^2\sigma^2$$

Recursively iterating the above expression, we have that

$$\begin{aligned}\mathbb{E}_t\|u_{t+1} - u^{\text{ps}}\|^2 &\leq \frac{1}{1+\eta\alpha}\left(\frac{1}{1+\eta\alpha}(\|u_{t-1} - u^{\text{ps}}\|^2 + \frac{4}{1+\eta\alpha}\eta^2\sigma^2\right) + \frac{4}{1+\eta\alpha}\eta^2\sigma^2\\ &\leq \left(\frac{1}{1+\eta\alpha}\right)^t \|u_0 - u^{\text{ps}}\|^2 + 4\eta^2\sigma^2\sum_{s=1}^{t}\left(\frac{1}{1+\eta\alpha}\right)^t\\ &\leq \left(\frac{1}{1+\eta\alpha}\right)^t \|u_0 - u^{\text{ps}}\|^2 + 4\eta^2\sigma^2\frac{1}{\eta\alpha}\end{aligned}$$

Given the choice of $\eta \leq \frac{\alpha}{4L_\ell^2}$, we have that

$$\mathbb{E}_t\|u_{t+1} - u^{\mathsf{ps}}\|^2 \leq \left(\frac{1}{1+\eta\alpha}\right)^t \|u_0 - u^{\mathsf{ps}}\|^2 + \frac{\sigma^2}{L_\ell^2} \leq \left(1 - \frac{\alpha\eta}{2}\right)^t \|u_0 - u^{\mathsf{ps}}\|^2 + \frac{4\eta\sigma^2}{\alpha}.$$

This completes the proof. □

As noted in the main it possible to employ any number of stage-based methods from stochastic optimization in order to obtain convergence to an ε-performatively stable Stackelberg equilibrium.

Corollary 3. Under the assumptions of Theorem 3, consider running the stochastic repeated gradient method in $k = 0, \ldots, K$ super-epochs, for T_k epochs each with constant step-size $\eta_k = 2^{-k}\eta_0$, and such that the last iterate of each epoch k is used as the first iterate in stage $k+1$. Fix a target accuracy $\varepsilon > 0$ and suppose the decision-maker has available $R \geq \|u_0 - u^{\mathsf{ps}}\|$. Set $\eta_0 := \frac{\alpha}{4L_\ell^2(1+L_{\mathsf{eq}}^2)}$, and

$$T_0 = \left\lceil \frac{2}{\alpha\eta_0} \log(\frac{2R^2}{\varepsilon}) \right\rceil, \quad T_k = \left\lceil \frac{2\log(4)}{\alpha\eta_k} \right\rceil \quad \text{for } k \geq 1, \quad \text{and } K = \left\lceil 1 + \log_2\left(\frac{\sigma^2}{L_\ell^2(1+L_{\mathsf{eq}}^2)\varepsilon}\right) \right\rceil.$$

Then

$$\mathbb{E}\,\|u_T - u^{\mathsf{ps}}\|^2 \leq \varepsilon \quad \text{and} \quad \mathbb{E}\,\|x_T - x^*(u^{\mathsf{ps}})\|^2 \leq 2(\epsilon_\tau + L_{\mathsf{eq}}\varepsilon)$$

in a total number of epochs

$$T = \sum_{k=1}^{K} T_k \quad \text{that is at most} \quad \mathcal{O}\left(\frac{L_\ell^2(1+L_{\mathsf{eq}}^2)}{\alpha^2} \log\left(\frac{2R^2}{\varepsilon}\right) + \frac{\sigma^2}{\alpha^2\varepsilon}\right).$$

Proof. The proof follows immediately from applying Lemma 3 with $\mathcal{A} \equiv \texttt{RGM}$ in Algorithm 2. Indeed, we set constants

$$\psi(\eta) = \frac{\alpha\eta}{2}, \quad C = 1, \quad D = \frac{4\sigma^2}{\alpha}, \quad \eta_0 = \frac{\alpha}{4L_\ell^2(1+L_{\mathsf{eq}}^2)}$$

so that

$$T = \sum_{k=0}^{K} T_k = \left\lceil \frac{8L_\ell^2(1+L_{\mathsf{eq}}^2)}{\alpha^2} \cdot \log\left(\frac{2R^2}{\varepsilon}\right) \right\rceil + \sum_{k=1}^{K} \left\lceil \frac{8L_\ell^2(1+L_{\mathsf{eq}}^2)\log(4)}{\alpha^2 2^{-k}} \right\rceil$$

and K is given as in the corollary statement. Applying Lemma 3 gives us that $\mathbb{E}\,\|u_T - u^{\mathsf{ps}}\|^2 \leq \varepsilon$ in T total stages. Then

$$\mathbb{E}\,\|x_T - x^*(u^{\mathsf{ps}})\|^2 \leq 2\,\mathbb{E}\,\|x_T - x^*(u_T)\|^2 + 2\,\mathbb{E}\,\|x^*(u_T) - x^*(u^{\mathsf{ps}})\|^2 \leq 2(\epsilon_\tau + L_{\mathsf{eq}}\varepsilon),$$

where ϵ_τ is given in the Theorem 5. □

I Proof Strategic Decision-Maker

In this appendix section, we put all the formal analysis for the strategic decision-maker. Let us introduce some needed notation. Let

$$\mathcal{L}_t^\delta(u_t) = \frac{d}{\delta} \mathop{\mathbb{E}}_{v\sim\mathbb{B}^d}\left[\mathop{\mathbb{E}}_{\xi\sim\mathcal{D}_o}[\ell(u_t + \delta v_t, \mathcal{A}(x_t, u_t + \delta v_t), +\xi)]\right]$$

denote the smoothed expected loss at time t, and let

$$\mathcal{L}^\delta(u) = \frac{d}{\delta} \mathop{\mathbb{E}}_{v\sim\mathbb{B}^d}\left[\mathop{\mathbb{E}}_{\xi\sim\mathcal{D}_o}[\ell(u_t + \delta v_t, x^*(u_t + \delta v_t) + \xi)]\right]$$

denote the smoothed expected risk. The smoothed expected risk is evaluated when the strategic agents are at the Nash equilibrium $x^*(u_t + \delta v_t)$ for the reported value $u_t + \delta v_t$. The estimate $\widehat{g}_t$ is an unbiased estimate of $\nabla\mathcal{L}_t^\delta$—i.e., $\mathbb{E}_{v\sim\mathbb{B}^d}[\widehat{g}_t] = \nabla\mathcal{L}_t^\delta(u_t)$.

I.1 Technical Lemmas

The following series of lemmas allow us to bound the error between the true gradient of $\mathcal{L}(u)$ and the zeroth-order gradient estimate g_t.

Lemma 8. Suppose that Assumptions 1 and 3 hold. The smoothed expected risk $\mathcal{L}^\delta(u)$ satisfies $\|\nabla\mathcal{L}(u) - \nabla\mathcal{L}^\delta(u)\| \leq L\delta$, where $L := L_\ell(1 + L_{\mathtt{eq}})$.

Proof. For any points $u, u' \in \mathcal{U}$, we successively estimate

$$\|\nabla\mathcal{L}^\delta(u) - \nabla\mathcal{L}^\delta(u')\| \leq \mathop{\mathbb{E}}_{w\sim\mathbb{B}} \|\nabla\mathcal{L}(u + \delta w) - \nabla\mathcal{L}(u' + \delta w)\| \leq L\|u - u'\|.$$

Therefore $\nabla\mathcal{L}^\delta$ is L-Lipschitz continuous. Next, we have that

$$\|\nabla\mathcal{L}(u) - \nabla\mathcal{L}^\delta(u)\| \leq \mathop{\mathbb{E}}_{w\sim\mathbb{B}} \|\nabla\mathcal{L}(\delta w) - \nabla\mathcal{L}(u)\| \leq L\delta \mathop{\mathbb{E}}_{w\sim\mathbb{B}} \|w\| \leq L\delta,$$

which concludes the proof. □

Lemma 9. Under Assumptions 3 and 5, by choosing $\delta \leq \frac{c\bar{\alpha}}{L_{\mathtt{H}}}$, for any $c \in (0, 1)$ the smoothed decision-dependent risk $\mathcal{L}^\delta(u)$ is $(1 - c)\bar{\alpha}$–strongly convex.

Proof. We first define $h(u) := \nabla\mathcal{L}^\delta(u) - \nabla\mathcal{L}(u)$. Observe that $\nabla h(u) = \mathbb{E}_{w\sim\mathbb{B}}[\nabla^2\mathcal{L}(u + \delta w) - \nabla^2\mathcal{L}(u)]$. Since $u \mapsto \nabla^2\mathcal{L}(u)$ is $L_{\mathtt{H}}$-Lipschitz continuous, we deduce that

$$\|\nabla h(u)\|_{\mathtt{op}} \leq \mathop{\mathbb{E}}_{w\sim\mathbb{B}} \|\nabla^2\mathcal{L}(u + \delta w) - \nabla^2\mathcal{L}(u)\|_{\mathtt{op}} \leq \delta L_{\mathtt{H}} \mathop{\mathbb{E}}_{w\sim\mathbb{B}} \|w\| \leq \delta L_{\mathtt{H}}.$$

We therefore compute that

$$\langle\nabla\mathcal{L}^\delta(u) - \nabla\mathcal{L}^\delta(u'), u - u'\rangle = \langle\nabla\mathcal{L}(u) - \nabla\mathcal{L}(u')\rangle + \langle h(u) - h(u'), u - u'\rangle \geq (\bar{\alpha} - L_{\mathtt{H}}\delta)\|u - u'\|^2,$$

which concludes the proof of the first statement. □

Now, let u^* be the optimal point for $\mathcal{L}$ over $\mathcal{U}$, and let u^δ be the optimal point of $\mathcal{L}^\delta$ on $(1 - \delta)\mathcal{U}$.

Lemma 10. Suppose Assumptions 3 and 5 hold. Choose any $\delta < \min\{r, \frac{\bar{\alpha}}{L_{\mathtt{H}}}\}$. Then the estimate holds:

$$\|u^\delta - u^*\| \leq \frac{\delta L}{\bar{\alpha}} + \left(\frac{\delta L}{\bar{\alpha}} + \delta\right)\|u^*\|.$$

Proof. There are two sources of perturbation: one replacing $\mathcal{U}$ with $(1 - \delta)\mathcal{U}$ and the other replacing $\mathcal{L}$ with $\mathcal{L}^\delta$.

Set $\phi = 1 - \delta$ and let $\tilde{u}$ be the optimal point for $\mathcal{L}$ on $\phi\mathcal{U}$. Thus $0 \in \nabla\mathcal{L}(\tilde{u}) + N_{\phi\mathcal{U}}(\tilde{u})$. Then

$$\|u^* - u^\delta\| \leq \|u^* - \tilde{u}\| + \|\tilde{u} - u^\delta\|.$$

The first term is bounded as

$$\bar{\alpha}\|\tilde{u} - \phi u^*\| \leq \operatorname{dist}(0, \nabla\mathcal{L}(\phi u^*) + N_{\phi\mathcal{U}}(\phi u^*)$$

since $u \mapsto \nabla\mathcal{L}(u) + N_{\phi\mathcal{U}}(u)$ is $\bar{\alpha}$–strongly convex. For the second term, since u^* is optimal, we have that $0 \in \nabla\mathcal{L}(u^*) + N_{\mathcal{U}}(u^*)$. Since $N_{\phi\mathcal{U}}(\phi u^*) = N_{\mathcal{U}}(u^*)$, we have that

$$\operatorname{dist}(0, \nabla\mathcal{L}(\phi u^*) + N_{\phi\mathcal{U}}(\phi u^*)) = \operatorname{dist}(0, \nabla\mathcal{L}(\phi u^*) + N_{\mathcal{U}}(u^*)) \leq \|\nabla\mathcal{L}(\phi u^*) - \nabla\mathcal{L}(u^*)\| \leq \delta L\|u^*\|.$$

We therefore have that

$$\|u^* - \tilde{u}\| \leq \|\tilde{u} - \phi u^*\| + \delta\|u^*\| \leq \delta\left(1 + \frac{L}{\bar{\alpha}}\right)\|u^*\|.$$

Since $\tilde{u}$ is optimal, we have that

$$\langle -\nabla\mathcal{L}(\tilde{u}), u - \tilde{u}\rangle \leq 0, \quad \forall u \in \phi\mathcal{U}.$$

Analogously, since u^δ is also optimal we have that

$$\langle -\nabla \mathcal{L}^\delta(u^\delta), u - u^\delta \rangle \le 0, \quad \forall u \in \phi\mathcal{U}.$$

Therefore

$$\begin{aligned}
\bar{\alpha}\|\tilde{u} - u^\delta\|^2 &\le \langle \nabla \mathcal{L}(\tilde{u}) - \nabla \mathcal{L}(u^\delta), \tilde{u} - u^\delta \rangle \\
&\le \langle \nabla \mathcal{L}^\delta(u^\delta) - \nabla \mathcal{L}(u^\delta), \tilde{u} - u^\delta \rangle \\
&\le \|\nabla \mathcal{L}^\delta(u^\delta) - \nabla \mathcal{L}(u^\delta)\| \|\tilde{u} - u^\delta\| \\
&\le L\delta \|\tilde{u} - u^\delta\|.
\end{aligned}$$

Combining the bounds yields the claim. $\square$

The next lemma bounds the error between the converged strategies $x_t^{\tau_t}(u_t)$ and the equilibrium $x^*(u_t)$ as a function of the previous iterates.

Lemma 11. Suppose the agents are employing deterministic algorithms satisfying Definition 1 with $\rho \in [0, 1)$ and $\sigma_{\mathsf{a}} = 0$. Under Assumptions 1, 3, 9 and 5, the estimate holds:

$$\|\mathcal{A}(x_t, u_t) - x^*(u_t)\| \le \rho^{\tau_t} \left(\rho^{t-1}\|x_0 - x^*(u_0)\| + \rho L_{\mathsf{eq}} \eta_0 \frac{\ell_* d}{\delta(1-\rho)} \right)$$

Proof. Given Definition 1, we have that

$$\|x_t^{\tau_t}(u_t) - x^*(u_t)\| \le \rho \|x_t^{\tau_t - 1}(u_t) - x^*(u_t)\|.$$

Iterating this expression we have that

$$\|x_t^{\tau_t}(u_t) - x^*(u_t)\| \le \rho^{\tau_t} \|x_{t-1}^{\tau_{t-1}}(u_{t-1}) - x^*(u_t)\|$$

Adding and subtracting appropriate terms we have that

$$\begin{aligned}
\|x_t^{\tau_t}(u_t) - x^*(u_t)\| &\le \rho^{\tau_t} \|x_{t-1}^{\tau_{t-1}}(u_{t-1}) - x^*(u_{t-1}) + x^*(u_{t-1}) - x^*(u_t)\| \\
&\le \rho^{\tau_t} \|x_{t-1}^{\tau_{t-1}}(u_{t-1}) - x^*(u_{t-1})\| + \rho^{\tau_t} L_{\mathsf{eq}} \|u_{t-1} - u_t\|
\end{aligned}$$

Continuing in this fashion we have that

$$\begin{aligned}
\|x_t^{\tau_t}(u_t) - x^*(u_t)\| &\le \rho^{\tau_t} \|x_{t-1}^{\tau_{t-1}}(u_{t-1}) - x^*(u_{t-1})\| + \rho^{\tau_t} L_{\mathsf{eq}} \|u_{t-1} - u_t\| \\
&\le \rho^{\tau_t} (\rho^{\tau_{t-1}} \|x_{t-2}^{\tau_{t-2}}(u_{t-2}) - x^*(u_{t-2})\| + \rho^{\tau_{t-1}} L_{\mathsf{eq}} \|u_{t-2} - u_{t-1}\|) + \rho^{\tau_t} L_{\mathsf{eq}} \|u_{t-1} - u_t\| \\
&\le \rho^{\tau_t} \rho^{t-1} \|x_0 - x^*(u_0)\| + L_{\mathsf{eq}} \rho^{\tau_t} \sum_{s=1}^{t} \rho^s \|u_{t-s} - u_{t-s-1}\|,
\end{aligned}$$

where in the last inequality we use the fact that $\rho^\tau \le \rho$ for any $\tau \ge 1$. Using the update for the decision maker, we have that

$$\begin{aligned}
\|x_t^{\tau_t}(u_t) - x^*(u_t)\| &\le \rho^{\tau_t} \left(\rho^{t-1} \|x_1 - x^*(u_1)\| + \rho L_{\mathsf{eq}} \sum_{s=0}^{t-1} \rho^s \eta_{t-1-s} \frac{\ell_* d}{\delta} \right) \\
&\le \rho^{\tau_t} \left(\rho^{t-1} \|x_0 - x^*(u_0)\| + \rho L_{\mathsf{eq}} \eta_0 \frac{\ell_* d}{\delta(1-\rho)} \right),
\end{aligned}$$

where we used the fact that $\eta_0 \ge \eta_{t-i}$ for all $t \ge 1$. This concludes the proof. $\square$

Next, we use Lemma 11 to bound the error between the gradient of the smoothed expected risk $\mathcal{L}^\delta(u)$ and the smoothed loss $\mathcal{L}_t^\delta(u_t)$.

Lemma 12. Suppose the agents are employing deterministic algorithms satisfying Definition 1 with $\rho \in [0, 1)$ and $\sigma_{\mathtt{a}} = 0$. Under Assumptions 1, 3, 9 and 5, the smoothed expected risk and the smoothed loss satisfy

$$\|\nabla \mathcal{L}_t^\delta(u_t) - \nabla \mathcal{L}^\delta(u_t)\|^2 \leq L_\ell^2 \left(\rho^{\tau_t}\left(\rho^{t-1}\|x_0 - x^*(u_0)\| + \frac{L_{\mathtt{eq}}\eta_0 \ell_* d}{\delta(1-\rho)}\right)\right)^2.$$

Proof. We have that

$$\begin{aligned}\|\nabla \mathcal{L}_t^\delta(u_t) - \nabla \mathcal{L}^\delta(u_t)\| &\leq \mathop{\mathbb{E}}_{v\sim\mathbb{B}^d}\Big\| \mathop{\mathbb{E}}_{\xi\sim\mathcal{D}_o}[\nabla\ell(u_t+\delta v_t, x_t^{\tau_t}(u_t+\delta v_t)+\xi) \\ &\qquad - \nabla\ell(u_t+\delta v, x^*(u_t+\delta v)+\xi)]\Big\| \\ &\leq L_\ell \mathop{\mathbb{E}}_{v\sim\mathbb{B}^d}\|x_t^{\tau_t}(u_t+\delta v_t) - x^*(u_t+\delta v_t)\|.\end{aligned}$$

Hence applying Lemma 11 gives the result. □

The above lemmas give us our main result, which establishes that the decision-maker's updates converge to the optimal parameter $u^* \in \mathcal{U}$ (and correspondingly, the agents' updates converge to the Nash equilibrium $x^*(u^*)$).

Let us define a useful quantity that we will use in the remaining proof:

$$C_t(\sigma_{\mathtt{a}}) := \left(\rho^{t-1}\|x_0 - x^*(u_0)\| + \frac{L_{\mathtt{eq}}\eta_0\ell_* d}{\delta(1-\rho)} + \frac{\rho\sigma_{\mathtt{a}} c}{(1-\rho)^2}\right), \tag{23}$$

so that

$$\|\nabla \mathcal{L}_t^\delta(u_t) - \nabla \mathcal{L}^\delta(u_t)\|^2 \leq L_\ell^2\left(\rho^{\tau_t}C_t(\sigma_{\mathtt{a}}) + \frac{\rho\sigma_{\mathtt{a}}c}{1-\rho}\right)^2.$$

Theorem 6 (Formal Statement of Theorem 4)**.** Suppose that Assumptions 1, 3, and 5 hold, and that we have available a constant $R > \|x_0 - x^*(u_0)\|$. Further, suppose the decision-maker runs Algorithm 1 with $\mathtt{Alg} := \mathtt{DFM}$ using step-size $\eta_t = \frac{4}{\bar{\alpha}(t+1)}$, query radius $\delta < \min\{b, \frac{\bar{\alpha}}{L_{\mathtt{H}}}\}$, and the agents employ a ρ–contracting algorithm $\mathcal{A}$ with $\rho \in [0, 1)$. Consider the following two cases:

- **Case 1:** The agents employ deterministic algorithms (i.e., $\sigma_{\mathtt{a}} = 0$) and the decision-maker receives a noisy observation $\mathcal{A}(x_{t-1}, u_t) + \xi$ in each round where ξ is zero mean and finite variance. In this case, set the epoch length such that $\tau \geq \log(\frac{2\delta L_\ell C_t(0)}{\sqrt{\eta_t}\bar{\alpha}\ell_* d})\frac{1}{\log(1/\rho)}$, constant $c = 32\ell_*^2 d^2$ and agent tolerance $\epsilon_t = C_t(0)\rho^{\tau_t}$.
- **Case 2:** The agents employ stochastic gradient play with $\sigma_{\mathtt{a}} \in (0, \infty)$ run stage-wise via Algorithm 2. In this case, set the epoch length to

$$\tau = \sum_{k=0}^{K} T_k = \left\lceil\left(1 + \frac{2L_{\mathtt{a}}^2}{\mu^2}\right)\log\left(\frac{2R}{\epsilon_t}\right)\right\rceil + \sum_{k=1}^{K}\left\lceil\left(1 + \frac{2^{k+1}L_{\mathtt{a}}^2}{\mu^2}\right)\log(4)\right\rceil, \tag{24}$$

where $K = \left\lceil 1 + \log_2(\frac{\sigma_{\mathtt{a}}^2}{\epsilon_t L_{\mathtt{a}}^2})\right\rceil$, tolerance $\epsilon_\tau = \frac{1}{\delta(t+1)}$, and constant $c = 16(\ell_*^2 d^2 + 1)$.

Then the following estimate holds:

$$\mathbb{E}\,\|u_t - u^*\|^2 \leq \frac{\max\{2\bar{\alpha}^2\delta^2\|u_0 - u^*\|^2, c\}}{\delta^2\bar{\alpha}^2(t+1)^\beta} + 2\delta^2\left(\left(1 + \frac{L}{\bar{\alpha}}\right)\|u^*\| + \frac{L}{\bar{\alpha}}\right).$$

The proof follows a similar structure to that of Theorem 3 (see Theorem 5 for the longer version).

Proof. Consider the error $\|u_{t+1} - u^*\|^2$. Add and subtract u^δ, and apply the triangle inequality and Lemma 9 to get the following estimate:

$$\frac{1}{2}\|u_{t+1} - u^*\|^2 \leq \|u_{t+1} - u^\delta\|^2 + \|u_\delta - u^*\|^2 \leq \|u_{t+1} - u^\delta\|^2 + \left(\frac{\delta L}{\bar{\alpha}} + \left(\frac{\delta L}{\bar{\alpha}} + \delta\right)\|u^*\|\right)^2.$$

Now, to bound the error $\|u_{t+1} - u^\delta\|^2$, we note by the nonexpansiveness of the projection mapping that

$$\begin{aligned}\mathbb{E}[\|u_{t+1} - u^\delta\|^2] &\leq \mathbb{E}[\|u_t - u^\delta - \eta_t g_t\|^2] \\ &\leq \mathbb{E}[\|u_t - u^\delta\|^2 - 2\eta_t \mathbb{E}\langle g_t, u_t - u^\delta\rangle + \eta_t^2 \mathbb{E}\|g_t\|^2] \\ &\leq \mathbb{E}[\|u_t - u^\delta\|^2 - 2\eta_t \mathbb{E}\langle \nabla \mathcal{L}_t^\delta(u_t), u_t - u^\delta\rangle + \eta_t^2 \mathbb{E}\|g_t\|^2]\end{aligned}$$

where we use the fact that $\mathbb{E}[g_t] = \nabla \mathcal{L}_t^\delta(u_t)$ in the last inequality, and the expectation is taken over the randomness in ξ and v_t up to time t.

Next, we add and subtract $\nabla \mathcal{L}^\delta(u_t)$ from the middle term to get that

$$\begin{aligned}\langle \nabla \mathcal{L}_t^\delta(u_t), u_t - u^\delta\rangle &= \langle \nabla \mathcal{L}^\delta(u_t), u_t - u^\delta\rangle + \langle \nabla \mathcal{L}_t^\delta(u_t) - \nabla \mathcal{L}^\delta(u_t), u_t - u^\delta\rangle \\ &\leq \frac{\bar{\alpha}}{2}\|u_t - u^\delta\|^2 + \langle \mathcal{L}_t^\delta(u_t) - \nabla \mathcal{L}^\delta(u_t), u_t - u^\delta\rangle,\end{aligned}$$

where we have used the fact that $\mathcal{L}^\delta(x)$ is $(1-c)\bar{\alpha}$–strong convex with $c = 1/2$ (Lemma 9). Hence, we deduce that

$$\begin{aligned}\mathbb{E}[\|u_{t+1} - u^\delta\|^2] &\leq \mathbb{E}[\|u_t - u^\delta - \eta_t g_t\|^2], \\ &\leq \mathbb{E}[\|u_t - u^\delta\|^2] - 2\eta_t \,\mathbb{E}\langle \nabla \mathcal{L}^\delta(u_t), u_t - u^\delta\rangle \\ &\quad - 2\eta_t \,\mathbb{E}\langle \nabla \mathcal{L}_t^\delta(u_t) - \nabla \mathcal{L}^\delta(u_t), u_t - u^\delta\rangle + \eta_t^2 \,\mathbb{E}\,\|g_t\|^2, \\ &\leq (1 - \eta_t\bar{\alpha})\,\mathbb{E}[\|u_t - u^\delta\|^2] - 2\eta_t \,\mathbb{E}\langle \nabla \mathcal{L}_t^\delta(u_t) - \nabla \mathcal{L}^\delta(u_t), u_t - u^\delta\rangle + \eta_t^2 \frac{\ell_*^2 d^2}{2\delta^2}.\end{aligned}$$

Here is where we split the proof into the two cases.

Proof of Case 1. We start by bounding the term $\langle \nabla \mathcal{L}_t^\delta(u_t) - \nabla \mathcal{L}^\delta(u_t), u_t - u^\delta\rangle$ and we apply Young's inequality[6] to this term to get that

$$\begin{aligned}\mathbb{E}\,|\langle \nabla \mathcal{L}_t^\delta(u_t) - \nabla \mathcal{L}^\delta(u_t), u_t - u^\delta\rangle| &\leq \frac{1}{2\nu_1}\,\mathbb{E}\,\|\nabla \mathcal{L}_t^\delta(u_t) - \nabla \mathcal{L}^\delta(u_t)\|^2 + \frac{\nu_1}{2}\,\mathbb{E}\,\|u_t - u^\delta\|^2 \\ &\leq \frac{1}{2\nu_1}\left(L_\ell^2\,(\rho^{\tau_t} C_t)^2\right) + \frac{\nu_1}{2}\,\mathbb{E}\,\|u_t - u^\delta\|^2,\end{aligned}$$

where since we are in **Case 1**, we have that $\sigma_{\mathsf{a}} = 0$ so that

$$\bar{C}_t := \left(\rho^{t-1}\|x_0 - x^*(u_0)\| + \frac{L_{\mathsf{eq}}\eta_0 \ell_* d}{\delta(1-\rho)}\right).$$

Setting $\nu_1 := \bar{\alpha}/2$, we deduce that

$$\begin{aligned}\mathbb{E}\|u_{t+1} - u^\delta\|^2 &\leq (1 - \eta_t\bar{\alpha})\,\mathbb{E}[\|u_t - u^\delta\|^2] + \frac{\eta_t^2 \ell_*^2 d^2}{2\delta^2} + 2\eta_t\left(\frac{1}{2\Delta_1}\left(L_\ell^2\,(\rho^{\tau_t}\bar{C}_t)^2\right) + \frac{\Delta_1}{2}\|u_t - u^\delta\|^2\right), \\ &\leq \left(1 - \frac{\bar{\alpha}\eta_t}{2}\right)\mathbb{E}[\|u_t - u^\delta\|^2] + \frac{\eta_t^2 \ell_*^2 d^2}{2\delta^2} + \frac{2\eta_t}{\bar{\alpha}}\left(L_\ell^2\,(\rho^{\tau_t}\bar{C}_t)^2\right).\end{aligned}$$

Hence, if it is the case that

$$\frac{2\eta_t}{\bar{\alpha}} L_\ell^2 \rho^{2\tau_t} \bar{C}_t^2 \leq \frac{\eta_t^2 \ell_*^2 d^2}{2\delta^2}, \tag{25}$$

then we conclude

$$\mathbb{E}\|u_{t+1} - u^\delta\|^2 \leq \left(1 - \eta_t\frac{\bar{\alpha}}{2}\right)\mathbb{E}[\|u_t - u^\delta\|^2] + \eta_t^2 \frac{\ell_*^2 d^2}{\delta^2} \tag{26}$$

Indeed, the bound in (25) is equivalent to

$$\tau_t \log(\rho^2) \leq \log\left(\frac{\eta_t \ell_*^2 d^2}{4\delta^2}\frac{\bar{\alpha}}{L_\ell^2 \bar{C}_t^2}\right) \iff \tau_t \geq \log\left(\frac{2\delta L_\ell \bar{C}_t}{\sqrt{\eta_t\bar{\alpha}}\ell_* d}\right)\frac{1}{\log(1/\rho)},$$

[6] Young's inequality for inner product spaces says that for two vectors $u, v \in V$ where V is an inner product space, we have $\langle u, v\rangle \leq \frac{\lambda}{2}\|u\|^2 + \frac{1}{2\lambda}\|v\|^2$ for any $\lambda > 0$.

which is precisely the assumed bound on τ_t.

Recall that $\eta_t = \frac{4}{\bar{\alpha}(t+1)}$. Hence we apply Lemma 1 to obtain the final bound in this case. Indeed, we have that

$$\mathbb{E}\,\|u_t - u^*\|^2 \leq \frac{\max\{4\bar{\alpha}^2\delta^2\|u_0 - u^\delta\|^2, 32d^2\ell_*^2\}}{\bar{\alpha}^2\delta^2(t+1)} + 2\delta^2\left(\left(1+\frac{L}{\bar{\alpha}}\right)\|u^*\| + \frac{L}{\bar{\alpha}}\right) \tag{27}$$

as claimed.

Now we proceed to **Case 2.**

Proof of Case 2. In this case the agents are running stage-based stochastic gradient play. In this case we set $\epsilon_t = (\delta^2(t+1))^{-1}$ and choose

$$\tau_t = \sum_{k=0}^{K} T_k = \left\lceil \left(1+\frac{2L_{\mathsf{a}}^2}{\mu^2}\right)\log\left(\frac{2R}{\epsilon_t}\right)\right\rceil + \sum_{k=1}^{K}\left\lceil\left(1+\frac{2^{k+1}L_{\mathsf{a}}^2}{\mu^2}\right)\log(4)\right\rceil,$$

and $K = \left\lceil 1+\log_2\left(\frac{\sigma_{\mathsf{a}}^2}{L_{\mathsf{a}}^2\epsilon_t}\right)\right\rceil$. Then we have that

$$\begin{aligned}
\mathbb{E}\|u_{t+1} - u^\delta\|^2 &\leq \left(1-\frac{\bar{\alpha}\eta_t}{2}\right)\mathbb{E}[\|u_t - u^\delta\|^2] + \frac{8\ell_*^2 d^2}{\delta^2\alpha^2(t+1)^2} + \frac{8}{\bar{\alpha}^2}\frac{1}{\delta^2(t+1)^2}\\
&\leq \left(1-\frac{2}{t+1}\right)\mathbb{E}[\|u_t - u^\delta\|^2] + \left(\ell_*^2 d^2+1\right)\frac{8}{\bar{\alpha}^2\delta^2(t+1)^2}.
\end{aligned}$$

Next, we claim that

$$\mathbb{E}\,\|u_t - u^\delta\|^2 \leq \frac{\max\{\bar{\alpha}^2\delta^2\|u_0 - u^\delta\|^2, 8(\ell_*^2 d^2+1)\}}{\delta^2\bar{\alpha}^2(t+1)}.$$

To see this, let $D_t = \mathbb{E}[\|u_t - u^\delta\|^2]$ so that we need to show the above claim given that

$$D_{t+1} \leq \left(1-\frac{2}{t+1}\right)D_t + \left(\ell_*^2 d^2+1\right)\frac{8}{\bar{\alpha}^2\delta^2(t+1)^2}.$$

Clearly the claim holds for $t=1$. Suppose it holds for some fixed $t>1$. Then we have that

$$\begin{aligned}
D_{t+1} &\leq \left(1-\frac{2}{t+1}\right)D_t + \left(\ell_*^2 d^2+1\right)\frac{8}{\bar{\alpha}^2\delta^2(t+1)^2}\\
&\leq \left(1-\frac{2}{t+1}\right)\frac{8(\ell_*^2 d^2+1)}{\delta^2\bar{\alpha}^2(t+1)} + \left(\ell_*^2 d^2+1\right)\frac{8}{\bar{\alpha}^2\delta^2(t+1)^2}\\
&\leq \frac{8(\ell_*^2 d^2+1)}{\delta^2\bar{\alpha}^2}\left(\frac{1}{(t+1)} - \frac{1}{(t+1)^2}\right)\\
&\leq \frac{8(\ell_*^2 d^2+1)}{\delta^2\bar{\alpha}^2}\frac{1}{(t+2)}
\end{aligned}$$

Therefore

$$\mathbb{E}\,\|u_t - u^*\|^2 \leq \frac{\max\{4\bar{\alpha}^2\delta^2\|u_0 - u^\delta\|^2, 16(\ell_*^2 d^2+1)\}}{\delta^2\bar{\alpha}^2(t+1)} + 2\delta^2\left(\left(1+\frac{L}{\bar{\alpha}}\right)\|u^*\| + \frac{L}{\bar{\alpha}}\right)$$

□

The preceding theorem allows us to obtain the following convergence guarantee.

Corollary 4. Suppose the assumptions of Theorem 6 hold. Fix target accuracy $\varepsilon < 4b^2\left(\left(1+\frac{L}{\bar{\alpha}}\right)B + \frac{L}{\bar{\alpha}}\right)^2$, and set $\delta = \bar{\alpha}\sqrt{\varepsilon/4}/((\bar{\alpha}+L)B+L)$ and $\eta_t = 4/(\bar{\alpha}(t+1))$. The iterates (u_t, x_t) converge to an approximate Stackelberg equilibrium $(u^*, x^*(u^*))$—i.e., the estimates $\mathbb{E}[\|u_t - u^*\|^2] \leq \varepsilon$ and $\mathbb{E}[\|x_t - x^*(u^*)\|^2] \leq 2(\epsilon_0 + L_{\mathsf{eq}}\varepsilon)$ hold for all

$$t \geq \frac{\max\{16\bar{\alpha}^4\varepsilon B^2, 8c((\bar{\alpha}+L)B+L)^2\}}{\bar{\alpha}^4\varepsilon^2}.$$

In the proceeding theorem, the lower bound on t is in terms of the number of epochs. In terms of total iterations $\sum_{s=1}^{t} \tau_s$ the rate is $\mathcal{O}\left(\frac{d^2}{\varepsilon^2}\log(1/\varepsilon)\right)$. This rate is equivalent to $\widetilde{\mathcal{O}}(T^{-1/2})$ in terms of iteration complexity where $T = \sum_{s=1}^{t} \tau_s$.

Proof of Corollary 4. The assumed upper bound on ε directly implies that $\delta \leq \bar{\alpha}/(2L_{\mathtt{H}})$ and $\delta < b$. Applying Theorem 6 yields

$$\begin{aligned}\mathbb{E}[\|x_t - x^*\|^2] &\leq \frac{\max\{2\bar{\alpha}^2\delta^2\|u_0 - u^\delta\|^2, c\}}{\delta^2\bar{\alpha}^2(t+1)} + 2\delta^2\left(\left(1+\frac{L}{\bar{\alpha}}\right)\|u^*\| + \frac{L}{\bar{\alpha}}\right)^2 \\ &\leq \frac{\max\{8\bar{\alpha}^4\varepsilon B^2, 4c((\bar{\alpha}+L)B+L)^2\}}{\varepsilon\bar{\alpha}^4(t+1)} + \frac{\varepsilon}{2}.\end{aligned}$$

Setting the right-hand side to ε and solving for t, concludes the proof. □