LEGO: Optimal Online Learning under Sequential Price Competition

(Authors' names blinded for peer review)

We consider price competition among multiple sellers over a selling horizon of T periods. In each period, sellers simultaneously offer their prices and subsequently observe their respective demand that is unobservable to competitors. The realized demand of each seller depends on the prices of all sellers following a private unknown linear model. We propose a least-squares estimation then gradient optimization (LEGO) policy, which does not require sellers to communicate demand information or coordinate price experiments throughout the selling horizon. We show that our policy, when employed by all sellers, leads at a fast convergence rate $O(1/\sqrt{T})$ to the Nash equilibrium prices that sellers would reach if they were fully informed. Meanwhile, each seller achieves an optimal order-of-$\sqrt{T}$ regret relative to a dynamic benchmark policy. Our analysis further shows that the *unknown individual price sensitivity contributes to the major difficulty of dynamic pricing in sequential competition* and forces regret to the order of $\sqrt{T}$ in the worst case. If each seller knows their individual price sensitivity coefficient, then a gradient optimization policy can achieve an optimal order-of-$\frac{1}{T}$ convergence rate to Nash equilibrium as well as an optimal order-of-$\log T$ regret.

Key words: sequential competition, dynamic pricing, demand learning, Nash equilibrium, regret analysis.

1. Introduction

Competition permeates various sectors, spanning from food retail to advanced manufacturing. In the bubble tea business, companies such as Fong Fu Tea, Gong Cha, and Möge Tee compete for market share (Qin 2023). The global fast-fashion industry, which was estimated worth \$1.7 trillion in 2023, also witnesses intense competition among famous brands such as Zara, H&M, and Uniqlo (McKinsey 2023). Meanwhile, emerging and disruptive entrants like Shein and TikTok Shop have gained leverage over their more established counterparts by offering less expensive products. In the personal computer (PC) market, notable manufacturers include Dell, HP, and Apple, whose price competition recently expanded to mini desktop PCs (Gershgorn 2024). Likewise, the electronic vehicles (EV) sector, comprising giant automotive manufacturers such as Tesla, BYD, and Volkswagen, showcases substantial competition within a global market worth \$57 trillion (Scott et al. 2023). Recently, new players including Huawei and Xiaomi launched new EV models, intensifying the "price wars" in China, the world's largest EV market (Ren 2023). These examples demonstrate that competition is a ubiquitous force shaping market dynamics across diverse sectors.

Meanwhile, there is an increasing trend for companies to adopt data-driven tools to determine prices in response to the evolving demand. Through sophisticated automated pricing algorithms or algorithmic recommendations, prices are sometimes adjusted on a daily or hourly basis. This fast-moving nature of pricing strategies, once primarily associated with airlines and hotels, has now permeated various sectors of commerce, including mobility service, sporting event tickets, and even fast-food restaurants. For instance, Uber has implemented a surge-pricing strategy, considering various factors such as the availability of drivers, geographical locations, and even the customer's phone battery level (which correlates with their patience) (Martin 2019). Wendy's recently announced plans to implement "Uber-like surge pricing" on its menu, leading to price fluctuations of items like the Frosty throughout the day (Valinsky 2024). The ascendancy of algorithms over human decision-makers in setting prices reflects a paradigm shift in contemporary commerce.

While the Internet has transformed the dissemination of price information, enhancing transparency and enabling real-time monitoring and rapid responses, there remains a notable gap in understanding how to incorporate competition into data-driven pricing algorithms within literature and business practices. Despite some companies strategically considering competitors, many pricing algorithms do not model competition at a day-to-day operational level. For example, Phillips (2021) remarks that "there does not appear to be a single pricing and revenue optimization system that explicitly attempts to forecast competitive response using game theory as part of its ongoing operation". Even in scenarios where competitors are acknowledged, individual sellers may still choose to employ a monopolistic model, hoping that individual demand data may somehow implicitly incorporate competition effects. Given the widespread use of monopoly models, it is unsurprising that much of the operations literature neglects competitors.

Indeed, several challenges are associated with designing a pricing algorithm that incorporates competition. Firstly, competitor information is often incomplete: it is typically impractical to implement real-time monitoring of competitor demand. Secondly, it is hard to impose centralized price experiments on sellers when estimating the demand model. For example, it is often impractical to request competitors to fix their prices to allow a certain seller to test their price sensitivity via individual price experiments. Thirdly, sellers lack access to competitor strategies before the competition begins, making it challenging to formulate optimal responses to the full contingent prices of competitors or ascertain how competitors would react to different strategies. Addressing the aforementioned challenges and designing a pricing algorithm that effectively utilizes (incomplete) competitor information constitutes the primary focus of this work.

1.1. Overview of Model

We consider N sellers, each selling a single type of product with unlimited inventories over a selling horizon of T periods. At the beginning of each period, each seller simultaneously posts their price and subsequently observes their private demand, which depends on the prices of all sellers following a noisy and unknown linear model. Sellers can observe the historical prices of competitors but do not know the demand of competitors. The demand of different sellers within a period can be correlated. When examining pricing policies, we focus on determining whether the revenue of each seller is maximized and whether seller prices converge to the Nash equilibrium that sellers would have reached if fully informed. Our performance measure for a seller is the cumulative regret over the selling horizon, defined as the worst-case difference in the seller's average revenue between our proposed policy (without knowledge of model parameters or competitor demand information) and an optimal policy in hindsight (where the competitor prices are fixed for all periods and the seller retrospectively optimizes individual prices for each period with access to model parameters).

Our work focuses on a linear demand model, which is robust against model misspecification in appropriate contexts of dynamic pricing with demand learning. For example, Besbes and Zeevi (2015) demonstrated that learning based on a simple parametric linear model, even if it deviates considerably from the true underlying model, does not necessarily result in significant revenue loss or deviation from the true optimal price, provided that the price adjustments toward the perceived "optimal" price align with the direction of the gradient of the true underlying revenue function. Indeed, the revenue function under linear demand is quadratic. If (each segment of) the true underlying revenue function can be approximated by a quadratic function, we expect that the fitted linear demand model can provide gradient feedback aligning with the true revenue gradient.

1.2. Key Results and Contribution

A novel decentralized phased pricing policy. We propose a decentralized online learning algorithm, entitled "least-squares estimation then gradient optimization (LEGO)" policy, for the dynamic pricing problem in a sequential competition under unknown linear demand; see Algorithm 1 and illustration in Figure 1. In our policy, each seller partitions the entire selling horizon into two phases: an exploration phase focused on estimating private parameters, and a gradient optimization phase focused on adjusting prices based on estimated gradient feedback. Sellers are allowed to have private exploration phase lengths. In contrast to batched bandit algorithms (such as explore-then-commit algorithms) derived from single-agent stationary settings, which ultimately commit to a fixed action (price) until the end of the problem horizon after some exploration, our algorithm allows sellers to continue adjusting prices after exploration (price experiments). This is particularly relevant within the competition context as it addresses the need for "optimal" decisions that dynamically adapt prices in response to evolving competitor prices. (For example, after one seller adjusts their price, other sellers may subsequently adjust prices to improve revenue, which forms a loop of response; as a result, an eventually fixed price can hardly guarantee maximal revenues.)

Optimal dynamic regret and fast convergence to equilibrium. We demonstrate that, if each seller privately employs an order-of-$\sqrt{T}$ policy (that is, their LEGO policy's exploration phase length is of the same order as $\sqrt{T}$), each seller achieves a worst-case regret of $O\big(N\sqrt{T}(N+\log T)\big)$ (Theorem 1). Furthermore, the joint prices of sellers at the end of the selling horizon will converge to Nash equilibrium prices at a rate of $O(\frac{N^2}{\sqrt{T}})$. The joint prices are also a vector of ϵ-Nash equilibrium prices with probability $1-O\big(\frac{N^2}{\epsilon\sqrt{T}}\big)$ for $\epsilon>0$. Our worst-case regret matches the problem lower bound regarding the order of problem horizon length T; see Remark 1 for a detailed discussion.

We investigate how different obstacles (unknown model parameters, uncertain competitor future prices, unobservable competitor demand, etc.) contribute to the difficulty of dynamic pricing in sequential competition and which obstacle forces the order-of-$\sqrt{T}$ regret. We show that if each seller knows their individual price sensitivity coefficient, a gradient optimization policy achieves an improved regret of $O(N\log T)$ while leading sellers to Nash equilibrium at a rate of $O(\frac{N}{T})$ (Theorem 2). This indicates that the *unknown individual price sensitivity contributes to the major difficulty of dynamic pricing in sequential competition* and forces an order-of-$\sqrt{T}$ regret in the worst case.

Novel analysis techniques for sequential competition. We propose a "Multi-Agent GradIent desCent" (MAGIC) framework for analyzing sequential competition problems. Our framework leverages the key observation that the played actions of agents (*i.e.*, estimators in the exploration phase or offered prices in the gradient optimization phase) constitute an embedded (time-inhomogeneous) Markovian process. We directly investigate the variation of played actions, distinguishing from the classical convex analysis, which typically centers on the variation of cost function values at played actions. Specifically, we respectively examine (i) how individual estimators progress closer (or further away) to the true parameters after each exploration period and (ii) how the offered prices of sellers progress toward Nash equilibrium in each gradient optimization period after all sellers complete exploration. We establish recursive inequalities describing how the estimators and prices vary each time a new period of observation is collected. Using these recursive inequalities, we can quantify the estimation error after the exploration phase and the deviation of the limiting prices of sellers from the Nash equilibrium. It is noteworthy that our techniques can be applied to analyzing other single- or multi-agent settings.

We assume that sellers apply the same class of pricing policies due to our goal of exploiting the problem's lower bounds, *i.e.*, to balance the price experiments and optimization to achieve the minimal worst-case regret. By demonstrating the optimal order-of-$\sqrt{T}$ regret in the general case and the optimal order-of-$\log T$ regret in the case of known price sensitivity, we convey the main message of this paper: unknown individual price sensitivity contributes to the major difficulty of dynamic pricing in sequential competition, forcing regret to at least the order of $\sqrt{T}$. From the perspective of individual sellers, our policy belongs to the popular class of phased pricing policies. Phased exploration and exploitation algorithms are widely used in both practice and theory of revenue management (Besbes and Zeevi 2012, Sauré and Zeevi 2013, Chen et al. 2021, Li et al. 2022, Chen and Shi 2023). Our policy does not require sellers to communicate throughout the selling horizon, making it relatively easy to implement. If sellers apply various classes of pricing policies, the system becomes challenging to track. As one of the earliest attempts to explore optimal pricing policies in sequential competition, we leave this investigation to future research.

1.3. Related Literature

Price competition under known demand. Cournot (1838) and Bertrand (1883) introduced the classic static models for competition in production output and market price, respectively. A static price competition entails a single decision epoch involving multiple sellers without collaboration. Here we summarize recent advancements in the operations literature. Gallego et al. (2006) examined static price competition under an attraction demand model, including the significant case of the multinomial logit (MNL) model. Their results demonstrate that, under suitable assumptions, a unique pure-strategy Nash equilibrium exists and is globally stable. Aksoy-Pierson et al. (2013) and Gallego and Wang (2014) respectively established conditions for the existence and uniqueness of Nash equilibrium in static price competition under mixed MNL and nested logit demand. Alptekinoğlu and Semple (2016) delved into an exponomial choice model and explored strategies to guide sellers toward Nash equilibrium. Federgruen and Hu (2015, 2021) examined joint price and assortment competition under linear demand and established conditions for the existence and global robust stability of Nash equilibrium. Besbes and Sauré (2016) investigated joint price and assortment competition under MNL demand, analyzing the existence of Nash equilibrium and its Pareto dominance. Wang et al. (2022) introduced an integrated framework to study joint decisions of price, product quality, and service duration in competitive environments. Morrow and Skerlos (2011) proposed numerical methods based on fixed-point equations to efficiently compute Bertrand-Nash equilibrium prices under mixed-logit demand. Allon and Gurvich (2010) proposed a framework that merged heavy-traffic analysis with Nash equilibrium analysis. Dynamic price competition across multiple decision epochs has also gained considerable attention. For instance, Gallego and Hu (2014), Federgruen and Hu (2016), Chen and Chen (2021) examined competition models with sellers possessing full information.

Sequential price competition with demand learning. A growing body of operations literature delves into the study of demand learning in competitive environments. Kirman (1975, 1983) conducted the earliest works in this area, examining a symmetric two-seller repeated price competition under unknown linear demand. Cooper et al. (2015) expanded upon findings of Kirman (1975) to include two asymmetric sellers and noisy linear demand. In their model, each seller estimates demand by neglecting the presence of competitors from both the model and data (*i.e.*, learning as if they were a monopolist). Such a "decoupling" learning process led sellers to Nash equilibrium *under the condition that sellers knew individual price sensitivity coefficients*. However, price sensitivity coefficients are typically unknown in practice; particularly, our Section 5 demonstrates that unknown individual price sensitivity contributes to the major difficulty of dynamic pricing in sequential competition. Convergence rate, policy regret, and applicability to an arbitrary number of sellers are not addressed in Cooper et al. (2015). In contrast, we propose a dynamic pricing policy that achieves a fast convergence rate and optimal regret relative to a dynamic benchmark policy in our sequential price competition with an arbitrary number of sellers. Kachani et al. (2007) also considered competition under linear demand and proposed a centralized joint pricing and learning algorithm, where sellers need to share demand information to accomplish the estimation step. Their policy has no theoretical guarantees for limiting prices or associated regret. Yang et al. (2024) introduced a *centralized* demand learning algorithm designed to lead multiple sellers to equilibrium under unknown demand. Their algorithm involves a virtual central planner coordinating the price experiments of all sellers. Within each stage, comprised of multiple periods, the central planer assigns a fixed reference price vector to all sellers, who sequentially offer their reference price or a higher experimental price; sellers are not allowed to offer experimental prices simultaneously. At the end of each stage, the central planner determines a Nash equilibrium price vector based on the current estimates, which are assigned to sellers as reference prices in the subsequent stage. In general, it is considerably challenging to enforce centralized price experiments across all sellers and request them to share sale information for determining Nash equilibrium prices after each stage.

Other literature in this area includes Goyal et al. (2023) and Li and Mehrotra (2024), which analyzed price competition under MNL demand *with known price sensitivity coefficients*. They established a weaker order-of-$T^{\frac{2}{3}}$ regret bound. In contrast, our Section 4 proves an optimal order-of-$\sqrt{T}$ regret under analogous settings for linear demand competition but *with unknown price sensitivity coefficients*. Golrezaei et al. (2020), Guo et al. (2023) investigated two-seller price competition with reference effects under noise-free linear and MNL demand, respectively, which demonstrated that adjusting prices using revenue gradient leads sellers to equilibrium. However, the gradient of revenue function is typically unavailable when model parameters are unknown. Additionally, sellers do not directly observe noise-free demand at specific prices in practice; instead, they only observe a random variable whose mean value is the average demand. (Li and Mehrotra (2024) show that noisy observation can significantly slow down convergence in a competition.) Birge et al. (2024) examined a platform interacting with multiple sellers and studied how to improve the platform revenue by revealing information and providing price-setting incentives to sellers.

Algorithmic collusion in competition. Our work centers on exploring whether an appropriate revenue-maximizing policy can lead sellers to Nash equilibrium prices. A distinct and intriguing direction is to investigate whether particular learning policies steer sellers toward collusion or cooperation prices instead of equilibrium prices; *e.g.*, a misapplied policy that overlooks the presence of competitors may inadvertently encourage collusion among sellers. Here we summarize recent findings in the literature. Through simulation studies, Tesauro and Kephart (2002), Waltman and Kaymak (2008), Klein (2018), Calvano et al. (2020), Klein (2021), Hettich (2021), Asker et al. (2022), Eschenbaum et al. (2022), Abada and Lambin (2023), Epivent and Lambin (2024) demonstrated that certain algorithms can lead sellers to collusion, although without providing theoretical guarantees. Hansen et al. (2021) studied a duopoly where both players neglected the competitor. They used simulation to demonstrate the correlation between the limiting prices and the demand noise. They also provided theoretical results for a particular prisoner's dilemma-type model under noiseless demand. Cartea et al. (2022a,b,c) studied algorithmic collusion in a market-making game, a repeated prisoner's dilemma game, and a repeated potential game with bounded rationality, respectively. Banchio and Mantegazza (2023) revealed that, driven by "endogenous statistical linkages" in estimation, certain algorithms periodically synchronize on collusive actions. Cont and Xiong (2024) investigated algorithmic collusion in a dealer market, wherein a decentralized learning algorithm is used to model how market makers adapt quotes.

Cooper et al. (2015) studied collusion in a two-seller price competition under linear demand. In their first collusion case (their Section 4.2), they showed that if *sellers are symmetric and know their intercept parameter* (*i.e.*, parameter α_i in our model (1)), they can converge to cooperative prices. However, this cooperation result hinges on seller symmetry. It is also hard for sellers to know the demand intercept parameter a priori. In their second collusion case (their Section 4.3), they showed that when all parameters are unknown, if each seller "pretends" they were a monopolist and uses a myopic pricing policy, then the limiting prices of sellers will depend on initial conditions. The limiting revenues can be Pareto superior, Pareto inferior, or unilateral superior to Nash equilibrium. However, Meylahn and V. den Boer (2022) remarked that "it is questionable whether this algorithm (of Cooper et al. 2015) is implemented in practice". den Boer and Zwart (2014) proved that the myopic pricing policy used by Cooper et al. (2015) fails to converge to optimal prices in a monopolistic setting; thus, it is not likely to be used by rational sellers who operate as monopolists.

Additional literature that investigates algorithmic collusion in a duopoly setting includes Aouad and den Boer (2021), Loots and den Boer (2021), den Boer et al. (2022), Meylahn and V. den Boer (2022), den Boer (2023), Meylahn (2023a,b). From the standpoint of regulators with concerns regarding algorithmic collusion, our study establishes that, through proper oversight of information exchange among competitors, the realization of a Nash equilibrium remains an achievable objective.

Non-parametric online learning of multiple agents. There are also intersections between our work and the machine learning literature concerning online learning in games; for instance, Bravo et al. (2018), Mertikopoulos and Zhou (2019), Golowich et al. (2020), Lin et al. (2020), Hsieh et al. (2021), Lin et al. (2021), Jordan et al. (2023). These studies investigate a common *non-parametric* modeling framework involving N agents, with each agent aiming to maximize individual cumulative payoff over a finite problem horizon. Typically, it is assumed that agents cannot observe the actions or payoffs of competitors. However, this assumption does not fully reflect the reality of our price competition, where competitor prices are often observable. Particularly, this paper demonstrates that by leveraging public price information, sellers can achieve optimal regret and fast convergence. The literature on multi-agent online learning conventionally evaluates learning policies against a static benchmark policy. In contrast, we employ a stronger dynamic benchmark policy; Li and Mehrotra (2024) show that, in the context of sequential competition, a static optimal policy can result in linear regret relative to a dynamic optimal policy. We illustrate that our learning policy achieves optimal regret relative to a dynamic benchmark policy in our sequential price competition. While the literature on multi-agent online learning typically assumes competition properties such as monotonicity or variational stability, our work does not make such assumptions. Li and Mehrotra (2024) study a general sequential competition model and show that a passive learning policy can lead learners to Nash equilibrium with order-of-$T^{\frac{2}{3}}$ regret relative to a dynamic benchmark policy, provided that the learner's data is informative regarding their best response to competitor actions. However, their informative feedback condition does not hold for our sequential price competition problem with unknown parameters. Furthermore, our work establishes stronger optimal regret and faster convergence for our proposed dynamic pricing policies.

1.4. Organization and Notation

The remainder of this paper is organized as follows. Section 2 introduces our sequential price competition model under linear demand. Section 3 proposes the decentralized phased LEGO policy for pricing in the sequential competition. Section 4 demonstrates that the LEGO policy can achieve fast convergence to Nash equilibrium as well as optimal regret relative to a dynamic benchmark policy. Section 5 analyzes the origin of regret and shows that unknown individual price sensitivity contributes to the major difficulty of dynamic pricing in sequential competition. Section 6 conducts comparative numerical experiments with benchmark results. Section 7 presents concluding remarks.

Appendix A summarizes major notation. $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ denote the Manhattan norm, the Euclidean norm, and the maximum norm. Hardy's notation "$\asymp$" stands for that function $f \asymp g$ is of the same order as g; *i.e.*, f is asymptotically bounded by g both above and below (with constant factors). The notation $\tilde{O}(\cdot)$ omits any logarithmic factors in $O(\cdot)$. A function $f(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R}$ is

a-strongly convex for some $a>0$ if $f(\mathbf{x}')\geqslant f(\mathbf{x})+\nabla f(\mathbf{x})^\top(\mathbf{x}'-\mathbf{x})+\frac{a}{2}\cdot\|\mathbf{x}'-\mathbf{x}\|_2^2$ for all $\mathbf{x},\mathbf{x}'$ in the domain. $f(\mathbf{x})$ is a-smooth for $a>0$ if $f(\mathbf{x}')\leqslant f(\mathbf{x})+\nabla f(\mathbf{x})^\top(\mathbf{x}'-\mathbf{x})+\frac{a}{2}\cdot\|\mathbf{x}'-\mathbf{x}\|_2^2$ for all $\mathbf{x},\mathbf{x}'$.

2. Model and Problem Formulation

2.1. Sequential Price Competition under Linear Demand

We consider N sellers, each selling a single type of product with unlimited inventories and aiming to maximize individual cumulative revenue over a selling horizon of T periods. We use $t\in\mathcal{T}:=\{1,2,\ldots,T\}$ to index time periods and $i\in\mathcal{N}:=\{1,2,\ldots,N\}$ to index sellers. At the beginning of each period, each seller *simultaneously* selects their price. For seller i, $p_i^{(t)}\in\mathcal{P}_i:=[\underline{p}_i,\bar{p}_i]$ denotes the price that seller i offers in period t, where price bounds $\underline{p}_i,\bar{p}_i\in[0,+\infty)$. Let $\mathbf{p}_{-i}^{(t)}:=(p_j^{(t)})_{j\in\mathcal{N}\setminus\{i\}}$ denote the competitor prices in period t, $\mathbf{p}^{(t)}:=(p_j^{(t)})_{j\in\mathcal{N}}$ denote the joint prices, and $\mathcal{P}:=\prod_{i\in\mathcal{N}}[\underline{p}_i,\bar{p}_i]$ denote the support of joint prices. The demand $y_i^{(t)}$ of seller i in period t depends on the offered prices of all sellers following a linear model:

$$y_i^{(t)}=\alpha_i-\beta_i p_i^{(t)}+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^{(t)}+\varepsilon_i^{(t)}=\alpha_i-\beta_i p_i^{(t)}+\sum_{j\in\mathcal{N}\setminus\{i\}}\gamma_{ij}p_j^{(t)}+\varepsilon_i^{(t)},\qquad t\in\mathcal{T}. \tag{1}$$

Here $\{\alpha_i,\beta_i,\boldsymbol{\gamma}_i\}$ are unknown model parameters and $\{\varepsilon_i^{(t)}\}_{t\in\mathcal{T}}$ are demand noises following independent and identical distributions. Particularly, $\mathbb{E}[\varepsilon_i^{(t)}]=0$ and $\mathbb{E}[(\varepsilon_i^{(t)})^2]\leqslant U_0$. $\varepsilon_i^{(t)}$ and $\varepsilon_j^{(t)}$ can be correlated with $i\neq j$. Parameter vector $\boldsymbol{\gamma}_i\in\mathbb{R}^{N-1}$ measures how seller i's demand is affected by competitor prices. The parameter space of seller i is defined by

$$\underline{\alpha}_i\leqslant\alpha_i\leqslant\bar{\alpha}_i,\qquad \underline{\beta}_i\leqslant\beta_i\leqslant\bar{\beta}_i,\qquad \|\boldsymbol{\gamma}_i\|_1\leqslant\bar{\gamma}_i, \tag{2}$$

where $\underline{\alpha}_i,\bar{\alpha}_i,\underline{\beta}_i,\bar{\beta}_i,\bar{\gamma}_i\in(0,+\infty)$ are known constants. Let $\mathbf{y}^{(t)}:=(y_i^{(t)})_{i\in\mathcal{N}}$ denote the joint demand vector and $\boldsymbol{\varepsilon}^{(t)}:=(\varepsilon_i^{(t)})_{i\in\mathcal{N}}$ denote the joint noise vector. We also use superscript-less notations $\mathbf{p}$, $\mathbf{p}_{-i}$, $\mathbf{y}$, and $\boldsymbol{\varepsilon}$ to represent generic vectors of joint prices, competitor prices, joint demand, and joint noise, respectively. We assume that the parameter space is appropriate so that the average demand $\mathbb{E}[y_i^{(t)}]$ is non-negative among all values of $\{\alpha_i,\beta_i,\boldsymbol{\gamma}_i,\mathbf{p}^{(t)}\}$ $(i\in\mathcal{N},t\in\mathcal{T})$; a similar assumption is found in Birge et al. (2024).

Seller $i\in\mathcal{N}$ aims to maximize the individual (cumulative) revenue, which can be expressed as

$$\mathsf{R}_i(T):=\mathbb{E}\sum_{t=1}^{T}[p_i^{(t)}y_i^{(t)}]=\mathbb{E}\sum_{t=1}^{T}\left[p_i^{(t)}\cdot(\alpha_i-\beta_i p_i^{(t)}+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^{(t)})\right].$$

Each seller's offered prices are public but their demand history is private; that is, when seller $i\in\mathcal{N}$ decides price $p_i^{(t)}$ in period $t\in\mathcal{T}$, they can utilize the information of joint price history $\{\mathbf{p}^{(u)}\}_{u<t}$ and individual demand history $\{y_i^{(u)}\}_{u<t}$. We let $\mathcal{F}_i^{(t)}:=\sigma((\mathbf{p}^{(u)},y_i^{(u)})_{1\leqslant u\leqslant t})=\sigma((\mathbf{p}^{(u)},\varepsilon_i^{(u)})_{1\leqslant u\leqslant t})$, $t\in\mathcal{T}$ be the filtration associated with the joint pricing and individual demand process up to period t, and $\mathcal{F}_i^{(0)}:=\varnothing$. Let $\mathcal{F}^{(t)}:=\sigma((\mathbf{p}^{(u)},\mathbf{y}^{(u)})_{1\leqslant u\leqslant t})=\sigma((\mathbf{p}^{(u)},\boldsymbol{\varepsilon}^{(u)})_{1\leqslant u\leqslant t})$, $t\in\mathcal{T}$ be the filtration associated with the joint pricing and demand process up to period t, and $\mathcal{F}^{(0)}:=\varnothing$. Then $\mathcal{F}_i^{(t)}\subseteq\mathcal{F}^{(t)}$.

2.2. Policy Performance Measures

Maximizing a seller's revenue can be reframed as minimizing their regret. Each seller competes with a dynamic optimal sequence of prices in hindsight while assuming that the other sellers would not have responded differently if this sequence of prices had been offered. Under such a dynamic benchmark, the objective of each learner is to minimize the following regret metric in hindsight:

$$\begin{aligned}\mathsf{Reg}_i(T) &:= \mathbb{E}\sum_{t=1}^{T}\sup_{p_i'^{(t)}\in\mathcal{P}_i}\mathbb{E}[p_i'^{(t)}y_i'^{(t)}\,|\,p_i'^{(t)},\mathbf{p}_{-i}^{(t)}]-\mathsf{R}_i(T)\\ &= \mathbb{E}\sum_{t=1}^{T}\Big[\sup_{p_i'^{(t)}\in\mathcal{P}_i}p_i'^{(t)}(\alpha_i-\beta_i p_i'^{(t)}+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^{(t)})-p_i^{(t)}(\alpha_i-\beta_i p_i^{(t)}+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^{(t)})\Big].\end{aligned} \tag{3}$$

Here $\mathbb{E}[p_i'^{(t)}y_i'^{(t)}\,|\,p_i'^{(t)},\mathbf{p}_{-i}^{(t)}]$ denotes the "counterfactual" revenue for seller i if seller i had offered price $p_i'^{(t)}$ in period t. We note that in a sequential competition framework, a dynamic benchmark policy is stronger than a static benchmark policy: Li and Mehrotra (2024) demonstrates that the static optimal policy can result in linear regret relative to the dynamic optimal policy. We do not assume that sellers possess a prior model of their competitors. As a result, a seller cannot define a priori optimal strategies or determine how competitors would have reacted to different strategies. Instead, as the competition progresses, a seller can retrospectively assess whether they could have performed better. Without a strategic representation of competitors, when we define the dynamic benchmark policy, competitors were assumed to be unresponsive if the posterior optimal sequence of prices had been offered.

In addition to individual revenue maximization (already reframed as regret minimization), we also investigate whether sellers converge to equilibrium if each of them employs a regret-minimizing policy. Our equilibrium prices are defined based on the static (one-shot) price competition. Consider a single period ($T=1$), wherein superscript t is omitted. The Nash equilibrium prices $\mathbf{p}^*=(p_i^*)_{i\in\mathcal{N}}\in\mathcal{P}$ are defined as a price vector under which unilateral deviation is not profitable for any seller. Specifically, $\mathbf{p}^*$ is a solution to the following balance equations:

$$p_i^*=\arg\max_{p_i\in\mathcal{P}_i}\mathbb{E}[p_iy_i]=\arg\max_{p_i\in\mathcal{P}_i}p_i\cdot(\alpha_i-\beta_ip_i+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^*),\qquad i\in\mathcal{N}, \tag{4}$$

where $\mathbf{p}_{-i}^*:=(p_j^*)_{j\in\mathcal{N}\setminus\{i\}}$. We note that, given any $\mathbf{p}_{-i}^*$, the maximization problem in (4) has a unique solution p_i^* because $\beta_i>0$. Lemma 1 shows that there exists a unique Nash equilibrium $\mathbf{p}^*$ under appropriate assumptions.

We further define ϵ-Nash equilibrium as a joint price vector $\mathbf{p}^\epsilon=(p_i^\epsilon)_{i\in\mathcal{N}}\in\mathcal{P}$ that submits to:

$$p_i^\epsilon\cdot(\alpha_i-\beta_ip_i^\epsilon+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^\epsilon)\geqslant p_i\cdot(\alpha_i-\beta_ip_i+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^\epsilon)-\epsilon,\qquad p_i\in\mathcal{P}_i,\, i\in\mathcal{N}. \tag{5}$$

Here $\epsilon\in\mathbb{R}_+$ and $\mathbf{p}_{-i}^\epsilon:=(\mathbf{p}_j^\epsilon)_{j\in\mathcal{N}\setminus\{i\}}$. (5) represents that, when the joint prices of sellers are ϵ-Nash equilibrium, unilateral deviation of any seller cannot increase individual revenue by more than ϵ.

3. Decentralized Phased LEGO Pricing Policy

3.1. Challenges and Overview

This section outlines a decentralized online learning algorithm for dynamic pricing in a sequential price competition under linear demand. We note that algorithms in the monopolist special case ($i.e.$, $N=1$) are provided in Broder and Rusmevichientong (2012), Keskin and Zeevi (2014). For the sequential price competition problem with multiple sellers, there are several primary challenges.

(i) *Competitor demand is not observable.* As a result, seller $i \in \mathcal{N}$ is not able to estimate competitor parameters $\{\alpha_j, \beta_j, \boldsymbol{\gamma}_j\}_{j\in\mathcal{N}\setminus\{i\}}$. It is hard to predict equilibrium prices or how competitors respond to seller i's offered prices.

(ii) *Lack of unbiased gradient feedback.* While applying gradient-based methods is feasible for some competition models, it is important to note that the sellers in our problem lack unbiased gradient feedback. Specifically, for seller i, the gradient of their average revenue in period t with respect to individual price $p_i^{(t)}$ is as follows, which involves unknown parameters:

$$\frac{\partial \mathbb{E}[p_i^{(t)} y_i^{(t)} \,|\, p_i^{(t)}, \mathbf{p}_{-i}^{(t)}]}{\partial p_i^{(t)}} = \frac{\partial \left[p_i^{(t)} \cdot (\alpha_i - \beta_i p_i^{(t)} + \boldsymbol{\gamma}_i^\top \mathbf{p}_{-i}^{(t)})\right]}{\partial p_i^{(t)}} = \alpha_i - 2\beta_i p_i^{(t)} + \boldsymbol{\gamma}_i^\top \mathbf{p}_{-i}^{(t)}. \tag{6}$$

(iii) *Demand curves are not "well-separated".* Due to the coexistence of unknown demand intercept α_i and unknown price coefficients $\{\beta_i, \boldsymbol{\gamma}_i\}$, the prices of seller i is "uninformative". According to Broder and Rusmevichientong (2012) that focus on monopolistic settings, in the presence of uninformative prices, a seller cannot effectively learn from demand responses at all price levels. Consequently, some forced price exploration is needed and the worst-case regret significantly escalates to the order of $\sqrt{T}$, in contrast to the order-of-$\log T$ regret exhibited by well-separated cases. (One example of well-separated cases is when α_i is known.) In our sequential competition settings, we anticipate a similar magnitude of regret escalation.

(iv) *Conflicts between individual price experiments and competitor revenue maximization.* One last, and perhaps the most significant, challenge is the conflict of price experiments among sellers: a seller's pursuit of revenue-maximizing prices can hinder the parameter estimation efforts of other sellers. For example, when a seller $i \in \mathcal{N}$ tries to maximize individual revenue using strategies such as greedy pricing, seller i is essentially trying to offer a best-response price to competitor prices. As a result, seller i's price $p_i^{(t)}$ becomes correlated with competitor prices $\mathbf{p}_{-i}^{(t)}$. In our sequential price competition model under linear demand, seller i's best-response price $p_i'^{(t)} = \frac{\alpha_i + \boldsymbol{\gamma}_i^\top \mathbf{p}_{-i}^{(t)}}{2\beta_i}$, which maximizes seller i's revenue in period t conditional on competitor prices $\mathbf{p}_{-i}^{(t)}$, is linear in $\mathbf{p}_{-i}^{(t)}$. As a result, when seller i tries to maximize individual revenue, their offered price $p_i^{(t)}$ should approach a linear combination of competitor prices $\mathbf{p}_{-i}^{(t)}$ and the constant intercept. This causes the notorious "collinearity" issue among variables $\{p_i^{(t)}\}_{i\in\mathcal{N}} \cup \{1\}$ for the parameter estimation task of seller $j \in \mathcal{N}\setminus\{i\}$.

We weave the online learning techniques from monopolistic pricing settings into our sequential price competition problem. In our pricing policy, each seller partitions the entire selling horizon into two phases: an exploration phase and a gradient optimization phase. Note that *we allow sellers to have private exploration phase lengths*; that is, seller i's exploration phase may overlap with the gradient optimization phase of seller $j \neq i$. Indeed, in practice, it is challenging for all sellers to coordinate price experiments and reach a consensus to conclude the exploration phase simultaneously. During the exploration phase, the seller conducts price experiments using randomized prices. By the end of the exploration phase, the seller acquires an estimator of individual parameters involved in the gradient (6). We establish that this estimator is consistent, leveraging a least square estimation model and appropriate assumptions. Subsequently, during the gradient optimization phase, the seller adjusts prices based on the estimated gradient; that is, using (6) with parameters estimated from the exploration phase. Figure 1 illustrates our learning policy.

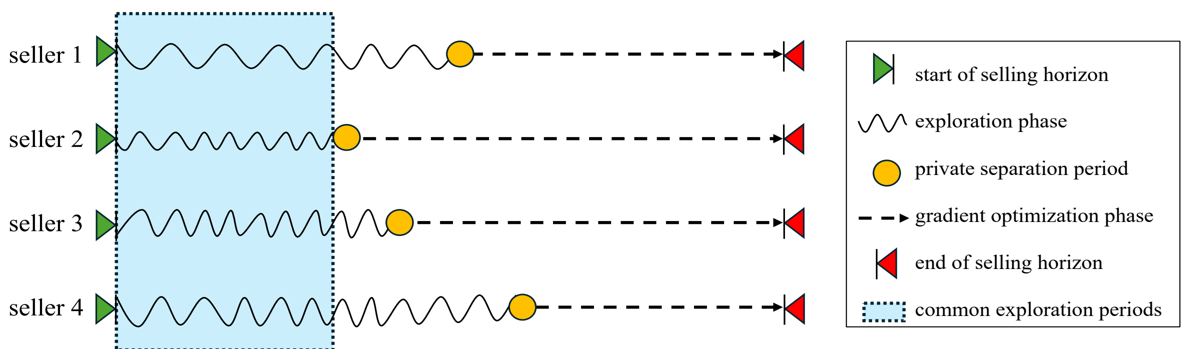


Figure 1 Illustration of the decentralized phased LEGO policy (Algorithm 1) with $N = 4$ sellers in sequential price competition under linear demand. LEGO means "least-squares estimation then gradient optimization". For each seller $i \in \{1, 2, 3, 4\}$, in their exploration phase (curved line) of a private length τ_i, they offer randomized prices following a private distribution $\mathcal{D}_i$. Seller i's price experiment ends in period $t = \tau_i$ (circle). Subsequently, in their gradient optimization phase (dashed line), seller i adjusts prices using an online gradient ascent approach till the end of the selling horizon. There is a duration of common exploration periods (dotted box), during which all sellers offer randomized prices as part of the pricing experiment.

3.2. Description of the Decentralized Phased LEGO Policy

Inputs and initialization. Given the problem horizon length T, we have two sets of inputs for each seller $i \in \mathcal{N}$ before implementing the learning algorithm: (i) a distribution $\mathscr{D}_i$ supported on $\mathcal{P}_i$ that seller i uses to sample their exploration prices $\{p_i^{(t)}\}_{t=1,2,\ldots,\tau_i}$, where τ_i is seller i's private length of exploration phase; and (ii) the sequence of step sizes $\{\eta_i^{(t)}\}_{t=\tau_i,\tau_i+1,\ldots,T-1}$ that seller i applies in the gradient optimization phase.

Exploration phase. In each period $t \in \{1, 2, \ldots, \tau_i\}$ of exploration phase, seller $i \in \mathcal{N}$ samples and offers an exploration price $p_i^{(t)}$ following distribution $\mathscr{D}_i$. At the end of the separation period τ_i, seller i estimates individual parameters $(\alpha_i, \beta_i, \boldsymbol{\gamma}_i)$ based on information of public prices and individual demand. Estimators are obtained by minimizing the squared loss of demand prediction:

$$\min_{\substack{\underline{\alpha}_i \leqslant \hat{\alpha}_i \leqslant \bar{\alpha}_i, \\ \underline{\beta}_i \leqslant \hat{\beta}_i \leqslant \bar{\beta}_i, \\ \|\hat{\boldsymbol{\gamma}}_i\|_1 \leqslant \bar{\gamma}_i}} \sum_{t=1}^{\tau_i} (\hat{\alpha}_i - \hat{\beta}_i p_i^{(t)} + \hat{\boldsymbol{\gamma}}_i^\top \mathbf{p}_{-i}^{(t)} - y_i^{(t)})^2. \tag{7}$$

For tractability, we assume that seller i practically solves the least square problem (7) using a stochastic gradient descent (SGD) approach. Specifically, seller i computes the following iterations at the end of separation period τ_i:

$$\begin{aligned}
\hat{\alpha}_i^{(t+1)} &:= \Pi_{[\underline{\alpha}_i, \bar{\alpha}_i]} \Big(\hat{\alpha}_i^{(t)} - \chi_i^{(t)} (\hat{\alpha}_i^{(t)} - \hat{\beta}_i^{(t)} p_i^{(t)} + \hat{\boldsymbol{\gamma}}_i^{(t)\top} \mathbf{p}_{-i}^{(t)} - y_i^{(t)}) \Big), \\
\hat{\beta}_i^{(t+1)} &:= \Pi_{[\underline{\beta}_i, \bar{\beta}_i]} \Big(\hat{\beta}_i^{(t)} + \chi_i^{(t)} (\hat{\alpha}_i^{(t)} - \hat{\beta}_i^{(t)} p_i^{(t)} + \hat{\boldsymbol{\gamma}}_i^{(t)\top} \mathbf{p}_{-i}^{(t)} - y_i^{(t)}) p_i^{(t)} \Big), \\
\hat{\boldsymbol{\gamma}}_i^{(t+1)} &:= \Pi_{\{\boldsymbol{\gamma} \in \mathbb{R}^{N-1} \,|\, \|\boldsymbol{\gamma}\|_1 \leqslant \bar{\gamma}_i\}} \Big(\hat{\boldsymbol{\gamma}}_i^{(t)} - \chi_i^{(t)} (\hat{\alpha}_i^{(t)} - \hat{\beta}_i^{(t)} p_i^{(t)} + \hat{\boldsymbol{\gamma}}_i^{(t)\top} \mathbf{p}_{-i}^{(t)} - y_i^{(t)}) \mathbf{p}_{-i}^{(t)} \Big),
\end{aligned} \tag{8}$$

where $(\hat{\alpha}_i^{(1)}, \hat{\beta}_i^{(1)}, \hat{\boldsymbol{\gamma}}_i^{(1)})$ is an arbitrary element in $[\underline{\alpha}_i, \bar{\alpha}_i] \times [\underline{\beta}_i, \bar{\beta}_i] \times \{\boldsymbol{\gamma} \in \mathbb{R}^{N-1} \,|\, \|\boldsymbol{\gamma}\|_1 \leqslant \bar{\gamma}_i\}$ and $\{\chi_i^{(t)}\}_{t=1,2,\ldots,\tau_i} \subseteq \mathbb{R}_+$ are step sizes in gradient descent computation. Π is a projection function:

$$\Pi_S(x) := \arg\min_{x' \in S} \|x' - x\|_2,$$

for $x \in \mathbb{R}^n$ and closed set $S \subseteq \mathbb{R}^n$. Seller i's estimator $(\hat{\alpha}_i, \hat{\beta}_i, \hat{\boldsymbol{\gamma}}_i)$ is obtained from the last iteration:

$$(\hat{\alpha}_i, \hat{\beta}_i, \hat{\boldsymbol{\gamma}}_i) = (\hat{\alpha}_i^{(\tau_i+1)}, \hat{\beta}_i^{(\tau_i+1)}, \hat{\boldsymbol{\gamma}}_i^{(\tau_i+1)}). \tag{9}$$

We note that the computation (8)–(9) is implemented by seller i only at the end of separation period τ_i and does not affect seller i's prices in the exploration phase (from period 1 to period τ_i).

We allow the exploration price distributions $\{\mathscr{D}_i\}_{i \in \mathcal{N}}$ to be correlated; *i.e.*, the exploration price $p_i^{(t)}$ can be correlated with $p_j^{(t)}$ with $j \neq i$ and $t \leqslant \min\{\tau_i, \tau_j\}$. We assume that each seller implements a sufficient price experiment in the exploration phase.

ASSUMPTION 1 **(Sufficient Price Experiments)**. *Let random variable p_i follow distribution $\mathcal{D}_i$, $i \in \mathcal{N}$. The matrix $\mathbb{E}\big[(1, \mathbf{p})^\top (1, \mathbf{p})\big]$ has a smallest eigenvalue of $\lambda_{\min} > 0$ and a maximal eigenvalue $\lambda_{\max} > 0$, where row vector $(1, \mathbf{p}) := (1, p_1, p_2, \ldots, p_N) \in \mathbb{R}^{N+1}$. At the end of exploration phase of seller i, the computation parameters are set as $\chi_i^{(t)} = \frac{\upsilon_i}{t}$ where $\upsilon_i > \frac{1}{2\lambda_{\min}}$.*

When the exploration price distributions $\{\mathscr{D}_i\}_{i \in \mathcal{N}}$ are independent, the smallest eigenvalue condition $\lambda_{\min} > 0$ in Assumption 1 is equivalent to that $\mathsf{Var}(p_i) > 0$ for all $i \in \mathcal{N}$, where Var denotes the variance. Indeed, if seller i offers constant prices in the exploration phase, the price sensitivity coefficient β_i is not identifiable. Let $\tau_{\min} := \min_{i\in\mathcal{N}} \tau_i$ denote the duration of common exploration periods. The smallest eigenvalue condition $\lambda_{\min} > 0$ in Assumption 1 requires that during the common exploration period $t \in \{1, 2, \ldots, \tau_{\min}\}$, the exploration prices of sellers are not collinear; *i.e.*, a seller's price is not a linear combination of prices of other sellers. Therefore, the individual demand and public price data up to period $\tau_{\min}$ (*i.e.*, the data subset $\{(\mathbf{p}^{(t)}, y_i^{(t)})\}_{t=1,2,\ldots,\tau_{\min}}$) provides sufficient price variation for seller i to estimate individual parameters. Although the data from seller i's remaining exploration periods (*i.e.*, the data subset $\{(\mathbf{p}^{(t)}, y_i^{(t)})\}_{t=\tau_{\min}+1,\tau_{\min}+2,\ldots,\tau_i}$) is also incorporated into seller i's estimator computation (9) and such data does not necessarily possess sufficient price variation as we discussed in (iv) of Subsection 3.1, we show that it does not significantly reduce the eventual estimator accuracy in Subsection 4.2. Our Theorem 1 and (12) will demonstrate that $\lambda_{\min}^{-1}$ and $\lambda_{\max}$ both affect the convergence rate to Nash equilibrium and the regret of sellers in a polynomial order.

Gradient optimization phase. In each period $t \in \{\tau_i + 1, \tau_i + 2, \ldots, T\}$ of gradient optimization phase, seller $i \in \mathcal{N}$ updates prices using a projected gradient ascent approach. Specifically, the estimated gradient in period t is defined as

$$\phi_i^{(t)} := y_i^{(t)} - \hat{\beta}_i p_i^{(t)}, \qquad t \in \{\tau_i + 1, \tau_i + 2, \ldots, T\},\, i \in \mathcal{N}. \tag{10}$$

Particularly, $\phi_i^{(\tau_i)} := 0$ for $i \in \mathcal{N}$. Then the price is updated by

$$p_i^{(t)} := \Pi_{\mathcal{P}_i}(p_i^{(t-1)} + \eta_i^{(t-1)} \phi_i^{(t-1)}), \qquad t \in \{\tau_i + 1, \tau_i + 2, \ldots, T\}, i \in \mathcal{N}. \tag{11}$$

We elaborate on the definition of the estimated gradient in (10). Recall that the exact gradient of seller i's average revenue with respect to individual price is $\frac{\partial \mathbb{E}[p_i^{(t)} y_i^{(t)} \mid p_i^{(t)}, \mathbf{p}_{-i}^{(t)}]}{\partial p_i^{(t)}} = \alpha_i - 2\beta_i p_i^{(t)} + \boldsymbol{\gamma}_i^\top \mathbf{p}_{-i}^{(t)}$, as shown in (6). When estimator $\hat{\beta}_i$ has a small error (*i.e.*, $\hat{\beta}_i \approx \beta_i$), we have that $\mathbb{E}[\phi_i^{(t)} \mid p_i^{(t)}, \mathbf{p}_{-i}^{(t)}] \approx \mathbb{E}[y_i^{(t)} \mid p_i^{(t)}, \mathbf{p}_{-i}^{(t)}] - \beta_i p_i^{(t)} = \alpha_i - \beta_i p_i^{(t)} + \boldsymbol{\gamma}_i^\top \mathbf{p}_{-i}^{(t)} - \beta_i p_i^{(t)}$, which approximately equals the exact gradient $\frac{\partial \mathbb{E}[p_i^{(t)} y_i^{(t)} \mid p_i^{(t)}, \mathbf{p}_{-i}^{(t)}]}{\partial p_i^{(t)}}$. Thus, (10) can be viewed as a noisy gradient feedback based on average revenue.

The LEGO policy for the sequential pricing competition problem is summarized in Algorithm 1.

4. Regret Analysis of Phased LEGO

Our main results (Theorem 1) consist of two instance-independent upper bounds concerning the policy regret associated with Algorithm 1 and its convergence rate to Nash equilibrium prices, respectively. Specifically, we demonstrate that an order-of-$\sqrt{T}$ policy (*i.e.*, $\tau_i \asymp \sqrt{T}$, $i \in \mathcal{N}$) will yield a cumulative regret of $O\big(N\sqrt{T}(N + \log T)\big)$ if this class of policy is employed by all sellers. Furthermore, the joint prices of sellers will converge to Nash equilibrium prices at a rate of $O(\frac{N^2}{\sqrt{T}})$. To establish our results, we also need Assumption 2–3 defined in Subsection 4.1.

Algorithm 1 Decentralized Phased LEGO: Least-Squares Estimation then Gradient Optimization

Input: Exploration phase length $\tau_i \in \mathbb{N}$, exploration price distribution $\mathscr{D}_i$ supported on $\mathcal{P}_i$, and step sizes $\{\eta_i^{(t)}\}_{t=\tau_i,\tau_i+1,\ldots,T-1} \subseteq \mathbb{R}_+$ for seller $i \in \mathcal{N}$.

Output: Price $p_i^{(t)}$ that seller $i \in \mathcal{N}$ offers in period $t \in \mathcal{T}$.

for $t \in \mathcal{T}$ **do**

Each seller $i \in \mathcal{N}$ samples a price $\tilde{p}_i^{(t)}$ following distribution $\mathscr{D}_i$.

Each seller $i \in \mathcal{N}$ offers a price based on

$$p_i^{(t)} := \begin{cases} \tilde{p}_i^{(t)}, & \text{when } t \leqslant \tau_i \text{ (seller } i\text{'s exploration phase);} \\ \Pi_{\mathcal{P}_i}(p_i^{(t-1)} + \eta_i^{(t-1)} \phi_i^{(t-1)}), & \text{when } t > \tau_i \text{ (seller } i\text{'s gradient optimization phase).} \end{cases}$$

Each seller $i \in \mathcal{N}$ observes individual realized demand $y_i^{(t)}$.

When $t = \tau_i$, seller $i \in \mathcal{N}$ constructs an estimator $(\hat{\alpha}_i, \hat{\beta}_i, \hat{\boldsymbol{\gamma}}_i) = (\hat{\alpha}_i^{(\tau_i+1)}, \hat{\beta}_i^{(\tau_i+1)}, \hat{\boldsymbol{\gamma}}_i^{(\tau_i+1)})$ using iterations (8) and data set $\{(\mathbf{p}^{(t)}, y_i^{(t)})\}_{t=1,2,\ldots,\tau_i}$.

If $t \geqslant \tau_i$, seller $i \in \mathcal{N}$ constructs a gradient feedback:

$$\phi_i^{(t)} = \begin{cases} 0, & \text{when } t = \tau_i; \\ y_i^{(t)} - \hat{\beta}_i p_i^{(t)}, & \text{when } t > \tau_i \text{, following (10).} \end{cases}$$

end for

Theorem 1 **(Optimal Order-Of-$\sqrt{T}$ Policy)**. *Suppose Assumptions 1–2 hold and each seller $i \in \mathcal{N}$ implements Algorithm 1:*

(A) for each $i \in \mathcal{N}$, the exploration phase length $\tau_i \in [\underline{\iota}\sqrt{T}, \bar{\iota}\sqrt{T}]$, where $\bar{\iota} \geqslant \underline{\iota} > 0$; and

(B) for $t \in \{1, 2, \ldots, T-1\}$, the step size $\eta_i^{(t)} = \frac{\zeta_i}{t}$, where $\nu := \frac{\max_{j \in \mathcal{N}} \zeta_j}{\min_{j \in \mathcal{N}} \zeta_j} < 4\kappa - 1$ and $\zeta_i > \frac{2\kappa}{(4\kappa - 1 - \nu) \min_{j \in \mathcal{N}} \underline{\beta}_j}$. (Here $\kappa \geqslant 1$ is the constant in Assumption 2.)

Then we have the following results for $T > 0$:

(i) convergence to Nash equilibrium prices:

$$\mathbb{E}\left[\|\mathbf{p}^{(T)} - \mathbf{p}^*\|_2^2\right] \leqslant \frac{\tilde{b} + 2\omega D}{T} + \frac{\bar{\iota} D + \rho C_1}{\sqrt{T}} \in O\left(\frac{N^2}{\sqrt{T}}\right);$$

(ii) individual sublinear regret:

$$\mathsf{Reg}_i(T) \leqslant 4\bar{\beta}_i(\tilde{b} + 2\omega D)\log T + (4\bar{\beta}_i \bar{\iota} D + 4\bar{\beta}_i \rho C_1 + 4\bar{\beta}_i \bar{\iota} D \log T)\sqrt{T} \in O\Big(N\sqrt{T}(N + \log T)\Big)$$

for each seller $i \in \mathcal{N}$ satisfying Assumption 3; and

(iii) $\mathbf{p}^{(T)}$ is a vector of ϵ-Nash equilibrium prices with (at least) probability $1 - \frac{\tilde{b} + 2\omega D}{\mu \epsilon T} - \frac{\bar{\iota} D + \rho C_1}{\mu \epsilon \sqrt{T}} = 1 - O\left(\frac{N^2}{\epsilon \sqrt{T}}\right)$ if each seller $i \in \mathcal{N}$ satisfies Assumption 3.

The associated constants are defined as

$$
\begin{aligned}
&d_i := (\bar{\alpha}_i - \underline{\alpha}_i)^2 + (\bar{\beta}_i - \underline{\beta}_i)^2 + 4\bar{\gamma}_i^2, \qquad U_1 := U_0 + \max_{i\in\mathcal{N}} \max\left\{\bar{\beta}_i^2 \bar{p}_i^2, \, (\bar{\alpha}_i - \underline{\beta}_i \underline{p}_i + \bar{\gamma}_i \max_{j\in\mathcal{N}} \bar{p}_j)^2\right\}, \\
&U_\ell := (\lambda_{\max} \max_{i\in\mathcal{N}} d_i + U_0)\cdot\big(1 + \sum_{j\in\mathcal{N}} \bar{p}_j^2\big), \qquad D := \sum_{i\in\mathcal{N}} (\bar{p}_i - \underline{p}_i)^2, \qquad \mu := \frac{1}{4\max_{i\in\mathcal{N}} \bar{\beta}_i}, \\
&\omega := \frac{4\kappa - 1 - \nu}{2\kappa} \cdot \min_{i\in\mathcal{N}} \zeta_i \cdot \min_{i\in\mathcal{N}} \underline{\beta}_i, \qquad h_i := \frac{\upsilon_i^2 U_\ell}{2\lambda_{\min}\upsilon_i - 1}, \qquad \tilde{b} := \frac{U_1 \sum_{i\in\mathcal{N}} \zeta_i^2}{\omega - 1}, \\
&C_0 := \max_{i\in\mathcal{N}} h_i + (2\lambda_{\min}\upsilon_i + 1) d_i, \qquad C_1 := \underline{\iota}^{-1} C_0 N + \underline{\iota}^{-1} U_\ell \sum_{i\in\mathcal{N}} \upsilon_i^2, \qquad \rho := \omega^{-1} \max_{i\in\mathcal{N}} \zeta_i^2 \bar{p}_i^2.
\end{aligned} \tag{12}
$$

Here $D, U_\ell, h_i, \tilde{b}, C_0 \in O(N)$ and $C_1 \in O(N^2)$ by definition. When sellers have the same step sizes ($\nu = 1$), the constraints of ζ_i in (B) of Theorem 1 can be simplified into $\zeta_i > \frac{\kappa}{(2\kappa-1)\min_{j\in\mathcal{N}} \underline{\beta}_j}$, $i \in \mathcal{N}$.

Remark 1 (Policy Optimality). Seller i's revenue maximization is cast into a regret minimization problem in our study. Let us consider $N = 1$, which is the special case of a monopolistic seller. The linear demand model in (1) includes an unknown potential market size and an unknown price sensitivity coefficient. In such a scenario, Broder and Rusmevichientong (2012), Keskin and Zeevi (2014) have demonstrated that demand curves are not well separated, resulting in a $\Omega(\sqrt{T})$ worst-case regret bound for any non-anticipating dynamic pricing policy. This lower bound is matched by our $\tilde{O}(\sqrt{T})$ regret bound as presented in result (ii) of Theorem 1, implying that our LEGO policy is optimal in the context of sequential price competition. It is also worth noting that we employ a much stronger dynamic benchmark policy to define the regret metric compared with Broder and Rusmevichientong (2012), Keskin and Zeevi (2014) using a static benchmark policy. Our work achieves the optimal regret under a dynamic benchmark policy because of the specific problem structure, particularly with linear demand functions and the assumption that sellers employ the same class of online learning policies. Our numerical experiments further show that our order-of-$\sqrt{T}$ policy achieves an optimal balance between exploration and exploitation; see Subsection 6.2.

Remark 2 (Components of Error and Regret). As stated in result (i) of Theorem 1, the error between eventual joint prices $\mathbf{p}^{(T)}$ and Nash equilibrium prices $\mathbf{p}^*$ has two components: $\frac{\tilde{b}+2\omega D}{T} \in O\big(\frac{N}{T}\big)$ and $\frac{\bar{\iota}D+\rho C_1}{\sqrt{T}} \in O\big(\frac{N^2}{\sqrt{T}}\big)$. As we shall prove in Section 5, the predominant error component $\frac{\bar{\iota}D+\rho C_1}{\sqrt{T}}$ can be attributed to the unknown individual price sensitivity β_i for each seller $i \in \mathcal{N}$. Specifically, if each seller $i \in \mathcal{N}$ has prior knowledge of β_i, then the convergence rate of $\mathbb{E}\big[\|\mathbf{p}^{(T)} - \mathbf{p}^*\|_2^2\big]$ can be enhanced from $\frac{\tilde{b}+2\omega D}{T} + \frac{\bar{\iota}D+\rho C_1}{\sqrt{T}} \in O\big(\frac{N^2}{\sqrt{T}}\big)$ to $\frac{b+(2\omega+1)D}{T} = \frac{\delta}{T} \in O\big(\frac{N}{T}\big)$; definitions of constants b and δ are provided in (17) of Section 5. Conversely, only a minor error component $\frac{\tilde{b}+2\omega D}{T}$ stems from other factors including uncertainty of competitor future prices, unknown individual parameters $(\alpha_i, \boldsymbol{\gamma}_i)$, unknown competitor parameters $(\alpha_j, \beta_j, \boldsymbol{\gamma}_j)$ where $j \neq i$, and unknown competitor demand. Analogous conclusions apply to the regret metric. When each seller $i \in \mathcal{N}$ knows their individual price sensitivity, the regret bound in result (ii) of Theorem 1 will relinquish its second term and reduce to $4\bar{\beta}_i(b+(2\omega+1)D)\log T + 4\bar{\beta}_i\delta \in O\big(N\log T\big)$; the order-of-$N\log T$ regret is delineated in Theorem 2 and is proved in Section 5. Consequently, the regret of $\tilde{O}(\sqrt{T})$ in the general case (Theorem 1) is likewise attributable to the unknown individual price sensitivity β_i.

Remark 3 (Step Size Choice and Its Minor Impact). To establish the optimal regret bound and fast convergence rate in Theorem 1, we presuppose in condition (B) that each seller's step size is $O(\frac{1}{t})$ and does not exceed ν times the step size of other sellers. We also assume that the step size coefficient $\zeta_i \in \Omega\big(\frac{1}{2\min_{j\in\mathcal{N}}\underline{\beta}_j}\big)$, where $2\min_{j\in\mathcal{N}}\underline{\beta}_j$ is the *minimal* strong concavity parameter across the sellers; recall that seller i's average revenue is $2\beta_i$-strongly concave in their price. Assuming the step size coefficient proportional to the inverse of the strong concavity parameter is not uncommon for online concave optimization to achieve the optimal convergence rate; see, *e.g.*, Hazan et al. (2016). In our numerical experiments, Subsection 6.3 illustrates that even when employing sequences of step sizes that do not adhere to these constraints, sellers still converge to the Nash equilibrium at a rapid speed. Particularly, Subsection 6.3 shows that using step sizes such as $\eta_i^{(t)} = \frac{\zeta_i}{\sqrt{t}}$ does not significantly impact the regret or convergence to Nash equilibrium. This is because it is already fast enough to employ step sizes between $O(\frac{1}{t}) \sim O(\frac{1}{\sqrt{t}})$ during the gradient optimization phase. Under these adequate step sizes, the rates of regret growth and convergence to the Nash equilibrium are predominantly constrained by the order-of-$\frac{1}{\sqrt{T}}$ estimation error induced by the exploration phase. Consequently, compared with the exploration phase length, the step sizes of sellers have a relatively minor impact on the regret and the convergence to Nash equilibrium.

Remark 4 (Novel Analysis Techniques for Sequential Competition). We propose a "Multi-Agent GradIent desCent" (MAGIC) framework for analyzing sequential competition problems. Our framework leverages the key observation that the played actions of agents (*i.e.*, estimators $(\hat{\alpha}_i^{(t)}, \hat{\beta}_i^{(t)}, \hat{\boldsymbol{\gamma}}_i^{(t)})$ in the exploration phase or prices $\mathbf{p}^{(t)}$ in the gradient optimization phase) constitute an embedded time-inhomogeneous Markovian process. We directly investigate the variation of played actions, distinguishing from the classical convex analysis which typically centers on the variation of cost function values at played actions. Specifically, we investigate (a) how the individual estimators $(\hat{\alpha}_i^{(t)}, \hat{\beta}_i^{(t)}, \hat{\boldsymbol{\gamma}}_i^{(t)})$ progress toward the true parameters $(\alpha_i, \beta_i, \boldsymbol{\gamma}_i)$ in each common exploration period $t \in \{1, 2, \ldots, \tau_{\min}\}$, (b) how the accuracy of individual estimators $(\hat{\alpha}_i^{(t)}, \hat{\beta}_i^{(t)}, \hat{\boldsymbol{\gamma}}_i^{(t)})$ varies in the private exploration period $t \in \{\tau_{\min}+1, \tau_{\min}+2, \ldots, \tau_i\}$, and (c) how the offered prices $\mathbf{p}^{(t)}$ of sellers progress toward Nash equilibrium $\mathbf{p}^*$ after all sellers complete exploration. For each investigation of (a)–(c), we first establish a recursive inequality describing how the estimators or prices vary after one period; see (42), (43), and (15) for investigations (a), (b), and (c), respectively. It is noteworthy that although seller i's estimators $(\hat{\alpha}_i, \hat{\beta}_i, \hat{\boldsymbol{\gamma}}_i) = (\hat{\alpha}_i^{(\tau_i+1)}, \hat{\beta}_i^{(\tau_i+1)}, \hat{\boldsymbol{\gamma}}_i^{(\tau_i+1)})$ are computed once at the end of the exploration phase rather than being updated in every period, they are computed in an iterative manner using an SGD approach (8)–(9) based on cumulative data from period 1 to period τ_i, which enables us to quantify the estimator variation each time a new period of observation is collected. By leveraging the recursive inequalities, we can quantify the estimation error $(\hat{\alpha}_i - \alpha_i, \hat{\beta}_i - \beta_i, \hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)$ after the exploration phase and the deviation $\mathbf{p}^{(T)} - \mathbf{p}^*$ of the eventual prices from the Nash equilibrium. A flowchart of the proof is given in Figure 2. Our analysis is applicable to analyzing other single- or multi-agent settings. Particularly, the result in (14) can be viewed as a sensitivity analysis that quantifies the impact of estimator error $\hat{\beta}_i - \beta_i$ on the deviation between limiting prices and Nash equilibrium.

Remark 5 (Inapplicability of a Myopic Pricing Approach). Our dynamic pricing policy adopts a structure that separates price experiments from price optimization. A natural question arises: is a greedy pricing-while-estimating policy applicable? Can a seller update individual parameter estimators in each period and subsequently compute an optimal price to offer based on their current estimates? Such a policy is also known as passive learning, myopic pricing, or certainty equivalent pricing, and it does not require sellers to know the selling horizon length T a priori. Despite the elegance of this policy structure, literature in both statistics and operations research has demonstrated that a myopic pricing policy does not converge to an optimal solution when the system involves two or more unknown parameters, as in our problem. For instance, Lai and Robbins (1982), den Boer and Zwart (2014) investigated a linear model with $N = 1$ (*i.e.*, the monopoly case) and proved that such a myopic policy does not necessarily converge to the optimal solution. Cooper et al. (2015) investigated the linear demand competition model with $N = 2$ (*i.e.*, the duopoly case) and demonstrated that a myopic pricing policy does not necessarily lead to Nash equilibrium or individual optimal prices; refer to their Section 4.3 for details. Since a myopic pricing policy lacks guarantees for converging to optimal prices, it is not likely to be implemented in practice.

4.1. Assumptions and Preliminaries

We make the following assumption on individual price sensitivity coefficient β_i $(i \in \mathcal{N})$.

Assumption 2 **(Appropriate Price Sensitivity)**. *There exists constant* $\kappa \geqslant 1$ *such that* $\beta_i \geqslant \kappa \sum_{j \in \mathcal{N}\setminus\{i\}} |\gamma_{ij}|$ *and* $\beta_i \geqslant \kappa \sum_{j \in \mathcal{N}\setminus\{i\}} |\gamma_{ji}|$ *for all* $i \in \mathcal{N}$.

A market that does not satisfy this assumption can yield impractical outcomes. For instance, consider the scenario where $\boldsymbol{\gamma}_i \geqslant 0$ and $\beta_i < \sum_{j \in \mathcal{N}\setminus\{i\}} |\gamma_{ij}| = \sum_{j \in \mathcal{N}\setminus\{i\}} \gamma_{ij}$ for all $i \in \mathcal{N}$. In this case, (1) implies that the demand of each seller i will tend toward infinity when all sellers simultaneously increase their prices to infinity, which is impractical. Consider the scenario where $\gamma_{ji} \geqslant 0$ for all

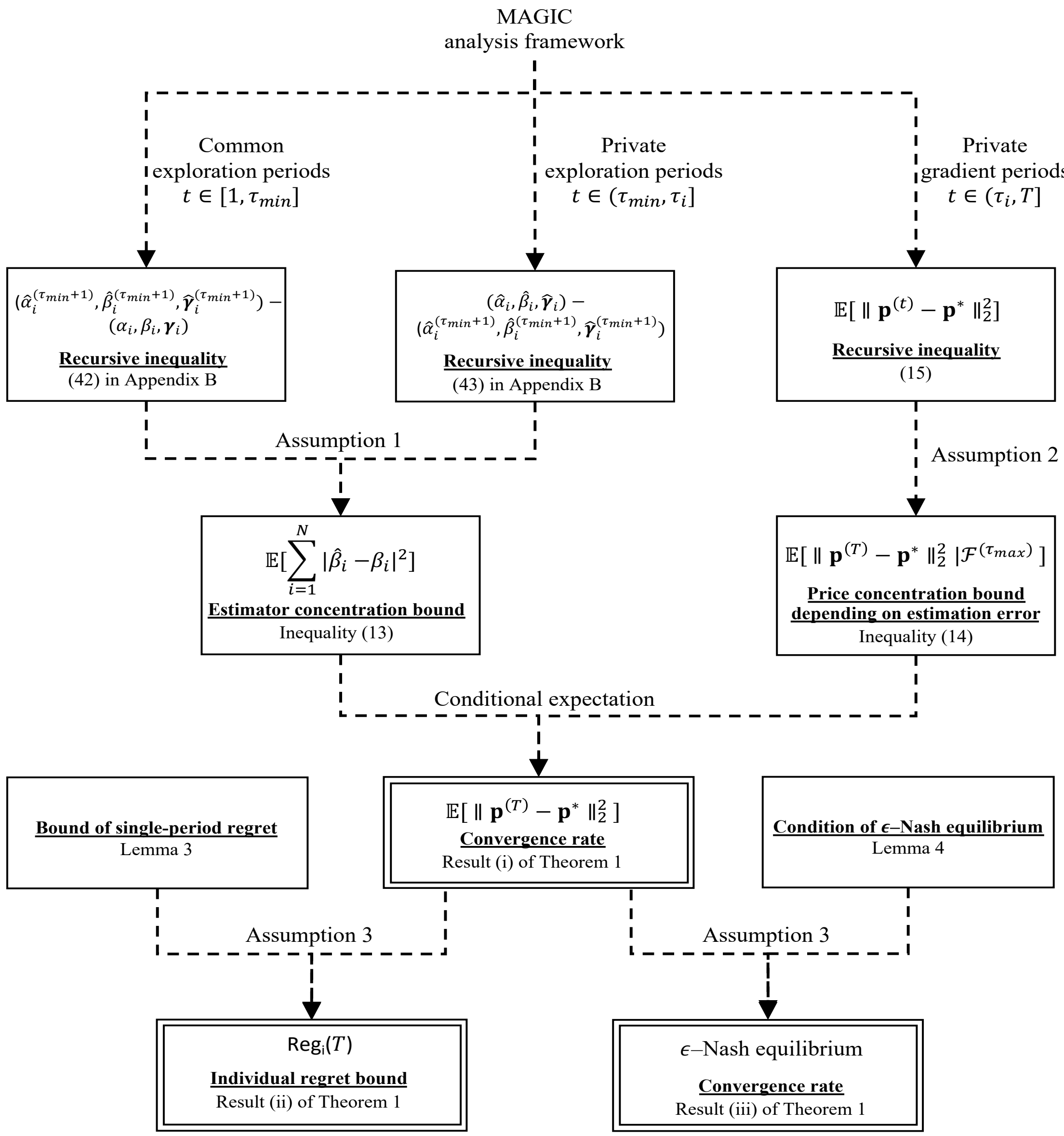


Figure 2 A flowchart of proving Theorems 1 for the LEGO policy's regret bounds and convergence rates using our MAGIC analysis framework. (A dashed arrow between A and B represents A indicating B. Double boxes highlight the main results. MAGIC means "Multi-Agent GradIent desCent". LEGO means "least-squares estimation then gradient optimization".)

$j \in \mathcal{N} \backslash \{i\}$ and $\beta_i < \sum_{j \in \mathcal{N} \backslash \{i\}} |\gamma_{ji}| = \sum_{j \in \mathcal{N} \backslash \{i\}} \gamma_{ji}$. In this case, (1) implies that the total demand of all sellers will tend toward infinity when seller i increases their price to infinity, which is impractical. Similar assumptions are found in Kachani et al. (2007). Assumption 2 ensures the existence of Nash equilibrium as in the following lemma:

LEMMA 1 **(Unique Nash Equilibrium Prices)**. *Under Assumption 2, there is a unique price vector* $\mathbf{p}^* \in \mathcal{P}$ *satisfying the balance equations* (4).

This result is based on the contraction mapping theorem. The proof is relegated to Appendix B.1.

To establish the optimal regret bound in result (ii) of Theorem 1, we assume that, given competitor prices, seller i's best response price is an interior point in $\mathcal{P}_i$:

Assumption 3 **(Sub-Optimal Boundary Prices)**. *A seller* $i \in \mathcal{N}$ *satisfies* (i) $\underline{p}_i \leqslant \max\big\{0, \frac{\alpha_i - \bar{\gamma}_i \max_{j\in\mathcal{N}\setminus\{i\}} \bar{p}_j}{2\beta_i}\big\}$ *and* (ii) $\bar{p}_i \geqslant \frac{\alpha_i + \bar{\gamma}_i \max_{j\in\mathcal{N}\setminus\{i\}} \bar{p}_j}{2\beta_i}$.

Here a sufficient condition for (i) is $\underline{p}_i = 0$, while a sufficient condition for (ii) is that for all $j \in \mathcal{N}$, $\bar{p}_j = \bar{p}$ with $\bar{p} \geqslant \frac{\alpha_j}{2\beta_j - \bar{\gamma}_j}$ and $\beta_j > \frac{\bar{\gamma}_j}{2}$. In a linear demand model, seller i's best response price is half of the choke price (also known as the null price, which reduces seller i's demand to 0). Assumption 3 essentially mandates that each seller can set prices higher than half of the choke price. This condition is weaker compared to the requirement found in the price competition literature, where sellers are assumed to be capable of offering choke prices; see, e.g., Gallego and Hu (2014).

The following lemmata provide useful bounds for proving our theorems.

Lemma 2 **(Bound of Feedback Moment)**. *For all* $t \geqslant \tau_i$, $t' < t$, *and* $i \in \mathcal{N}$, *we have that* $\mathbb{E}[(\phi_i^{(t)})^2 \,|\, \mathcal{F}^{(t')}] \leqslant U_1$. *Particularly,* $\mathbb{E}[(\phi_i^{(t)})^2] \leqslant U_1$.

Lemma 3 **(Bound of Single-Period Regret)**. *For all* $t \in \mathcal{T}$, *if seller* $i \in \mathcal{N}$ *satisfies Assumption 3, then* $\mathsf{Reg}_i(t) - \mathsf{Reg}_i(t-1) \leqslant 4\bar{\beta}_i \cdot \mathbb{E}\big[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2\big]$. *In other words, the regret of seller i increases no more than* $4\bar{\beta}_i \cdot \mathbb{E}\big[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2$ *in period t.*

Lemma 4 **(Sufficient Conditions of ϵ-Nash Equilibrium)**. *If each seller* $i \in \mathcal{N}$ *satisfies Assumption 3 and* $\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2 \leqslant \mu\epsilon$, $t \in \mathcal{T}$, *then* $\mathbf{p}^{(t)}$ *is a vector of ϵ-Nash equilibrium prices.*

The proof is relegated to Appendices B.2–B.4. Lemma 3 and 4 both leverage the properties that (i) the (single-period) average revenue of seller i is $2\beta_i$-smooth in individual price, and (ii) when seller i's price $p^{(t)}$ is the best response to competitor prices in period t, the gradient of average revenue is 0 and consequently, adjusting seller i's price to $p'^{(t)}$ reduces the revenue by $O\big((p'^{(t)} - p^{(t)})^2\big)$.

4.2. Proof of Main Theorem

We analyze the regret associated with the exploration phase and the gradient optimization phase, respectively.

Concentration bounds of estimators. Because the exploration phase does not adjust prices to increase revenue, the associated regret is $O(\tau_i)$ for each seller $i \in \mathcal{N}$. Nevertheless, seller i obtains an estimator of price coefficient β_i at the end of the exploration phase. The following inequality shows that the error of estimator $\hat{\beta}_i$ is $O(\tau_{\min}^{-1})$, or equivalently $O(T^{-\frac{1}{2}})$ as sellers apply order-of-$\sqrt{T}$ policies with $\tau_i \asymp \sqrt{T}$:

$$\mathbb{E}\big[\sum_{i\in\mathcal{N}} |\hat{\beta}_i - \beta_i|^2\big] \leqslant \frac{C_0 N}{\tau_{\min}} + \frac{U_\ell \sum_{i\in\mathcal{N}} \upsilon_i^2}{\tau_{\min}} \leqslant \frac{C_1}{\sqrt{T}}. \tag{13}$$

The proof is relegated to Appendix B.5. We interpret (13) as follows:

- Using the price and demand data collected up to the period $\tau_{\min}$ (*i.e.*, the end of *common* exploration periods), each seller $i \in \mathcal{N}$ has already improved their estimator $(\hat{\alpha}_i^{(\tau_{\min}+1)}, \hat{\beta}_i^{(\tau_{\min}+1)}, \hat{\boldsymbol{\gamma}}_i^{(\tau_{\min}+1)})$ to a good accuracy level with an average squared error less than $\frac{C_0 N}{\tau_{\min}}$. This part of error bound is established using the convexity in the least square estimation and the price variation provided by Assumption 1.
- The remaining data collected from period $\tau_{\min}$ to period τ_i (*i.e.*, the end of individual exploration phase) is also incorporated into seller i's estimation. Such remaining data cannot further improve seller i's estimator accuracy due to possible price collinearity as discussed in (*iv*) of Subsection 3.1. Nevertheless, their eventual estimator $(\hat{\alpha}_i, \hat{\beta}_i, \hat{\boldsymbol{\gamma}}_i) = (\hat{\alpha}_i^{(\tau_i+1)}, \hat{\beta}_i^{(\tau_i+1)}, \hat{\boldsymbol{\gamma}}_i^{(\tau_i+1)})$ does not lose a significant amount of accuracy compared with $(\hat{\alpha}_i^{(\tau_{\min}+1)}, \hat{\beta}_i^{(\tau_{\min}+1)}, \hat{\boldsymbol{\gamma}}_i^{(\tau_{\min}+1)})$; specifically, the average squared error increases no more than $\frac{U_\ell \sum_{i\in\mathcal{N}} v_i^2}{\tau_{\min}}$.

It is worth noting that it is hard for seller i to stop exploration exactly in period $\tau_{\min}$ or directly use $(\hat{\alpha}_i^{(\tau_{\min}+1)}, \hat{\beta}_i^{(\tau_{\min}+1)}, \hat{\boldsymbol{\gamma}}_i^{(\tau_{\min}+1)})$ instead of $(\hat{\alpha}_i^{(\tau_i+1)}, \hat{\beta}_i^{(\tau_i+1)}, \hat{\boldsymbol{\gamma}}_i^{(\tau_i+1)})$ as their eventual estimator $(\hat{\alpha}_i, \hat{\beta}_i, \hat{\boldsymbol{\gamma}}_i)$ for the gradient optimization phase. This is because exploration phase lengths are private and $\tau_{\min}$ is unknown to sellers.

Convergence in the gradient optimization phase. The regret associated with the gradient optimization phase depends on the error of estimators obtained from the exploration phase. Let us define random variable $\Delta := \sum_{i\in\mathcal{N}} |\hat{\beta}_i - \beta_i|^2$. Because all estimators $\{\hat{\beta}_i\}_{i\in\mathcal{N}}$ are already determined by the end of period $\tau_{\max} := \max_{i\in\mathcal{N}} \tau_i$, Δ is $\mathcal{F}^{(\tau_{\max})}$-measurable. We next establish the following inequality to quantify the convergence of prices in the gradient optimization phase:

$$\mathbb{E}\big[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2 \,|\, \mathcal{F}^{(\tau_{\max})}\big] \leqslant \frac{\tilde{b} + 2\omega D + \tau_{\max} D}{t} + \rho\Delta, \qquad t \in \{\tau_{\max}+1, \tau_{\max}+2, \ldots, T\}. \tag{14}$$

This inequality indicates that given the estimator error Δ, the joint price vector $\mathbf{p}^{(t)}$ in the gradient optimization phase eventually converges to a ball of radius $\rho\Delta$ centered at the equilibrium prices $\mathbf{p}^*$ if the gradient optimization phase is sufficiently long (*i.e.*, $t \to \infty$). A longer exploration phase reduces the error Δ, thereby diminishing the distance between the eventual joint prices $\mathbf{p}^{(T)}$ and the Nash equilibrium prices $\mathbf{p}^*$.

Our proof of (14) uses the following recursive inequality (15). For notation simplicity, we write the conditional expectation $\tilde{\mathbb{E}}[\cdot] := \mathbb{E}[\cdot \,|\, \mathcal{F}^{(\tau_{\max})}]$. For $t \in \{\tau_{\max}+1, \tau_{\max}+2, \ldots, T-1\}$, we have

$$\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t+1)} - \mathbf{p}^*\|_2^2 - \frac{\tilde{b}}{t+1}\big] \leqslant (1 - \frac{\omega}{t}) \cdot \tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2 - \frac{\tilde{b}}{t}\big] + \frac{\rho\Delta}{t}. \tag{15}$$

The proof of (15) is relegated to Appendix B.6. Define $T^* := \max\{\lceil \omega \rceil, \tau_{\max}+1\} \geqslant 2$. Then for $t \in \{T^*, T^*+1, \ldots, T-1\}$, we have that

$$\begin{aligned}
\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t+1)} - \mathbf{p}^*\|_2^2\big] - \frac{\tilde{b}}{t+1} \leqslant & \Big\{\prod_{u=T^*}^{t}(1-\frac{\omega}{u})\Big\} \cdot \tilde{\mathbb{E}}\big[\|\mathbf{p}^{(T^*)} - \mathbf{p}^*\|_2^2 - \frac{\tilde{b}}{T^*}\big] + \sum_{u=T^*}^{t}\frac{\rho\Delta}{u}\prod_{r=u+1}^{t}(1-\frac{\omega}{r}) \\
& (\text{by substituting (15) up to } T^*. \text{ Note } T^* \geqslant \omega,\ 1-\frac{\omega}{u} \geqslant 0 \text{ for } u \geqslant T^*) \\
\leqslant & \Big\{\prod_{u=T^*}^{t}(1-\frac{\omega}{u})\Big\} \cdot \tilde{\mathbb{E}}\big[\|\mathbf{p}^{(T^*)} - \mathbf{p}^*\|_2^2\big] + \sum_{u=T^*}^{t}\frac{\rho\Delta}{u}\prod_{r=u+1}^{t}(1-\frac{\omega}{r}) \\
\leqslant & \Big\{\prod_{u=T^*}^{t}(1-\frac{1}{u})\Big\} \cdot D + \sum_{u=T^*}^{t}\frac{\rho\Delta}{u}\prod_{r=u+1}^{t}(1-\frac{1}{r}) \\
& (\text{due to } \omega > 1 \text{ and } D = \sum_{i\in\mathcal{N}}(\bar{p}_i - \underline{p}_i)^2) \\
\leqslant & \frac{T^*-1}{t}D + \sum_{u=T^*}^{t}\frac{\rho\Delta}{u}\cdot\frac{u}{t} \leqslant \frac{T^*D}{t+1} + \rho\Delta.
\end{aligned}$$

This suggests $\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2\big] \leqslant \frac{\tilde{b}}{t} + \frac{T^*D}{t} + \rho\Delta = \frac{\tilde{b}+T^*D}{t} + \rho\Delta$ for $t \in \{T^*+1, T^*+2, \ldots, T\}$. For $t \geqslant \tau_{\max}+1$ such that $t \leqslant T^*$, we also have $\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2\big] \leqslant D \leqslant \frac{T^*D}{t} \leqslant \frac{\tilde{b}+T^*D}{t} + \rho\Delta$. Thus, for $t \in \{\tau_{\max}+1, \tau_{\max}+2, \ldots, T\}$,

$$\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2\big] \leqslant \frac{\tilde{b}+T^*D}{t} + \rho\Delta \leqslant \frac{\tilde{b}+D(2\omega+\tau_{\max})}{t} + \rho\Delta = \frac{\tilde{b}+2\omega D+\tau_{\max}D}{t} + \rho\Delta.$$

The second inequality is due to $T^* = \max\{\lceil\omega\rceil, \tau_{\max}+1\} \leqslant \omega + \tau_{\max} + 1 \leqslant 2\omega + \tau_{\max}$. By definition of $\tilde{\mathbb{E}}$, $\mathbb{E}\big[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2 \,|\, \mathcal{F}^{(\tau_{\max})}\big] \leqslant \frac{\tilde{b}+2\omega D+\tau_{\max}D}{t} + \rho\Delta$ for $t \in \{\tau_{\max}+1, \tau_{\max}+2, \ldots, T\}$, which is (14).

Regret and convergence to equilibrium. Taking expectation $\mathbb{E}[\cdot]$ for both sides of (14) with $t > \tau_{\max}$ and noticing $\tau_{\max} \leqslant \bar{\iota}\sqrt{T}$, we have that

$$\begin{aligned}
\mathbb{E}[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2] \leqslant & \frac{\tilde{b}+2\omega D+\bar{\iota}D\sqrt{T}}{t} + \rho\mathbb{E}[\Delta] = \frac{\tilde{b}+2\omega D+\bar{\iota}D\sqrt{T}}{t} + \rho\mathbb{E}[\sum_{i\in\mathcal{N}}|\hat{\beta}_i - \beta_i|^2] \\
\leqslant & \frac{\tilde{b}+2\omega D+\bar{\iota}D\sqrt{T}}{t} + \rho\frac{C_1}{\sqrt{T}}, \qquad t > \tau_{\max}.
\end{aligned} \tag{16}$$

The equality is due to definition $\Delta = \sum_{i\in\mathcal{N}}|\hat{\beta}_i - \beta_i|^2$ and the last row is due to $\mathbb{E}\big[\sum_{i\in\mathcal{N}}|\hat{\beta}_i - \beta_i|^2\big] \leqslant \frac{C_1}{\sqrt{T}}$ in (13). If $T > \tau_{\max}$, plugging $t = T$ into (16) yields $\mathbb{E}[\|\mathbf{p}^{(T)} - \mathbf{p}^*\|_2^2] \leqslant \frac{\tilde{b}+2\omega D}{T} + \frac{\bar{\iota}D+\rho C_1}{\sqrt{T}}$. Otherwise, $T = \tau_{\max} \leqslant \bar{\iota}\sqrt{T}$ and $\sqrt{T} \leqslant \bar{\iota}$, which also yield $\mathbb{E}[\|\mathbf{p}^{(T)} - \mathbf{p}^*\|_2^2] \leqslant D \leqslant \frac{\bar{\iota}D}{\sqrt{T}} \leqslant \frac{\tilde{b}+2\omega D}{T} + \frac{\bar{\iota}D+\rho C_1}{\sqrt{T}}$. Because $\tilde{b}, D \in O(N)$ and $C_1 \in O(N^2)$, we have $\mathbb{E}[\|\mathbf{p}^{(T)} - \mathbf{p}^*\|_2^2] \in O(\frac{N^2}{\sqrt{T}})$. This is (i) of Theorem 1.

With the convergence rate (16), Lemma 3 implies the regret bound in (ii) of Theorem 1:

$$\mathsf{Reg}_i(T) \leqslant 4\bar{\beta}_i \cdot \sum_{t=1}^{T}\mathbb{E}\big[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2\big] = 4\bar{\beta}_i \cdot \sum_{t=1}^{\tau_{\max}}\mathbb{E}\big[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2\big] + 4\bar{\beta}_i \cdot \sum_{t=\tau_{\max}+1}^{T}\mathbb{E}\big[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2\big]$$

$$
\begin{aligned}
&\leqslant 4\bar{\beta}_i D\tau_{\max} + 4\bar{\beta}_i \cdot \sum_{t=\tau_{\max}+1}^{T} \Big(\frac{\tilde{b} + 2\omega D + \bar{\iota} D\sqrt{T}}{t} + \rho \frac{C_1}{\sqrt{T}} \Big) \qquad \text{(due to (16))} \\
&\leqslant 4\bar{\beta}_i D\bar{\iota}\sqrt{T} + 4\bar{\beta}_i \cdot \sum_{t=2}^{T} \Big(\frac{\tilde{b} + 2\omega D + \bar{\iota} D\sqrt{T}}{t} + \rho \frac{C_1}{\sqrt{T}} \Big) \\
&\leqslant 4\bar{\beta}_i D\bar{\iota}\sqrt{T} + 4\bar{\beta}_i \cdot \Big((\tilde{b} + 2\omega D + \bar{\iota} D\sqrt{T}) \log T + \rho C_1 \sqrt{T} \Big) \\
&= 4\bar{\beta}_i (\tilde{b} + 2\omega D) \log T + (4\bar{\beta}_i \bar{\iota} D + 4\bar{\beta}_i \rho C_1 + 4\bar{\beta}_i \bar{\iota} D \log T)\sqrt{T} \in O\big(N\sqrt{T}(N + \log T)\big).
\end{aligned}
$$

The order in the last row is due to $\tilde{b}, D \in O(N)$ and $C_1 \in O(N^2)$.

Consider $\epsilon > 0$. By the Markov inequality,

$$
\mathbb{P}\big[\|\mathbf{p}^{(T)} - \mathbf{p}^*\|_2^2 > \mu\epsilon\big] \leqslant \frac{\mathbb{E}[\|\mathbf{p}^{(T)} - \mathbf{p}^*\|_2^2]}{\mu\epsilon} \leqslant \frac{\tilde{b} + 2\omega D}{\mu\epsilon T} + \frac{\bar{\iota} D + \rho C_1}{\mu\epsilon\sqrt{T}}.
$$

The second inequality is due to (i) of Theorem 1. Thus, with probability $1 - \frac{\tilde{b}+2\omega D}{\mu\epsilon T} - \frac{\bar{\iota} D + \rho C_1}{\mu\epsilon\sqrt{T}}$, $\|\mathbf{p}^{(T)} - \mathbf{p}^*\|_2^2 \leqslant \mu\epsilon$ and $\mathbf{p}^{(T)}$ is a vector of ϵ-Nash equilibrium prices by Lemma 4. Because $\tilde{b}, D \in O(N)$ and $C_1 \in O(N^2)$, we have $1 - \frac{\tilde{b}+2\omega D}{\mu\epsilon T} - \frac{\bar{\iota} D + \rho C_1}{\mu\epsilon\sqrt{T}} = 1 - O(\frac{N^2}{\epsilon\sqrt{T}})$. This is (iii) of Theorem 1.

5. Improved Regret of Phased LEGO with Known Individual Price Sensitivity

In the general case studied in Sections 2–4, a seller i's unknown demand $y_i^{(t)}$ comprises two components, as shown in (1): the self-influenced demand $-\beta_i p_i^{(t)}$ and the potential demand size $\alpha_i + \boldsymbol{\gamma}_i^\top \mathbf{p}_{-i}^{(t)} + \varepsilon_i^{(t)}$, where the latter component depends on competitor prices when $N > 1$. These two demand components also constitute two significant obstacles to dynamic pricing in sequential competition, which force a policy to have an order-of-$\sqrt{T}$ worst-case regret (including the monopolist setting $N = 1$). The first obstacle is that a seller is uncertain regarding how their price affects the demand: they do not know what price maximizes the revenue even if the potential demand size $\alpha_i + \boldsymbol{\gamma}_i^\top \mathbf{p}_{-i}^{(t)} + \varepsilon_i^{(t)}$ is known. The second and seemingly more significant obstacle is that a seller is uncertain regarding competitor price patterns or how competitor prices affect the demand. This obstacle is compounded by the presence of demand noise: sellers do not directly observe the average noise-free demand at specific prices; instead, they only observe a random variable whose mean value corresponds to the average demand. Consequently, these two obstacles render prices uninformative, as no pricing policy can effectively reduce demand uncertainty.

Given this observation, a natural question arises: how much does each of the two obstacles contribute to the difficulty of dynamic pricing in sequential competition? Which obstacle forces the order-of-$\sqrt{T}$ worst-case regret? The following theorem addresses this issue by considering a condition in which each seller knows their individual price sensitivity coefficient. Under this assumption, we demonstrate that a gradient optimization policy achieves an improved regret of $O(\log T)$.

THEOREM 2 **(Optimal Policy with Known Individual Price Sensitivity)**. *Under Assumption 2, suppose that each seller $i \in \mathcal{N}$ knows individual price sensitivity β_i and implements a modified Algorithm 1:*

(A) $\tau_i = 1$, $\tilde{p}_i^{(1)}$ is arbitrary in $\mathcal{P}_i$;

(B) for $t \in \{1, 2, \ldots, T-1\}$, the step size $\eta_i^{(t)} = \frac{\zeta_i}{t}$, where $\nu := \frac{\max_{j\in\mathcal{N}} \zeta_j}{\min_{j\in\mathcal{N}} \zeta_j} < 4\kappa - 1$ and $\zeta_i > \frac{\kappa}{(4\kappa-1-\nu)\min_{j\in\mathcal{N}} \underline{\beta}_j}$; and

(C) the estimation step (7) is replaced by $\hat{\beta}_i = \beta_i$.

Then we have

(i) convergence to Nash equilibrium prices:

$$\mathbb{E}\big[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2\big] \leqslant \frac{\delta}{t} \in O\Big(\frac{N}{t}\Big), \qquad t \in \mathcal{T};$$

(ii) individual sublinear regret:

$$\mathsf{Reg}_i(T) \leqslant 4\bar{\beta}_i \delta \log T + 4\bar{\beta}_i \delta \in O\big(N \log T\big)$$

for each seller $i \in \mathcal{N}$ satisfying Assumption 3 and $t \in \mathcal{T}$; and

(iii) $\mathbf{p}^{(t)}$ is a vector of ϵ-Nash equilibrium prices with (at least) probability $1 - \frac{\delta}{\mu\epsilon t} = 1 - O\Big(\frac{N}{\epsilon t}\Big)$ for all $t \in \mathcal{T}$ if each seller $i \in \mathcal{N}$ satisfies Assumption 3.

The associated constants are defined as

$$b := \frac{U_1 \sum_{i\in\mathcal{N}} \zeta_i^2}{2\omega - 1}, \qquad \delta := b + (2\omega + 1)D. \tag{17}$$

Because $b, D \in O(N)$, we have $\delta \in O(N)$.

Comparing the order-of-$\log T$ regret in Theorem 2 with the order-of-$\sqrt{T}$ regret in Theorem 1 (which matches the problem lower bound as discussed in Remark 1), we can conclude that *the unknown individual price sensitivity β_i contributes to the major difficulty of dynamic pricing in sequential competition* and forces regret to the order of $\sqrt{T}$ in the worst case.

REMARK 6 (POLICY OPTIMALITY UNDER KNOWN INDIVIDUAL PRICE SENSITIVITY). We delve into policy optimality from two perspectives. Firstly, for each seller $i \in \mathcal{N}$, their revenue maximization problem with known individual price sensitivity β_i can be viewed as strongly concave optimization. Particularly, with β_i known, (noisy) gradient feedback becomes available, as we discussed when defining (10). Then the $O(\frac{1}{t})$ convergence rate in result (i) of Theorem 2 aligns with the $\Omega(\frac{1}{t})$ lower bound established in Nemirovskij and Yudin (1983) for strongly concave optimization. Secondly, seller i's revenue maximization is framed as a regret minimization problem within this work. For the special case of a single seller with known price sensitivity, Broder and Rusmevichientong (2012), Keskin and Zeevi (2014) have established a lower regret bound of $\Omega(\log T)$.

This lower bound is matched by our $O(\log T)$ regret bound in result (ii) of Theorem 2. Here we adopt a stronger dynamic benchmark to define regret compared with Broder and Rusmevichientong (2012), Keskin and Zeevi (2014) using a static benchmark. Our work achieves the optimal regret under a dynamic benchmark because of the specific problem structure and the assumption that sellers employ the same class of online learning policies.

Proof of Theorem 2: For all $t \in \{2,3,\dots,T-1\}$, we have that

$$\begin{aligned}
\|\mathbf{p}^{(t+1)} - \mathbf{p}^*\|_2^2 &= \sum_{i\in\mathcal{N}} (p_i^{(t+1)} - p_i^*)^2 = \sum_{i\in\mathcal{N}} \left[\Pi_{\mathcal{P}_i}\left(p_i^{(t)} + \eta_i^{(t)}\phi_i^{(t)}\right) - p_i^*\right]^2 \quad \text{(due to (11))}\\
&\leqslant \sum_{i\in\mathcal{N}} \left(p_i^{(t)} + \eta_i^{(t)}\phi_i^{(t)} - p_i^*\right)^2 \quad \text{(due to definition of projection } \Pi_{\mathcal{P}_i}(\cdot))\\
&= \sum_{i\in\mathcal{N}} \left[\left(p_i^{(t)} - p_i^*\right)^2 + \left(\eta_i^{(t)}\phi_i^{(t)}\right)^2 + 2\eta_i^{(t)}\phi_i^{(t)}\cdot(p_i^{(t)} - p_i^*)\right]\\
&= \|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2 + \frac{\sum_{i\in\mathcal{N}}\zeta_i^2(\phi_i^{(t)})^2}{t^2} + \frac{2}{t}\sum_{i\in\mathcal{N}}\zeta_i\phi_i^{(t)}\cdot(p_i^{(t)} - p_i^*).
\end{aligned} \tag{18}$$

Here the last row is due to $\eta_i^{(t)} = \frac{\zeta_i}{t}$. Taking expectations for both sides yields that

$$\mathbb{E}\left[\|\mathbf{p}^{(t+1)} - \mathbf{p}^*\|_2^2\right] \leqslant \mathbb{E}\left[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2\right] + \frac{U_1\sum_{i\in\mathcal{N}}\zeta_i^2}{t^2} + \frac{2}{t}\mathbb{E}\Big[\sum_{i\in\mathcal{N}}\zeta_i\phi_i^{(t)}\cdot(p_i^{(t)} - p_i^*)\Big]. \tag{19}$$

The second term on the right-hand side of (19) is due to Lemma 2. We next show the last term in (19) satisfies

$$\mathbb{E}\Big[\sum_{i\in\mathcal{N}}\zeta_i\phi_i^{(t)}\cdot(p_i^{(t)} - p_i^*)\Big] \leqslant -\omega\cdot\mathbb{E}\left[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2\right]. \tag{20}$$

We recall that $\mathbf{p}^{(t)}$ is $\mathcal{F}^{(t-1)}$-measurable. To prove (20), we only need to show that

$$\mathbb{E}\Big[\sum_{i\in\mathcal{N}}\zeta_i\phi_i^{(t)}\cdot(p_i^{(t)} - p_i^*)\,\Big|\,\mathcal{F}^{(t-1)}\Big] \leqslant -\omega\cdot\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2. \tag{21}$$

This is because taking expectation $\mathbb{E}[\cdot] = \mathbb{E}[\cdot\,|\,\mathcal{F}^{(0)}]$ for both sides of (21) yields (20). Recall (I) $\phi_i^{(t)} = y_i^{(t)} - \hat{\beta}_i p_i^{(t)} = y_i^{(t)} - \beta_i p_i^{(t)}$ by (10) and (C) of Theorem 2, (II) $\mathbf{p}^{(t)}$ is $\mathcal{F}^{(t-1)}$-measurable, and (III) $\mathbb{E}[y_i^{(t)}\,|\,\mathcal{F}^{(t-1)}] = \alpha_i - \beta_i p_i^{(t)} + \boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^{(t)}$ due to (1). Because of (I)–(III), (21) is equivalent to

$$\sum_{i\in\mathcal{N}}\zeta_i\left(\alpha_i - 2\beta_i p_i^{(t)} + \boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^{(t)}\right)(p_i^{(t)} - p_i^*) \leqslant -\omega\cdot\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2. \tag{22}$$

Define the following vector function $\boldsymbol{\Psi}:\mathbb{R}^N \mapsto \mathbb{R}^N$ with variable $\mathbf{p} = (p_i)_{i\in\mathcal{N}} \in \mathcal{P}$.

$$\boldsymbol{\Psi}(\mathbf{p}) = \left(\Psi_i(\mathbf{p})\right)_{i\in\mathcal{N}} := \Big(\zeta_i\cdot\left(\alpha_i - 2\beta_i p_i + \boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}\right)\Big)_{i\in\mathcal{N}}.$$

Then (22) can be written as

$$\boldsymbol{\Psi}(\mathbf{p}^{(t)})^\top(\mathbf{p}^{(t)} - \mathbf{p}^*) \leqslant -\omega\cdot\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2. \tag{23}$$

We also have

$$\Psi_i(\mathbf{p}^*) \cdot (p_i^{(t)} - p_i^*) \leqslant 0, \qquad i \in \mathcal{N}. \tag{24}$$

This is because $\frac{\Psi_i(\mathbf{p}^*)}{\zeta_i} = \alpha_i - 2\beta_i p_i^* + \boldsymbol{\gamma}_i^\top \mathbf{p}_{-i}^*$ is the average revenue gradient of seller i at Nash equilibrium prices $\mathbf{p}^*$, as shown in (6). This implies (24): (I) if $p_i^* \in (\underline{p}_i, \bar{p}_i)$, then $\Psi_i(\mathbf{p}^*) = 0$ and $\Psi_i(\mathbf{p}^*) \cdot (p_i^{(t)} - p_i^*) \leqslant 0$; (II) if $p_i^* = \bar{p}_i$, then $\Psi_i(\mathbf{p}^*) \geqslant 0$, $p_i^{(t)} - p_i^* \leqslant 0$, and $\Psi_i(\mathbf{p}^*) \cdot (p_i^{(t)} - p_i^*) \leqslant 0$; and (III) if $p_i^* = \underline{p}_i$, then $\Psi_i(\mathbf{p}^*) \leqslant 0$, $p_i^{(t)} - p_i^* \geqslant 0$, and $\Psi_i(\mathbf{p}^*) \cdot (p_i^{(t)} - p_i^*) \leqslant 0$. Summing (24) over $i \in \mathcal{N}$ yields $\boldsymbol{\Psi}(\mathbf{p}^*)^\top (\mathbf{p}^{(t)} - \mathbf{p}^*) \leqslant 0$, which implies that, to show (23), we only need to prove $\big(\boldsymbol{\Psi}(\mathbf{p}^{(t)}) - \boldsymbol{\Psi}(\mathbf{p}^*)\big)^\top (\mathbf{p}^{(t)} - \mathbf{p}^*) \leqslant -\omega \cdot \|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2$, or equivalently,

$$\big(\boldsymbol{\Psi}(\mathbf{p}^{(t)}) + \omega \mathbf{p}^{(t)} - \boldsymbol{\Psi}(\mathbf{p}^*) - \omega \mathbf{p}^*\big)^\top (\mathbf{p}^{(t)} - \mathbf{p}^*) \leqslant 0. \tag{25}$$

Taking the derivative of Ψ_i in p_i yields that $\frac{\partial \Psi_i(\mathbf{p})}{\partial p_i} = -2\zeta_i\beta_i$. Taking the derivative of Ψ_i in p_j $(j \neq i)$ yields that $\frac{\partial \Psi_i(\mathbf{p})}{\partial p_j} = \zeta_i \gamma_{ij}$. Then we have

$$\Big|\frac{\partial \Psi_i(\mathbf{p})}{\partial p_i}\Big| - \sum_{j \in \mathcal{N}\setminus\{i\}} \Big|\frac{\partial \Psi_i(\mathbf{p})}{\partial p_j}\Big| = \zeta_i(2\beta_i - \sum_{j \in \mathcal{N}\setminus\{i\}} |\gamma_{ij}|) \geqslant \zeta_i\Big(2\beta_i - \frac{\beta_i}{\kappa}\Big) \geqslant \zeta_i\beta_i\Big(2 - \frac{1}{\kappa}\Big), \tag{26}$$

where the first inequality is due to $\beta_i \geqslant \kappa \sum_{j \in \mathcal{N}\setminus\{i\}} |\gamma_{ij}|$ in Assumption 2, and

$$\Big|\frac{\partial \Psi_i(\mathbf{p})}{\partial p_i}\Big| - \sum_{j \in \mathcal{N}\setminus\{i\}} \Big|\frac{\partial \Psi_j(\mathbf{p})}{\partial p_i}\Big| = 2\zeta_i\beta_i - \sum_{j \in \mathcal{N}\setminus\{i\}} \zeta_j |\gamma_{ji}| \geqslant 2\zeta_i\beta_i - \frac{\nu\zeta_i\beta_i}{\kappa} \geqslant \zeta_i\beta_i\Big(2 - \frac{\nu}{\kappa}\Big), \tag{27}$$

where the first inequality is due to $\nu = \frac{\max_{i \in \mathcal{N}} \zeta_i}{\min_{i \in \mathcal{N}} \zeta_i}$ in (B) of Theorem 2 and $\beta_i \geqslant \kappa \sum_{j \in \mathcal{N}\setminus\{i\}} |\gamma_{ji}|$ in Assumption 2. Combining (26) and (27) yields that $\frac{\partial[\boldsymbol{\Psi}(\mathbf{p}) + \boldsymbol{\Psi}^\top(\mathbf{p})]}{\partial \mathbf{p}}$ is diagonal dominant: in row i, the magnitude of the diagonal entry is larger than the sum of the magnitudes of all the other entries in that row by $\zeta_i\beta_i \cdot \frac{4\kappa - 1 - \nu}{\kappa} \geqslant 2\omega > 0$, where the inequalities are due to (12) and (B) of Theorem 2. Thus, we have $\frac{\partial\left[\boldsymbol{\Psi}(\mathbf{p}) + \boldsymbol{\Psi}^\top(\mathbf{p})\right]}{\partial \mathbf{p}} \preccurlyeq -2\omega \cdot \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, and

$$\frac{\partial \frac{\boldsymbol{\Psi}(\mathbf{p}) + \boldsymbol{\Psi}^\top(\mathbf{p})}{2}}{\partial \mathbf{p}} \preccurlyeq -\omega \cdot \mathbf{I}. \tag{28}$$

Define $f(z) = \Big(\boldsymbol{\Psi}\big(z\mathbf{p}^{(t)} + (1-z)\mathbf{p}^*\big) + z\omega\mathbf{p}^{(t)} - \boldsymbol{\Psi}(\mathbf{p}^*) - z\omega\mathbf{p}^*\Big)^\top (\mathbf{p}^{(t)} - \mathbf{p}^*)$. Then we have

$$\begin{aligned}
&\big(\boldsymbol{\Psi}(\mathbf{p}^{(t)}) + \omega\mathbf{p}^{(t)} - \boldsymbol{\Psi}(\mathbf{p}^*) - \omega\mathbf{p}^*\big)^\top (\mathbf{p}^{(t)} - \mathbf{p}^*) = f(1) - f(0) = f'(z') \quad (z' \text{ is some element in } [0,1]) \\
=&(\mathbf{p}^{(t)} - \mathbf{p}^*)^\top \frac{\partial \boldsymbol{\Psi}(\mathbf{p})}{\partial \mathbf{p}}\Big|_{\mathbf{p} = z'\mathbf{p}^{(t)} + (1-z')\mathbf{p}^*} (\mathbf{p}^{(t)} - \mathbf{p}^*) + \omega \cdot (\mathbf{p}^{(t)} - \mathbf{p}^*)^\top (\mathbf{p}^{(t)} - \mathbf{p}^*) \\
=&(\mathbf{p}^{(t)} - \mathbf{p}^*)^\top \frac{\partial \frac{\boldsymbol{\Psi}(\mathbf{p}) + \boldsymbol{\Psi}^\top(\mathbf{p})}{2}}{\partial \mathbf{p}}\Big|_{\mathbf{p} = z'\mathbf{p}^{(t)} + (1-z')\mathbf{p}^*} (\mathbf{p}^{(t)} - \mathbf{p}^*) + \omega \cdot (\mathbf{p}^{(t)} - \mathbf{p}^*)^\top (\mathbf{p}^{(t)} - \mathbf{p}^*) \\
\leqslant& -(\mathbf{p}^{(t)} - \mathbf{p}^*)^\top \omega \cdot \mathbf{I}(\mathbf{p}^{(t)} - \mathbf{p}^*) + \omega \cdot (\mathbf{p}^{(t)} - \mathbf{p}^*)^\top (\mathbf{p}^{(t)} - \mathbf{p}^*) = 0. \quad (\text{due to } (28))
\end{aligned}$$

Recall (12) and (B) of Theorem 2, which imply $\zeta_i > \frac{\kappa}{(4\kappa-1-\nu)\min_{j\in\mathcal{N}}\underline{\beta}_j} = \frac{\min_{j\in\mathcal{N}}\zeta_j}{2\omega}$ for all $i\in\mathcal{N}$. Thus, $2\omega>1$ and $b>0$ by definition. Plugging (20) into (19) yields that

$$\begin{aligned}
\mathbb{E}\big[\|\mathbf{p}^{(t+1)}-\mathbf{p}^*\|_2^2\big] &\leqslant \Big(1-\frac{2\omega}{t}\Big)\mathbb{E}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big] + \frac{U_1\sum_{i\in\mathcal{N}}\zeta_i^2}{t^2}\\
&= \Big(1-\frac{2\omega}{t}\Big)\mathbb{E}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big] + \frac{b(2\omega-1)}{t^2}\\
&\leqslant \Big(1-\frac{2\omega}{t}\Big)\mathbb{E}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big] + \frac{2\omega b}{t^2} - \frac{b}{t(t+1)}\\
&= \Big(1-\frac{2\omega}{t}\Big)\mathbb{E}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big] + \frac{2\omega b}{t^2} - \frac{b}{t} + \frac{b}{t+1}.
\end{aligned}$$

This can be written as $\mathbb{E}\big[\|\mathbf{p}^{(t+1)}-\mathbf{p}^*\|_2^2-\frac{b}{t+1}\big] \leqslant (1-\frac{2\omega}{t})\cdot\mathbb{E}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2-\frac{b}{t}\big]$ for $t\in\{2,3,\ldots,T-1\}$. Therefore, for $t\in\{T^*,T^*+1,\ldots,T-1\}$ where $T^*:=\lceil 2\omega\rceil\geqslant 2$, we have that

$$\begin{aligned}
\mathbb{E}\big[\|\mathbf{p}^{(t+1)}-\mathbf{p}^*\|_2^2\big]-\frac{b}{t+1} &\leqslant \Big\{\prod_{u=T^*}^{t}(1-\frac{2\omega}{u})\Big\}\cdot\mathbb{E}\big[\|\mathbf{p}^{(T^*)}-\mathbf{p}^*\|_2^2-\frac{b}{T^*}\big] \quad (\text{due to } T^*=\lceil 2\omega\rceil\geqslant 2\omega)\\
&\leqslant \Big\{\prod_{u=T^*}^{t}(1-\frac{2\omega}{u})\Big\}\cdot\mathbb{E}\big[\|\mathbf{p}^{(T^*)}-\mathbf{p}^*\|_2^2\big] \leqslant \Big\{\prod_{u=T^*}^{t}(1-\frac{2\omega}{u})\Big\}\cdot D\\
&\leqslant \Big\{\prod_{u=T^*}^{t}(1-\frac{1}{u})\Big\}\cdot D = \frac{T^*-1}{t}D \leqslant \frac{T^*D}{t+1}. \qquad (\text{due to } 2\omega>1)
\end{aligned}$$

This suggests $\mathbb{E}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big]\leqslant \frac{b}{t}+\frac{T^*D}{t}=\frac{b+T^*D}{t}$ for $t\in\{T^*+1,T^*+2,\ldots,T\}$. For $t\leqslant T^*$, we also have $\mathbb{E}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big]\leqslant D\leqslant\frac{T^*D}{t}\leqslant\frac{b+T^*D}{t}$. Thus, for $t\in\mathcal{T}$, $\mathbb{E}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big]\leqslant\frac{b+T^*D}{t}\leqslant\frac{b+(2\omega+1)D}{t}=\frac{\delta}{t}$ (due to $T^*=\lceil 2\omega\rceil\leqslant 2\omega+1$). Because $\delta\in O(N)$, we have $\mathbb{E}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big]\in O(\frac{N}{t})$. This completes the proof of (i) of Theorem 2.

With the above convergence rate, Lemma 3 implies the regret bound in (ii) of Theorem 2:

$$\mathsf{Reg}_i(T)\leqslant 4\bar{\beta}_i\cdot\sum_{t=1}^{T}\mathbb{E}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big]\leqslant 4\bar{\beta}_i\cdot\sum_{t=1}^{T}\frac{\delta}{t}\leqslant 4\bar{\beta}_i\delta(\log T+1)=4\bar{\beta}_i\delta\log T+4\bar{\beta}_i\delta\in O(N\log T).$$

The order in the last step is due to $\delta\in O(N)$.

Consider $\epsilon>0$. By the Markov inequality, $\mathbb{P}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2>\mu\epsilon\big]\leqslant\frac{\mathbb{E}[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2]}{\mu\epsilon}\leqslant\frac{\delta}{\mu\epsilon t}$. With probability $1-\frac{\delta}{\mu\epsilon t}$, we have $\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\leqslant\mu\epsilon$ and $\mathbf{p}^{(t)}$ is a vector of ϵ-Nash equilibrium prices by Lemma 4. Because $\delta\in O(N)$, $1-\frac{\delta}{\mu\epsilon t}=1-O\Big(\frac{N}{\epsilon t}\Big)$. This completes the proof of (iii) of Theorem 2. □

6. Numerical Experiments

We test the performance of our LEGO policy (Algorithm 1) in the sequential price competition with $N\in\{2,5,10\}$ sellers. For each seller $i\in\mathcal{N}$, their price is supported on $[0,1]$. Their price sensitivity β_i is sampled from $[10,12]$. The other price coefficient γ_{ij} $(j\neq i)$ is sampled from $[0,1]$ subject to the condition that $\sum_{j\in\mathcal{N}\setminus\{i\}}\gamma_{ij}\leqslant 3$. Their potential market size parameter α_i is sampled from $[13,17]$. The demand noise follows a uniform distribution on $[-1,1]$. When evaluating a policy, we independently repeat trials 800 times to obtain average performance.

6.1. Performance of LEGO Policy

Each seller applies our LEGO policy. Seller $i \in \mathcal{N}$ has an exploration phase of private length $\tau_i = \iota_i \sqrt{T}$, where ι_i is sampled from $[1,5]$, and uses a private step size $\eta_i^{(t)} = \frac{\zeta_i}{t}$, where ζ_i is sampled from $[1,10]$. Figure 3 presents the policy performance. The joint prices of sellers converge to the Nash equilibrium prices rapidly. Meanwhile, each seller achieves a sublinear regret. These experiment results are consistent with our Theorem 1. As the number of sellers increases, it takes the market more time to reach the Nash equilibrium.

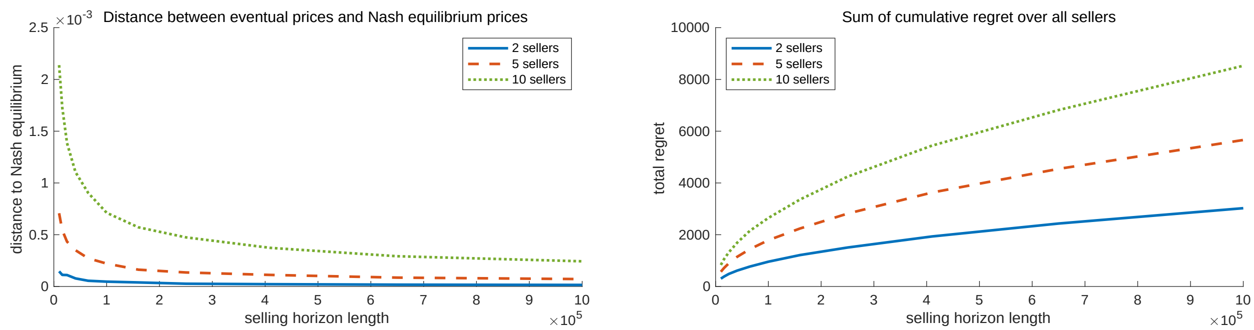
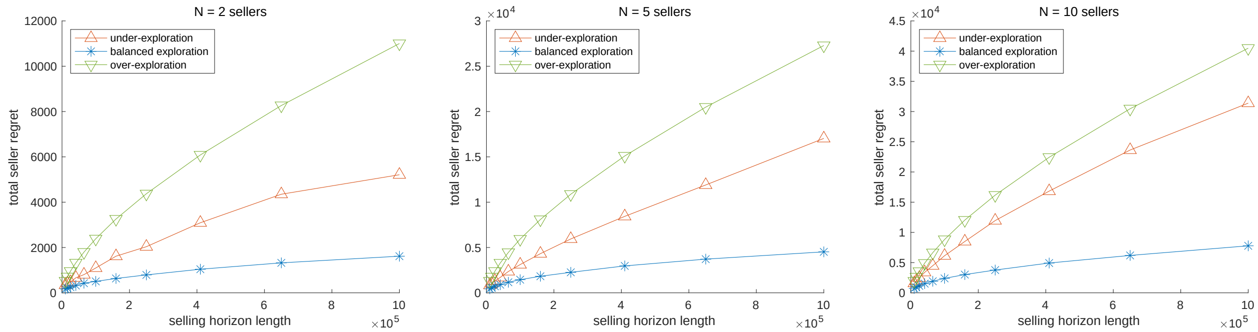


Figure 3 Performance of LEGO Policy in sequential price competition with $N \in \{2,5,10\}$ sellers. The left sub-figure presents $\|\mathbf{p}^{(T)} - \mathbf{p}^*\|_2^2$, *i.e*, distance between the eventual prices and the Nash equilibrium prices. The right sub-figure presents $\sum_{i\in\mathcal{N}} \mathsf{Reg}_i(T)$, *i.e.*, the sum of cumulative regret over all sellers.

6.2. Comparison with Policies of Under-Exploration and Over-Exploration

Our Theorem 1 shows that an optimal order-of-$\sqrt{T}$ regret is achieved when each seller has a balanced exploration phase length of $O(\sqrt{T})$. This subsection further shows that under-exploration and over-exploration both cause larger regret. Specifically, we let each seller apply the LEGO policy with exploration phase length $\tau_i = \iota_i T^{\frac{1}{3}}$ for under-exploration, $\tau_i = \iota_i T^{\frac{1}{2}}$ for balanced exploration, and $\tau_i = \iota_i T^{\frac{2}{3}}$ for over exploration. ι_i is sampled from $[1,2]$. Other settings are the same as those in Subsection 6.1. Figure 4 presents the policy performance. When sellers employ balanced exploration, their regret increases relatively slowly. Particularly, the log-log plot (b) of Figure 4 indicates that the regret is $\tilde{O}(\sqrt{T})$ under balanced exploration, which is consistent with Theorem 1. In contrast, the regret is approximately $O(T^{\frac{2}{3}})$ (as suggested by the slope in the log-log plot) under under- or over-exploration. This is consistent with our analysis: under-exploration is associated with an order-of-$T^{-\frac{1}{3}}$ estimation error that causes regret of $O(T \cdot T^{-\frac{1}{3}}) = O(T^{\frac{2}{3}})$ in the gradient optimization phase, while over-exploration directly causes regret of $O(T^{\frac{2}{3}})$ in the exploration phase.

(a) Sum of cumulative regret over all sellers



(b) Log–Log plot of the sum of cumulative regret

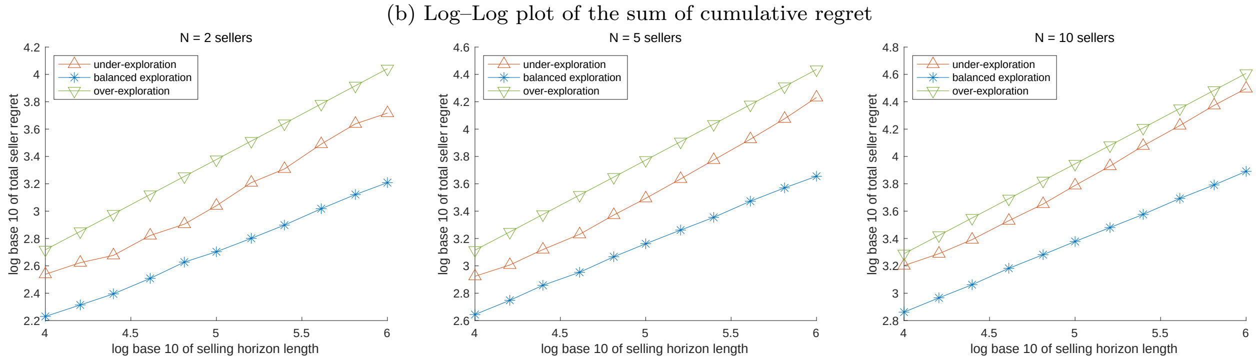


Figure 4 Impact of exploration phase length on the performance of LEGO policy in sequential price competition. Under-exploration, balanced exploration, and over-exploration respectively mean that each seller applies the LEGO policy with exploration phase length $\tau_i = \iota_i T^{\frac{1}{3}}$, $\tau_i = \iota_i T^{\frac{1}{2}}$, and $\tau_i = \iota_i T^{\frac{2}{3}}$ $(i \in \mathcal{N})$. Performance is measured by $\sum_{i\in\mathcal{N}} \mathsf{Reg}_i(T)$, *i.e.*, the sum of cumulative regret over all sellers. Sub-figure (b) is a log-log plot with base 10 of Sub-figure (a). In Sub-figure (b), the slopes of curves for under-exploration, balanced exploration, and over-exploration are $(0.59, 0.49, 0.66)$ when $N = 2$, $(0.65, 0.51, 0.66)$ when $N = 5$, and $(0.65, 0.51, 0.66)$ when $N = 10$.

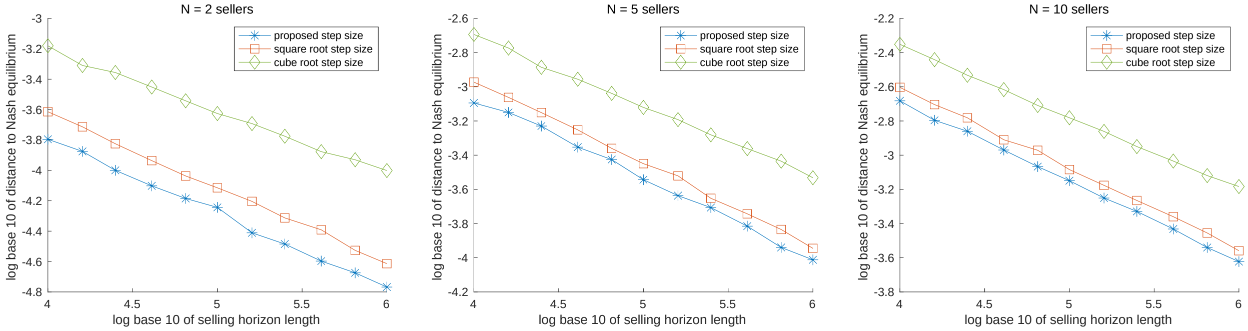
6.3. Robustness of LEGO Policy in Step Sizes

Our numerical experiments show that seller step sizes have a minor impact on the order-of-$\frac{1}{\sqrt{T}}$ convergence rate to Nash equilibrium and the order-of-$\sqrt{T}$ regret in Theorem 1. We test our LEGO policy with sellers applying proposed step sizes $\eta_i^{(t)} = \frac{1}{2t}$ (as in Theorem 1), square root step sizes $\eta_i^{(t)} = \frac{1}{2\sqrt{t}}$, and cube root step sizes $\eta_i^{(t)} = \frac{1}{2\sqrt[3]{t}}$ $(i \in \mathcal{N}, t > \tau_i)$. Other settings are the same as those in Subsection 6.1. Figure 5 indicates similar policy performance under different step sizes.

7. Concluding Remarks

We consider multiple sellers selling a single type of product with unlimited inventories over a selling horizon of T periods. In each period, each seller simultaneously posts their price and observes their private demand, which depends on the prices of all sellers following a noisy and unknown linear model. Sellers can observe the historical prices of competitors but do not know the demand of competitors. We propose a decentralized phased LEGO policy. Under our policy, each seller

(a) Log–Log plot of the distance from eventual prices to Nash equilibrium prices



(b) Log–Log plot of the sum of cumulative regret over all sellers

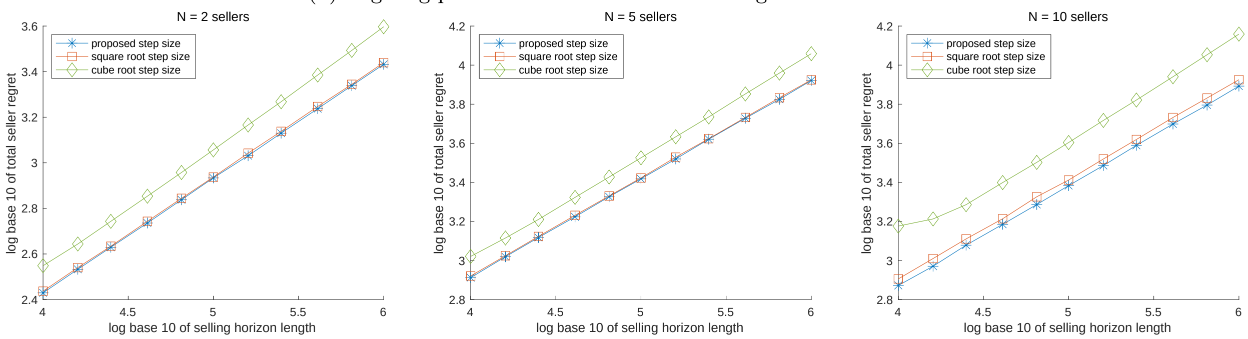


Figure 5 Robustness of LEGO policy in sequential price competition against different step sizes. Proposed, square root, and cube root step sizes respectively mean that each seller applies the LEGO policy with step size $\eta_i^{(t)} = \frac{1}{2t}$, $\eta_i^{(t)} = \frac{1}{2\sqrt{t}}$, and $\eta_i^{(t)} = \frac{1}{2\sqrt[3]{t}}$ $(i \in \mathcal{N}, t > \tau_i)$. In Sub-figure (b), the curve of proposed step sizes overlaps with the curve of square root step sizes. Sub-figure (a) presents $\|\mathbf{p}^{(T)} - \mathbf{p}^*\|_2^2$, *i.e*, distance between the eventual prices and the Nash equilibrium prices. Sub-figure (b) presents $\sum_{i\in\mathcal{N}} \mathsf{Reg}_i(T)$, *i.e.*, the sum of cumulative regret over all sellers.

partitions the selling horizon into two phases: an exploration phase focused on estimating private parameters, and a gradient optimization phase focused on adjusting prices based on estimated gradient feedback. Sellers may have private exploration phase lengths. We demonstrate that, if each seller privately employs an order-of-$\sqrt{T}$ policy (that is, their exploration phase length is of the same order as $\sqrt{T}$), each seller achieves a worst-case regret of $\tilde{O}(\sqrt{T})$, which matches the problem lower bound. Furthermore, the joint prices of sellers at the end of the selling horizon will converge to Nash equilibrium prices at a rate of $O(\frac{1}{\sqrt{T}})$. The joint prices are also a vector of ϵ-Nash equilibrium prices with probability $1 - O(\frac{1}{\epsilon\sqrt{T}})$ for $\epsilon > 0$. Our analysis further demonstrates that if each seller knows their individual price sensitivity coefficient, a gradient optimization policy achieves an improved regret of $O(\log T)$ while seller prices converge to Nash equilibrium at a rate of $O(\frac{1}{T})$. This indicates that the *unknown individual price sensitivity contributes to the major difficulty of dynamic pricing in sequential competition* and forces regret to the order of $\sqrt{T}$ in the worst case.

References

Abada I, Lambin X (2023) Artificial intelligence: Can seemingly collusive outcomes be avoided? *Management Science* 69(9):5042–5065.

Aksoy-Pierson M, Allon G, Federgruen A (2013) Price competition under mixed multinomial logit demand functions. *Management Science* 59(8):1817–1835.

Allon G, Gurvich I (2010) Pricing and dimensioning competing large-scale service providers. *Manufacturing & Service Operations Management* 12(3):449–469.

Alptekinoğlu A, Semple JH (2016) The exponomial choice model: A new alternative for assortment and price optimization. *Operations Research* 64(1):79–93.

Aouad A, den Boer AV (2021) Algorithmic collusion in assortment games. *Available at SSRN 3930364* .

Asker J, Fershtman C, Pakes A (2022) Artificial intelligence, algorithm design, and pricing. *AEA Papers and Proceedings*, volume 112, 452–56.

Banchio M, Mantegazza G (2023) Artificial intelligence and spontaneous collusion. *Available at SSRN* .

Bertrand J (1883) Théorie mathématique de la richesse sociale. *Journal des savants* 67(1883):499–508.

Besbes O, Sauré D (2016) Product assortment and price competition under multinomial logit demand. *Production and Operations Management* 25(1):114–127.

Besbes O, Zeevi A (2012) Blind network revenue management. *Operations research* 60(6):1537–1550.

Besbes O, Zeevi A (2015) On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Science* 61(4):723–739.

Birge JR, Chen H, Keskin NB, Ward A (2024) To interfere or not to interfere: Information revelation and price-setting incentives in a multiagent learning environment. *Operations Research* .

Bravo M, Leslie D, Mertikopoulos P (2018) Bandit learning in concave n-person games. *Advances in Neural Information Processing Systems* 31.

Broder J, Rusmevichientong P (2012) Dynamic pricing under a general parametric choice model. *Operations Research* 60(4):965–980.

Calvano E, Calzolari G, Denicolo V, Pastorello S (2020) Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review* 110(10):3267–3297.

Cartea Á, Chang P, Penalva J (2022a) Algorithmic collusion in electronic markets: The impact of tick size. *Available at SSRN 4105954* .

Cartea A, Chang P, Penalva J, Waldon H (2022b) The algorithmic learning equations: Evolving strategies in dynamic games. *Available at SSRN 4175239* .

Cartea Á, Chang P, Penalva J, Waldon H (2022c) Algorithms can learn to collude: A folk theorem from learning with bounded rationality. *Available at SSRN 4293831* .

Chen B, Chao X, Shi C (2021) Nonparametric learning algorithms for joint pricing and inventory control with lost sales and censored demand. *Mathematics of Operations Research* 46(2):726–756.

Chen N, Chen YJ (2021) Duopoly competition with network effects in discrete choice models. *Operations Research* 69(2):545–559.

Chen Y, Shi C (2023) Network revenue management with online inverse batch gradient descent method. *Production and Operations Management* 32(7):2123–2137.

Cont R, Xiong W (2024) Dynamics of market making algorithms in dealer markets: Learning and tacit collusion. *Mathematical Finance* 34(2):467–521.

Cooper WL, Homem-de Mello T, Kleywegt AJ (2015) Learning and pricing with models that do not explicitly incorporate competition. *Operations research* 63(1):86–103.

Cournot AA (1838) *Recherches sur les principes mathématiques de la théorie des richesses par Augustin Cournot* (Paris, France: chez L. Hachette).

den Boer AV (2023) A (mathematical) definition of algorithmic collusion. *Available at SSRN 4636488* .

den Boer AV, Meylahn JM, Schinkel MP (2022) Artificial collusion: Examining supracompetitive pricing by q-learning algorithms. *Amsterdam Law School Research Paper* (2022-25).

den Boer AV, Zwart B (2014) Simultaneously learning and optimizing using controlled variance pricing. *Management science* 60(3):770–783.

Epivent A, Lambin X (2024) On algorithmic collusion and reward-punishment schemes. *Economics Letters* 111661.

Eschenbaum N, Mellgren F, Zahn P (2022) Robust algorithmic collusion. *arXiv preprint arXiv:2201.00345* .

Federgruen A, Hu M (2015) Multi-product price and assortment competition. *Operations Research* 63(3):572–584.

Federgruen A, Hu M (2016) Sequential multiproduct price competition in supply chain networks. *Operations Research* 64(1):135–149.

Federgruen A, Hu M (2021) Global robust stability in a general price and assortment competition model. *Operations Research* 69(1):164–174.

Gallego G, Hu M (2014) Dynamic pricing of perishable assets under competition. *Management Science* 60(5):1241–1259.

Gallego G, Huh WT, Kang W, Phillips R (2006) Price competition with the attraction demand model: Existence of unique equilibrium and its stability. *Manufacturing & Service Operations Management* 8(4):359–375.

Gallego G, Wang R (2014) Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. *Operations Research* 62(2):450–461.

Gershgorn D (2024) The best mini desktop pcs. *The New York Times* URL https://www.nytimes.com/wirecutter/reviews/best-mini-desktop-pcs/.

Golowich N, Pattathil S, Daskalakis C (2020) Tight last-iterate convergence rates for no-regret learning in multi-player games. *Advances in neural information processing systems* 33:20766–20778.

Golrezaei N, Jaillet P, Liang JCN (2020) No-regret learning in price competitions under consumer reference effects. *Advances in Neural Information Processing Systems* 33:21416–21427.

Goyal V, Li S, Mehrotra S (2023) Learning to price under competition for multinomial logit demand. *Available at SSRN 4572453* .

Guo MA, Ying D, Lavaei J, Shen ZJM (2023) Last-iterate convergence in no-regret learning: Games with reference effects under logit demand. *SSRN* .

Hansen KT, Misra K, Pai MM (2021) Frontiers: Algorithmic collusion: Supra-competitive prices via independent algorithms. *Marketing Science* 40(1):1–12.

Hazan E, et al. (2016) Introduction to online convex optimization. *Foundations and Trends® in Optimization* 2(3-4):157–325.

Hettich M (2021) Algorithmic collusion: Insights from deep learning. *Available at SSRN 3785966* .

Hsieh YG, Antonakopoulos K, Mertikopoulos P (2021) Adaptive learning in continuous games: Optimal regret bounds and convergence to nash equilibrium. *Conference on Learning Theory*, 2388–2422 (PMLR).

Jordan MI, Lin T, Zhou Z (2023) Adaptive, doubly optimal no-regret learning in strongly monotone and exp-concave games with gradient feedback. *arXiv preprint arXiv:2310.14085* .

Kachani S, Perakis G, Simon C (2007) Modeling the transient nature of dynamic pricing with demand learning in a competitive environment. *Network science, nonlinear science and infrastructure systems* 223–267.

Keskin NB, Zeevi A (2014) Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations research* 62(5):1142–1167.

Kirman A (1983) On mistaken beliefs and resultant equilibria. *Individual forecasting and aggregate outcomes* 147–166.

Kirman AP (1975) Learning by firms about demand conditions. *Adaptive economic models*, 137–156 (Elsevier).

Klein T (2018) Assessing autonomous algorithmic collusion: Q-learning under short-run price commitments. Technical report, Tinbergen Institute Discussion Paper.

Klein T (2021) Autonomous algorithmic collusion: Q-learning under sequential pricing. *The RAND Journal of Economics* 52(3):538–558.

Lai TL, Robbins H (1982) Iterated least squares in multiperiod control. *Advances in Applied Mathematics* 3(1):50–73.

Li S, Luo Q, Huang Z, Shi C (2022) Online learning for constrained assortment optimization under markov chain choice model. *Available at SSRN 4079753* .

Li S, Mehrotra S (2024) Adaptive learning in uncertain and sequential competition. *Available at SSRN 4722908* .

Lin T, Zhou Z, Ba W, Zhang J (2021) Doubly optimal no-regret online learning in strongly monotone games with bandit feedback. *arXiv preprint arXiv:2112.02856* .

Lin T, Zhou Z, Mertikopoulos P, Jordan M (2020) Finite-time last-iterate convergence for multi-agent learning in games. *International Conference on Machine Learning*, 6161–6171 (PMLR).

Loots T, den Boer AV (2021) Data-driven collusion and competition in a pricing duopoly with multinomial logit demand. *Production and Operations Management* .

Martin N (2019) Uber charges more if they think you're willing to pay more. *Forbes* URL https://www.forbes.com/sites/nicolemartin1/2019/03/30/uber-charges-more-if-they-think-youre-willing-to-pay-more/?sh=1f8993647365.

McKinsey (2023) What is fast fashion? *McKinsey & Company* URL https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-fast-fashion.

Mertikopoulos P, Zhou Z (2019) Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming* 173:465–507.

Meylahn JM (2023a) Does an intermediate price facilitate algorithmic collusion? *Available at SSRN 4594415* .

Meylahn JM (2023b) Weak acyclicity in games with unique best-responses and implications for algorithmic collusion. *Available at SSRN 4589989* .

Meylahn JM, V den Boer A (2022) Learning to collude in a pricing duopoly. *Manufacturing & Service Operations Management* 24(5):2577–2594.

Morrow WR, Skerlos SJ (2011) Fixed-point approaches to computing bertrand-nash equilibrium prices under mixed-logit demand. *Operations research* 59(2):328–345.

Nemirovskij AS, Yudin DB (1983) Problem complexity and method efficiency in optimization .

Phillips RL (2021) *Pricing and revenue optimization* (Stanford university press).

Qin H (2023) Boba's boom: Reshaping the u.s. beverage landscape. *Michigan Journal Of Economics* URL https://sites.lsa.umich.edu/mje/2023/12/04/bobas-boom-reshaping-the-u-s-beverage-landscape/.

Ren D (2023) Huawei and xiaomi launch new ev models in china, reigniting worries about price wars in the world's largest ev market. *South China Morning Post* URL https://www.scmp.com/business/china-business/article/3246414/huawei-and-xiaomi-launch-new-ev-models-china-reigniting-worries-about-price-wars-worlds-largest-ev.

Sauré D, Zeevi A (2013) Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management* 15(3):387–404.

Scott M, Stillman A, Simon Z, Tanakasempipat P, Burkhardt P, Lorinc J (2023) Ev market's surge toward $57 trillion sparks global flashpoints. *Bloomberg* URL https://www.bloomberg.com/news/features/2023-11-07/the-57-trillion-ev-market-is-a-battleground-for-china-us-eu?itm_source=record&itm_campaign=EV_Slowdown&itm_content=$57_Trillion_Market-4.

Tesauro G, Kephart JO (2002) Pricing in agent economies using multi-agent q-learning. *Autonomous agents and multi-agent systems* 5:289–304.

Valinsky J (2024) Wendy's will test new menus that change prices throughout the day. *CNN* URL https://www.cnn.com/2024/02/27/food/wendys-test-surge-pricing/index.html.

Waltman L, Kaymak U (2008) Q-learning agents in a cournot oligopoly model. *Journal of Economic Dynamics and Control* 32(10):3275–3293.

Wang R, Ke C, Cui S (2022) Product price, quality, and service decisions under consumer choice models. *Manufacturing & Service Operations Management* 24(1):430–447.

Yang Y, Lee YC, Chen PA (2024) Competitive demand learning: A non-cooperative pricing algorithm with coordinated price experimentation. *Production and Operations Management* .

Appendix A: Summary of Major Notation

Table 1 summarizes the major mathematical notation used in the manuscript.

Table 1: Major notation and their definitions

Notation	Definition
$\alpha_i, \underline{\alpha}_i, \bar{\alpha}_i$	linear demand parameter for seller i and bounds
$\beta_i, \underline{\beta}_i, \bar{\beta}_i$	linear demand parameter for seller i and bounds
$\tilde{b}$	constant defined in (12)
b	constant defined in (17)
C_1, C_0	constants in (13), defined in (12)
$\chi_i^{(t)}$	step size of seller i when computing estimators
d_i	constant in (33), defined in (12)
D	constant defined in (12)
$\mathscr{D}_i$	exploration price distribution of seller i
Δ	L^2-normed error of estimators $\{\hat{\beta}_i\}_{i\in\mathcal{N}}$
$\eta_i^{(t)}$	step size of seller i in period t when updating prices
$\varepsilon_i^{(t)}, \boldsymbol{\varepsilon}^{(t)}$	demand noise
$\mathcal{F}^{(t)}/\mathcal{F}_i^{(t)}$	information filtration of all sellers/seller i
$\boldsymbol{\gamma}_i, \bar{\gamma}_i$	linear demand parameters for seller i and bound
h_i	constant defined in (12)
$\underline{\iota}, \bar{\iota}$	constants in bounds of exploration phase length in Theorem 1
κ	constant in Assumption 2
$\lambda_{\min}, \lambda_{\max}$	minimal and maximal eigenvalues in Assumption 1
δ	constant defined in (17)
μ	constants in Lemma 4, defined in (12)
$N, \mathcal{N}$	number of sellers and the set of sellers
ν	step size coefficient ratio in Theorems 1 and 2
ω	constant defined in (12)
$p_i^{(t)}/\mathbf{p}^{(t)}/\mathbf{p}_{-i}^{(t)}$	price (vector) of seller i/all sellers/competitors in period t
$p_i^*/\mathbf{p}^*/\mathbf{p}_{-i}^*$	price (vector) of seller i/all sellers/competitors at Nash equilibrium
$p_i^\epsilon/\mathbf{p}^\epsilon/\mathbf{p}_{-i}^\epsilon$	price (vector) of seller i/all sellers/competitors at ϵ-Nash equilibrium
$\underline{p}_i, \bar{p}_i$	price bounds of seller i
$\mathcal{P}/\mathcal{P}_i$	support of joint prices/seller i's price
Π	projection function
$\phi_i^{(t)}$	estimated gradient of seller i in period t
R_i	cumulative average revenue of seller i
Reg_i	regret of seller i relative to dynamic benchmark
ρ	constant in (14), defined in (12)
$T, \mathcal{T}$	selling horizon length and the set of periods
τ_i	separation period of seller i
$\tau_{\min}, \tau_{\max}$	minimum and maximum of τ_i over i
U_0	bound of demand noise
U_1	bound of feedback second moment in Lemma 2, defined in (12)
U_ℓ	constant in (39), defined in (12)
υ_i	step size coefficient of seller i when updating estimators
$y_i^{(t)}/\mathbf{y}^{(t)}$	demand (vector) of seller i/all sellers
ζ_i	step size coefficient of seller i when updating prices

Appendix B: Proof of Theorems and Lemmas

B.1. Proof of Lemma 1

We leverage the contraction mapping theorem. For $\mathbf{p}=(p_i,\mathbf{p}_{-i})\in\mathcal{P}$ and $i\in\mathcal{N}$, define

$$\Gamma_i(\mathbf{p}):=\frac{\alpha_i+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}}{2\beta_i},\qquad \Gamma_i^*(\mathbf{p}):=\Pi_{\mathcal{P}_i}\big(\Gamma_i(\mathbf{p})\big)=\Pi_{\mathcal{P}_i}\Big(\frac{\alpha_i+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}}{2\beta_i}\Big). \tag{29}$$

Given the joint price vector $\mathbf{p}\in\mathcal{P}$, the average revenue of seller i is $p_i\cdot(\alpha_i-\beta_ip_i+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i})$, which is a strongly concave function in p_i due to $\beta_i>0$. When competitor prices are fixed, $p_i=\Gamma_i(\mathbf{p})$ would maximize the average revenue. Mapping this to seller i's price support yields that, to maximize individual average revenue, the best-response price of seller i is $\Gamma_i^*(\mathbf{p})=\Pi_{\mathcal{P}_i}\big(\Gamma_i(\mathbf{p})\big)$. This implies that the Nash equilibrium prices $\mathbf{p}^*$ defined in (4) are also solutions to $p_i^*=\Gamma_i^*(\mathbf{p}^*), i\in\mathcal{N}$. In other words, $\mathbf{p}^*$ are a fixed point for the mapping $\boldsymbol{\Gamma}^*:=(\Gamma_i^*)_{i\in\mathcal{N}}$ on the compact space $\mathcal{P}$; *i.e.*, $\boldsymbol{\Gamma}^*(\mathbf{p}^*)=\mathbf{p}^*$.

To show that Nash equilibrium prices $\mathbf{p}^*$ exist and are unique, we only need to prove that $\boldsymbol{\Gamma}^*$ is a contraction mapping. For any $\mathbf{p},\mathbf{p}'\in\mathcal{P}$, we have that $|\Gamma_i^*(\mathbf{p})-\Gamma_i^*(\mathbf{p}')|\leqslant|\Gamma_i(\mathbf{p})-\Gamma_i(\mathbf{p}')|=\left|\frac{\boldsymbol{\gamma}_i^\top(\mathbf{p}_{-i}-\mathbf{p}'_{-i})}{2\beta_i}\right|\leqslant\frac{\|\boldsymbol{\gamma}_i\|_1\cdot\|\mathbf{p}_{-i}-\mathbf{p}'_{-i}\|_\infty}{2\beta_i}\leqslant\frac{\|\boldsymbol{\gamma}_i\|_1}{2\beta_i}\cdot\|\mathbf{p}-\mathbf{p}'\|_\infty\leqslant\frac{1}{2\kappa}\cdot\|\mathbf{p}-\mathbf{p}'\|_\infty$, $i\in\mathcal{N}$, where the last inequality is due to Assumption 2. This implies

$$\|\boldsymbol{\Gamma}^*(\mathbf{p})-\boldsymbol{\Gamma}^*(\mathbf{p}')\|_\infty\leqslant\frac{1}{2\kappa}\cdot\|\mathbf{p}-\mathbf{p}'\|_\infty,\qquad \mathbf{p},\mathbf{p}'\in\mathcal{P}. \tag{30}$$

Since $\kappa\geqslant1$, $\boldsymbol{\Gamma}^*$ is a contraction mapping on the compact space $\mathcal{P}$ under the supremum norm.

Because $\boldsymbol{\Gamma}^*(\mathbf{p}^*)=\mathbf{p}^*$, (30) indicates

$$\|\boldsymbol{\Gamma}^*(\mathbf{p})-\mathbf{p}^*\|_\infty\leqslant\frac{1}{2\kappa}\cdot\|\mathbf{p}-\mathbf{p}^*\|_\infty,\qquad \mathbf{p}\in\mathcal{P}. \tag{31}$$

B.2. Proof of Lemma 2

Consider $i\in\mathcal{N}$. When $t=\tau_i$, we have $\phi_i^{(t)}=0$ and $\mathbb{E}[(\phi_i^{(t)})^2\,|\,\mathcal{F}^{(t')}]=0\leqslant U_1$. Then we only need to prove for all $t>\tau_i$. For simplicity of notation, let us write the conditional expectation $\tilde{\mathbb{E}}[\cdot]:=\mathbb{E}[\cdot\,|\,\mathcal{F}^{(t')}]$. By (10), we have

$$\begin{aligned}
\tilde{\mathbb{E}}[\phi_i^{(t)}]^2=&\tilde{\mathbb{E}}\big[(y_i^{(t)}-\hat{\beta}_ip_i^{(t)})^2\big]=\tilde{\mathbb{E}}\Big\{\mathbb{E}\big[(y_i^{(t)}-\hat{\beta}_ip_i^{(t)})^2\big|\mathcal{F}^{(t-1)},\mathbf{p}^{(t)}\big]\Big\}\\
=&\tilde{\mathbb{E}}\Big\{\mathsf{Var}\big[y_i^{(t)}-\hat{\beta}_ip_i^{(t)}\big|\mathcal{F}^{(t-1)},\mathbf{p}^{(t)}\big]+\mathbb{E}^2\big[y_i^{(t)}-\hat{\beta}_ip_i^{(t)}\big|\mathcal{F}^{(t-1)},\mathbf{p}^{(t)}\big]\Big\}\\
=&\tilde{\mathbb{E}}\Big\{\mathsf{Var}\big[y_i^{(t)}-\hat{\beta}_ip_i^{(t)}\big|\mathcal{F}^{(t-1)},\mathbf{p}^{(t)}\big]\Big\}+\tilde{\mathbb{E}}\Big\{\mathbb{E}^2\big[y_i^{(t)}-\hat{\beta}_ip_i^{(t)}\big|\mathcal{F}^{(t-1)},\mathbf{p}^{(t)}\big]\Big\}\\
=&\tilde{\mathbb{E}}\Big\{\mathsf{Var}\big[\varepsilon_i^{(t)}\big|\mathcal{F}^{(t-1)},\mathbf{p}^{(t)}\big]\Big\}+\tilde{\mathbb{E}}\Big\{\mathbb{E}^2\big[\alpha_i-\beta_ip_i^{(t)}+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^{(t)}-\hat{\beta}_ip_i^{(t)}\big|\mathcal{F}^{(t-1)},\mathbf{p}^{(t)}\big]\Big\}\\
\leqslant&U_0+\tilde{\mathbb{E}}\Big\{\mathbb{E}^2\big[\max\{\alpha_i-\beta_ip_i^{(t)}+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^{(t)},\hat{\beta}_ip_i^{(t)}\}\big|\mathcal{F}^{(t-1)},\mathbf{p}^{(t)}\big]\Big\}\\
&\text{(due to }\mathbb{E}[\varepsilon_i^{(t)}]=0,\ \mathbb{E}[\varepsilon_i^{(t)}]^2\leqslant U_0,\ \alpha_i-\beta_ip_i^{(t)}+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^{(t)}\geqslant0,\text{ and }\hat{\beta}_ip_i^{(t)}\geqslant0)\\
\leqslant&U_0+\big(\max\{\bar{\alpha}_i-\underline{\beta}_i\underline{p}_i+\bar{\gamma}_i\max_{j\in\mathcal{N}}\bar{p}_j,\bar{\beta}_i\bar{p}_i\}\big)^2\leqslant U_0+\max\big\{(\bar{\alpha}_i-\underline{\beta}_i\underline{p}_i+\bar{\gamma}_i\max_{j\in\mathcal{N}}\bar{p}_j)^2,\bar{\beta}_i^2\bar{p}_i^2\big\}\leqslant U_1.
\end{aligned}$$

By recalling the definition of $\tilde{\mathbb{E}}[\cdot]$, we have $\mathbb{E}[(\phi_i^{(t)})^2\,|\,\mathcal{F}^{(t')}]\leqslant U_1$.

B.3. Proof of Lemma 3

Consider a price vector $\mathbf{p}^{(t)} \in \mathcal{P}$ in period $t \in \mathcal{T}$. For seller $i \in \mathcal{N}$, fix the competitor prices $\mathbf{p}_i^{(t)}$ and define function $f(p) = p \cdot (\alpha_i - \beta_i p + \boldsymbol{\gamma}_i^\top \mathbf{p}_{-i}^{(t)})$, which represents the average revenue of seller i in period t if seller i has offered a price p. The average revenue function f is quadratic, strongly concave, and $2\beta_i$-smooth. $f(p)$ is maximized when $p = \tilde{p} := \frac{\alpha_i + \boldsymbol{\gamma}_i^\top \mathbf{p}_{-i}^{(t)}}{2\beta_i}$. By definition, we have $\tilde{p} \geqslant \max\left\{\frac{\alpha_i - \bar{\gamma}_i \max_{j\in\mathcal{N}\setminus\{i\}} \bar{p}_j}{2\beta_i}, 0\right\}$ and $\tilde{p} \leqslant \frac{\alpha_i + \bar{\gamma}_i \max_{j\in\mathcal{N}\setminus\{i\}} \bar{p}_j}{2\beta_i}$. Due to Assumption 3, we have $\tilde{p} \geqslant \underline{p}_i$ and $\tilde{p} \leqslant \bar{p}_i$. Thus, $\tilde{p} \in \mathcal{P}_i$ is a valid price of seller i and $f'(\tilde{p}) = 0$. Then we have

$$
\begin{aligned}
f(\tilde{p}) - f(p_i^{(t)}) &= -f'(\tilde{p})(p_i^{(t)} - \tilde{p}) - \frac{f''(\tilde{p})}{2}(p_i^{(t)} - \tilde{p})^2 \leqslant \beta_i (p_i^{(t)} - \tilde{p})^2 \qquad (\text{due to } f'(\tilde{p}) = 0,\ f''(\tilde{p}) = -2\beta_i) \\
&= \beta_i \cdot \left(p_i^{(t)} - \Gamma_i^*(\mathbf{p}^{(t)})\right)^2 \qquad (\text{due to definitions of } \tilde{p} \text{ and } \Gamma_i^*, \text{ and } \tilde{p} \in \mathcal{P}_i) \\
&\leqslant \beta_i \cdot \|\mathbf{p}^{(t)} - \boldsymbol{\Gamma}^*(\mathbf{p}^{(t)})\|_\infty^2 \leqslant \beta_i \cdot (\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_\infty + \|\boldsymbol{\Gamma}^*(\mathbf{p}^{(t)}) - \mathbf{p}^*\|_\infty)^2 \\
&\leqslant \beta_i \cdot (2\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_\infty)^2 = 4\beta_i \cdot \|\mathbf{p}^{(t)} - \mathbf{p}^*\|_\infty^2 \qquad (\text{due to (31)}) \\
&\leqslant 4\beta_i \cdot \|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2 \leqslant 4\bar{\beta}_i \cdot \|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2. \qquad (32)
\end{aligned}
$$

By the definition of regret Reg_i in (3), we have that $\mathsf{Reg}_i(t) - \mathsf{Reg}_i(t-1) = \mathbb{E}\left[f(\tilde{p}) - f(p_i^{(t)})\right] \leqslant 4\bar{\beta}_i \cdot \mathbb{E}\left[\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2\right]$.

B.4. Proof of Lemma 4

In Lemma 3, (32) has shown that, for seller i, $f(\tilde{p}) - f(p_i^{(t)}) \leqslant 4\bar{\beta}_i \cdot \|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2$. Because $\tilde{p}$ maximizes f, we have $f(p_i^{(t)}) \geqslant f(p_i') - 4\bar{\beta}_i \cdot \|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2$ for all $p_i' \in \mathcal{P}_i$. Because $\|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2 \leqslant \mu\epsilon$ and $\mu = \dfrac{1}{4\max_{i\in\mathcal{N}} \bar{\beta}_i}$, we have $f(p_i^{(t)}) \geqslant f(p_i') - 4\bar{\beta}_i \cdot \|\mathbf{p}^{(t)} - \mathbf{p}^*\|_2^2 \geqslant f(p_i') - \epsilon$ for all $p_i' \in \mathcal{P}_i$. By recalling the definitions of average revenue function f and ϵ-Nash equilibrium prices, this set of inequalities suggests $\mathbf{p}^{(t)}$ is a vector of ϵ-Nash equilibrium prices.

B.5. Proof of Inequality (13)

Consider $i \in \mathcal{N}$. For simplicity of notation, let us write

$$
\boldsymbol{\theta}_i^* := (\alpha_i, \beta_i, \boldsymbol{\gamma}_i)^\top, \qquad \hat{\boldsymbol{\theta}}_i^{(t)} := (\hat{\alpha}_i^{(t)}, \hat{\beta}_i^{(t)}, \hat{\boldsymbol{\gamma}}_i^{(t)})^\top, \qquad \mathbf{x}^{(t)} := (1, -p_i^{(t)}, \mathbf{p}_{-i}^{(t)})^\top.
$$

$\boldsymbol{\theta}_i^*$, $\hat{\boldsymbol{\theta}}_i^{(t)}$, and $\mathbf{x}^{(t)}$ are all column vectors in $\mathbb{R}^{N+1}$. We write the parameter space as

$$
\Theta_i := \{(\alpha, \beta, \boldsymbol{\gamma}) \,|\, \underline{\alpha}_i \leqslant \alpha \leqslant \bar{\alpha}_i, \underline{\beta}_i \leqslant \beta \leqslant \bar{\beta}_i, \|\boldsymbol{\gamma}\|_1 \leqslant \bar{\gamma}_i\}.
$$

Then we have

$$
\|\boldsymbol{\theta}' - \boldsymbol{\theta}''\|_2^2 \leqslant d_i, \qquad \boldsymbol{\theta}', \boldsymbol{\theta}'' \in \Theta_i. \tag{33}
$$

Subsubsections B.5.1–B.5.2 will respectively prove the following two inequalities.

$$\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\tau_{\min}+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] \leqslant \frac{C_0}{\tau_{\min}}, \tag{34}$$

$$\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\tau_i+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] \leqslant \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\tau_{\min}+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] + \frac{\upsilon_i^2 U_\ell}{\tau_{\min}}. \tag{35}$$

Recall that $\hat{\boldsymbol{\theta}}_i^{(t)} = (\hat{\alpha}_i^{(t)}, \hat{\beta}_i^{(t)}, \hat{\boldsymbol{\gamma}}_i^{(t)})^\top$. Therefore, inequalities (34) and (35) respectively have the following interpretations. (i) Seller i's estimator $(\hat{\alpha}_i^{(t)}, \hat{\beta}_i^{(t)}, \hat{\boldsymbol{\gamma}}_i^{(t)})$ converges at a rate of $\frac{C_0}{t}$ during the common exploration periods from $t=1$ to $t=\tau_{\min}$. (ii) After period $\tau_{\min}$, although seller i's estimator $(\hat{\alpha}_i^{(t)}, \hat{\beta}_i^{(t)}, \hat{\boldsymbol{\gamma}}_i^{(t)})$ may not continue to become more accurate, it does not become significantly less accurate either: from period $t=\tau_{\min}$ to $t=\tau_i$, the error of $(\hat{\alpha}_i^{(t)}, \hat{\beta}_i^{(t)}, \hat{\boldsymbol{\gamma}}_i^{(t)})$ increases by $O(\frac{1}{\tau_{\min}})$.

Combining (34)–(35), we have that

$$\mathbb{E}[|\hat{\beta}_i - \beta_i|^2] = \mathbb{E}[|\hat{\beta}_i^{(\tau_i+1)} - \beta_i|^2] \leqslant \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\tau_i+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] \leqslant \frac{C_0}{\tau_{\min}} + \frac{\upsilon_i^2 U_\ell}{\tau_{\min}}, \tag{36}$$

and $\mathbb{E}\big[\sum_{i\in\mathcal{N}} |\hat{\beta}_i - \beta_i|^2\big] \leqslant \frac{C_0 N}{\tau_{\min}} + \frac{U_\ell \sum_{i\in\mathcal{N}} \upsilon_i^2}{\tau_{\min}} \leqslant \frac{C_0 N}{\iota\sqrt{T}} + \frac{U_\ell \sum_{i\in\mathcal{N}} \upsilon_i^2}{\iota\sqrt{T}} = \frac{C_1}{\sqrt{T}}$. This completes the proof of (13).

Before presenting the proof of (34)–(35), we first define useful functions and constants. For $t \in \mathcal{T}$, define the cost function

$$\ell^{(t)}(\boldsymbol{\theta}_i) = \frac{1}{2} \cdot (\boldsymbol{\theta}_i^\top \mathbf{x}^{(t)} - y_i^{(t)})^2, \qquad \boldsymbol{\theta}_i \in \Theta_i.$$

The gradient and the Hessian of the cost function are respectively

$$\nabla \ell^{(t)}(\boldsymbol{\theta}_i) = (\boldsymbol{\theta}_i^\top \mathbf{x}^{(t)} - y_i^{(t)}) \cdot \mathbf{x}^{(t)}, \qquad \nabla^2 \ell^{(t)}(\boldsymbol{\theta}_i) = \mathbf{x}^{(t)} \mathbf{x}^{(t)\top}, \qquad \boldsymbol{\theta}_i \in \Theta_i. \tag{37}$$

We note that the update rule (8) can be recast as

$$\hat{\boldsymbol{\theta}}_i^{(t+1)} = \Pi_{\Theta_i}\big(\hat{\boldsymbol{\theta}}_i^{(t)} - \chi_i^{(t)} \nabla \ell^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)})\big), \qquad t \in \{1, 2, \ldots, \tau_i\}. \tag{38}$$

As proved in Subsubsection B.5.3, we have

$$\mathbb{E}\big[\|\nabla \ell^{(t)}(\boldsymbol{\theta}_i)\|_2^2\big] \leqslant U_\ell, \qquad \boldsymbol{\theta}_i \in \Theta_i,\, t \in \mathcal{T}. \tag{39}$$

According to Assumption 1, we have

$$\lambda_{\min} \mathbf{I} \preccurlyeq \mathbb{E}[\mathbf{x}^{(t)} \mathbf{x}^{(t)\top} \mid \mathcal{F}^{(t-1)}] \preccurlyeq \lambda_{\max} \mathbf{I}, \qquad t \in \{1, 2, \ldots, \tau_{\min}\}. \tag{40}$$

B.5.1. Proof of Inequality (34). For all $t \in \{1,2,\ldots,\tau_i\}$, we have that

$$
\begin{aligned}
\|\hat{\boldsymbol{\theta}}_i^{(t+1)} - \boldsymbol{\theta}_i^*\|_2^2 =& \|\Pi_{\Theta_i}(\hat{\boldsymbol{\theta}}_i^{(t)} - \chi_i^{(t)}\nabla\ell^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)})) - \boldsymbol{\theta}_i^*\|_2^2 \leqslant \|\hat{\boldsymbol{\theta}}_i^{(t)} - \chi_i^{(t)}\nabla\ell^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)}) - \boldsymbol{\theta}_i^*\|_2^2 \\
& \text{(due to (38) and definition of projection } \Pi_{\Theta_i}(\cdot)) \\
=& \|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2 + \|\chi_i^{(t)}\nabla\ell^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)})\|_2^2 - 2\chi_i^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)^\top\nabla\ell^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)}) \\
=& \|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2 + \|\chi_i^{(t)}\nabla\ell^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)})\|_2^2 - 2\chi_i^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)^\top\mathbf{x}^{(t)}\cdot(\hat{\boldsymbol{\theta}}_i^{(t)\top}\mathbf{x}^{(t)} - y_i^{(t)}) \qquad \text{(by (37))} \\
=& \|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2 + \|\chi_i^{(t)}\nabla\ell^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)})\|_2^2 - 2\chi_i^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)^\top\mathbf{x}^{(t)}\cdot((\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)^\top\mathbf{x}^{(t)} - \varepsilon_i^{(t)}) \\
=& \|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2 + \|\chi_i^{(t)}\nabla\ell^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)})\|_2^2 - 2\chi_i^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)^\top\mathbf{x}^{(t)}\mathbf{x}^{(t)\top}(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*) \\
& + 2\chi_i^{(t)}\varepsilon_i^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)^\top\mathbf{x}^{(t)}.
\end{aligned} \tag{41}
$$

For all $t \in \{1,2,\ldots,\tau_{\min}\}$, taking expectations for both sides of (41) yields that

$$
\begin{aligned}
\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] \leqslant& \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + {\chi_i^{(t)}}^2\mathbb{E}\big[\|\nabla\ell^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)})\|_2^2\big] - 2\chi_i^{(t)}\mathbb{E}\Big[(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)^\top\mathbf{x}^{(t)}\mathbf{x}^{(t)\top}(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)\Big] \\
& \text{(due to independence of } \varepsilon_i^{(t)} \text{ and } \mathbb{E}[\varepsilon_i^{(t)}] = 0) \\
\leqslant& \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + {\chi_i^{(t)}}^2 U_\ell - 2\chi_i^{(t)}\mathbb{E}\Big[\mathbb{E}\big[(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)^\top\mathbf{x}^{(t)}\mathbf{x}^{(t)\top}(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)\,|\,\mathcal{F}^{(t-1)}\big]\Big] \\
& \text{(due to inequality (39))} \\
=& \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + {\chi_i^{(t)}}^2 U_\ell - 2\chi_i^{(t)}\mathbb{E}\Big[(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)^\top\mathbb{E}[\mathbf{x}^{(t)}\mathbf{x}^{(t)\top}\,|\,\mathcal{F}^{(t-1)}](\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)\Big] \\
& \text{(due to } (\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*) \text{ is } \mathcal{F}^{(t-1)}\text{-measurable)} \\
\leqslant& \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + {\chi_i^{(t)}}^2 U_\ell - 2\chi_i^{(t)}\mathbb{E}\Big[(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)^\top(\lambda_{\min}\mathbf{I})(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)\Big] \\
& \text{(due to inequality (40))} \\
=& \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + \frac{\upsilon_i^2 U_\ell}{t^2} - \frac{2\lambda_{\min}\upsilon_i}{t}\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] \qquad \text{(due to definition of } \chi_i^{(t)}) \\
=& \Big(1 - \frac{2\lambda_{\min}\upsilon_i}{t}\Big)\cdot\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + \frac{\upsilon_i^2 U_\ell}{t^2}.
\end{aligned} \tag{42}
$$

Recall the definition of $h_i = \frac{\upsilon_i^2 U_\ell}{2\lambda_{\min}\upsilon_i - 1}$ in (12). Plugging this into (41) yields that, for all $t \in \{1,2,\ldots,\tau_{\min}\}$,

$$
\begin{aligned}
\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] \leqslant& \Big(1 - \frac{2\lambda_{\min}\upsilon_i}{t}\Big)\cdot\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + \frac{h_i(2\lambda_{\min}\upsilon_i - 1)}{t^2} \\
=& \Big(1 - \frac{2\lambda_{\min}\upsilon_i}{t}\Big)\cdot\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + \frac{2\lambda_{\min}\upsilon_i h_i}{t^2} - \frac{h_i}{t^2} \\
\leqslant& \Big(1 - \frac{2\lambda_{\min}\upsilon_i}{t}\Big)\cdot\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + \frac{2\lambda_{\min}\upsilon_i h_i}{t^2} - \frac{h_i}{t} + \frac{h_i}{t+1}.
\end{aligned}
$$

This can be written as

$$
\Big(\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] - \frac{h_i}{t+1}\Big) \leqslant \Big(1 - \frac{2\lambda_{\min}\upsilon_i}{t}\Big)\cdot\Big(\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] - \frac{h_i}{t}\Big), \qquad t \in \{1,2,\ldots,\tau_{\min}\}.
$$

Therefore, if $\tau_{\min} \geqslant \lceil 2\lambda_{\min}\upsilon_i \rceil$, we have that

$$\begin{aligned}
\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\tau_{\min}+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] - \frac{h_i}{\tau_{\min}+1} \leqslant & \Big\{\prod_{u=\lceil 2\lambda_{\min}\upsilon_i\rceil}^{\tau_{\min}} (1-\frac{2\lambda_{\min}\upsilon_i}{u})\Big\} \cdot \Big(\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\lceil 2\lambda_{\min}\upsilon_i\rceil)} - \boldsymbol{\theta}_i^*\|_2^2\big] - \frac{h_i}{\lceil 2\lambda_{\min}\upsilon_i\rceil}\Big) \\
& \text{(due to } 1-\frac{2\lambda_{\min}\upsilon_i}{u} \geqslant 0 \text{ for } u \geqslant \lceil 2\lambda_{\min}\upsilon_i\rceil) \\
\leqslant & \Big\{\prod_{u=\lceil 2\lambda_{\min}\upsilon_i\rceil}^{\tau_{\min}} (1-\frac{2\lambda_{\min}\upsilon_i}{u})\Big\} \cdot \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\lceil 2\lambda_{\min}\upsilon_i\rceil)} - \boldsymbol{\theta}_i^*\|_2^2\big] \\
\leqslant & \Big\{\prod_{u=\lceil 2\lambda_{\min}\upsilon_i\rceil}^{\tau_{\min}} (1-\frac{2\lambda_{\min}\upsilon_i}{u})\Big\} \cdot d_i \leqslant \Big\{\prod_{u=\lceil 2\lambda_{\min}\upsilon_i\rceil}^{\tau_{\min}} (1-\frac{1}{u})\Big\} \cdot d_i \\
& \text{(due to (33) and } 2\lambda_{\min}\upsilon_i > 1 \text{ as in Assumption 1)} \\
= & \frac{\lceil 2\lambda_{\min}\upsilon_i\rceil - 1}{\tau_{\min}} d_i \leqslant \frac{(2\lambda_{\min}\upsilon_i+1)d_i}{\tau_{\min}}.
\end{aligned}$$

This suggests $\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\tau_{\min}+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] \leqslant \frac{h_i}{\tau_{\min}+1} + \frac{(2\lambda_{\min}\upsilon_i+1)d_i}{\tau_{\min}} \leqslant \frac{h_i+(2\lambda_{\min}\upsilon_i+1)d_i}{\tau_{\min}}$. If $\tau_{\min} < \lceil 2\lambda_{\min}\upsilon_i\rceil$, we also have $\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\tau_{\min}+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] \leqslant d_i \leqslant \frac{(2\lambda_{\min}\upsilon_i+1)d_i}{\tau_{\min}} \leqslant \frac{h_i+(2\lambda_{\min}\upsilon_i+1)d_i}{\tau_{\min}}$. Thus, we have $\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\tau_{\min}+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] \leqslant \frac{h_i+(2\lambda_{\min}\upsilon_i+1)d_i}{\tau_{\min}} \leqslant \frac{C_0}{\tau_{\min}}$, which is (34).

B.5.2. Proof of Inequality (35). For all $t \in \{1,2,\ldots,\tau_i\}$, taking expectations for both sides of (41) yields that

$$\begin{aligned}
\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] \leqslant & \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + {\chi_i^{(t)}}^2 \mathbb{E}\big[\|\nabla\ell^{(t)}(\hat{\boldsymbol{\theta}}_i^{(t)})\|_2^2\big] - 2\chi_i^{(t)}\mathbb{E}\Big[(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)^\top \mathbf{x}^{(t)}\mathbf{x}^{(t)\top}(\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)\Big] \\
& \text{(due to independence of } \varepsilon_i^{(t)} \text{ and } \mathbb{E}[\varepsilon_i^{(t)}]=0) \\
\leqslant & \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + {\chi_i^{(t)}}^2 U_\ell - 2\chi_i^{(t)}\mathbb{E}\Big[\big((\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*)^\top\mathbf{x}^{(t)}\big)^2\Big] \leqslant \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + {\chi_i^{(t)}}^2 U_\ell \\
= & \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(t)} - \boldsymbol{\theta}_i^*\|_2^2\big] + \frac{\upsilon_i^2 U_\ell}{t^2}. \qquad \text{(due to definition of } \chi_i^{(t)})
\end{aligned} \tag{43}$$

Summing the above inequality over $t = \tau_{\min}+1$ to τ_i yields that

$$\begin{aligned}
\mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\tau_i+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] \leqslant & \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\tau_{\min}+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] + \upsilon_i^2 U_\ell \sum_{u=\tau_{\min}+1}^{\tau_i} \frac{1}{u^2} \\
\leqslant & \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\tau_{\min}+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] + \upsilon_i^2 U_\ell \sum_{u=\tau_{\min}+1}^{\tau_i} \Big(\frac{1}{u-1} - \frac{1}{u}\Big) \\
\leqslant & \mathbb{E}\big[\|\hat{\boldsymbol{\theta}}_i^{(\tau_{\min}+1)} - \boldsymbol{\theta}_i^*\|_2^2\big] + \frac{\upsilon_i^2 U_\ell}{\tau_{\min}}.
\end{aligned}$$

This completes the proof of (35).

B.5.3. Proof of Inequality (39). We have

$$\begin{aligned}
\mathbb{E}\big[\|\nabla\ell^{(t)}(\boldsymbol{\theta}_i)\|_2^2\big] = & \mathbb{E}\big[(\boldsymbol{\theta}_i^\top\mathbf{x}^{(t)} - y_i^{(t)})^2 \cdot \mathbf{x}^{(t)\top}\mathbf{x}^{(t)}\big] \qquad \text{(due to (37))} \\
\leqslant & \mathbb{E}\Big[(\boldsymbol{\theta}_i^\top\mathbf{x}^{(t)} - y_i^{(t)})^2 \cdot \big(1 + \sum_{j\in\mathcal{N}} (p_j^{(t)})^2\big)\Big] \qquad \text{(due to } \mathbf{x}^{(t)} := (1, -p_i^{(t)}, \mathbf{p}_{-i}^{(t)})^\top)
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot\mathbb{E}\big[(\boldsymbol{\theta}_i^\top\mathbf{x}^{(t)}-y_i^{(t)})^2\big]=\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot\mathbb{E}\Big[\big((\boldsymbol{\theta}_i-\boldsymbol{\theta}_i^*)^\top\mathbf{x}^{(t)}-\varepsilon_i^{(t)}\big)^2\Big]\\
&=\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot\mathbb{E}\Big[\big((\boldsymbol{\theta}_i-\boldsymbol{\theta}_i^*)^\top\mathbf{x}^{(t)}\big)^2+(\varepsilon_i^{(t)})^2\Big]\quad(\text{by }\mathbb{E}[\varepsilon_i^{(t)}]=0\text{ and independence of }\varepsilon_i^{(t)})\\
&\leqslant \big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot\mathbb{E}\Big[\big((\boldsymbol{\theta}_i-\boldsymbol{\theta}_i^*)^\top\mathbf{x}^{(t)}\big)^2\Big]+\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot U_0\qquad(\text{due to }\mathbb{E}[(\varepsilon_i^{(t)})^2]\leqslant U_0)\\
&=\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot\mathbb{E}\big[(\boldsymbol{\theta}_i-\boldsymbol{\theta}_i^*)^\top\mathbf{x}^{(t)}\mathbf{x}^{(t)\top}(\boldsymbol{\theta}_i-\boldsymbol{\theta}_i^*)\big]+\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot U_0\\
&=\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot(\boldsymbol{\theta}_i-\boldsymbol{\theta}_i^*)^\top\mathbb{E}\big[\mathbf{x}^{(t)}\mathbf{x}^{(t)\top}\big](\boldsymbol{\theta}_i-\boldsymbol{\theta}_i^*)+\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot U_0\\
&\leqslant \big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot(\boldsymbol{\theta}_i-\boldsymbol{\theta}_i^*)^\top\big(\lambda_{\max}\mathbf{I}\big)(\boldsymbol{\theta}_i-\boldsymbol{\theta}_i^*)+\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot U_0\qquad(\text{due to }(40))\\
&=\lambda_{\max}\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot\|\boldsymbol{\theta}_i-\boldsymbol{\theta}_i^*\|_2^2+\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot U_0\qquad(\text{due to }(33))\\
&\leqslant \lambda_{\max}\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot d_i+\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\cdot U_0=(\lambda_{\max}d_i+U_0)\cdot\big(1+\sum_{j\in\mathcal{N}}\bar{p}_j^2\big)\leqslant U_\ell.
\end{aligned}
$$

This completes the proof of (39).

B.6. Proof of Inequality (15)

Define $\tilde{\phi}_i^{(t)}:=y_i^{(t)}-\beta_i p_i^{(t)}$ for $t\in\{\tau_i+1,\tau_i+2,\ldots,T\}$ and $i\in\mathcal{N}$. Then we have

$$
\phi_i^{(t)}=\tilde{\phi}_i^{(t)}+(\beta_i-\hat{\beta}_i)p_i^{(t)},\qquad t\in\{\tau_i+1,\tau_i+2,\ldots,T\}. \tag{44}
$$

Similar to (18), for all $t\in\{\tau_{\max}+1,\tau_{\max}+2,\ldots,T-1\}$, we have that

$$
\begin{aligned}
\|\mathbf{p}^{(t+1)}-\mathbf{p}^*\|_2^2\leqslant&\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2+\frac{\sum_{i\in\mathcal{N}}\zeta_i^2(\phi_i^{(t)})^2}{t^2}+\frac{2}{t}\sum_{i\in\mathcal{N}}\zeta_i\phi_i^{(t)}\cdot(p_i^{(t)}-p_i^*)\\
=&\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2+\frac{\sum_{i\in\mathcal{N}}\zeta_i^2(\phi_i^{(t)})^2}{t^2}+\frac{2}{t}\sum_{i\in\mathcal{N}}\zeta_i\tilde{\phi}_i^{(t)}\cdot(p_i^{(t)}-p_i^*)+\frac{2}{t}\sum_{i\in\mathcal{N}}\zeta_i p_i^{(t)}(\beta_i-\hat{\beta}_i)(p_i^{(t)}-p_i^*).
\end{aligned}
$$

Taking conditional expectations $\tilde{\mathbb{E}}[\cdot]$ for both sides yields that

$$
\begin{aligned}
\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t+1)}-\mathbf{p}^*\|_2^2\big]\leqslant&\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big]+\frac{\sum_{i\in\mathcal{N}}\zeta_i^2\tilde{\mathbb{E}}[\phi_i^{(t)}]^2}{t^2}+\tilde{\mathbb{E}}\Big[\frac{2}{t}\sum_{i\in\mathcal{N}}\zeta_i\mathbb{E}[\tilde{\phi}_i^{(t)}\cdot(p_i^{(t)}-p_i^*)|\mathcal{F}^{(t-1)}]\Big]\\
&+\tilde{\mathbb{E}}\Big[\frac{2}{t}\sum_{i\in\mathcal{N}}\zeta_i p_i^{(t)}(\beta_i-\hat{\beta}_i)(p_i^{(t)}-p_i^*)\Big]\\
\leqslant&\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big]+\frac{U_1\sum_{i\in\mathcal{N}}\zeta_i^2}{t^2}+\tilde{\mathbb{E}}\Big[\frac{2}{t}\sum_{i\in\mathcal{N}}\zeta_i\big(\alpha_i-2\beta_i p_i^{(t)}+\boldsymbol{\gamma}_i^\top\mathbf{p}_{-i}^{(t)}\big)(p_i^{(t)}-p_i^*)\Big]\\
&+\tilde{\mathbb{E}}\Big[\frac{2}{t}\sum_{i\in\mathcal{N}}\frac{\omega^{-1}\zeta_i^2(p_i^{(t)})^2(\beta_i-\hat{\beta}_i)^2+\omega(p_i^{(t)}-p_i^*)^2}{2}\Big]
\end{aligned}
$$

(due to the bound of $\tilde{\mathbb{E}}[\phi_i^{(t)}]^2$ in Lemma 2, definition of $\tilde{\phi}_i^{(t)}$, and Cauchy-Schwarz inequality)

$$
\begin{aligned}
&\leqslant \tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big]+\frac{U_1\sum_{i\in\mathcal{N}}\zeta_i^2}{t^2}-\frac{2\omega}{t}\cdot\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big]+\tilde{\mathbb{E}}\Big[\frac{\rho}{t}\sum_{i\in\mathcal{N}}(\beta_i-\hat{\beta}_i)^2\Big]\\
&\quad+\frac{\omega}{t}\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big] \qquad \text{(due to (22))}\\
&\leqslant\big(1-\frac{\omega}{t}\big)\cdot\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big]+\frac{(\omega-1)\tilde{b}}{t^2}+\frac{\rho\Delta}{t}\\
&\quad\Big(\text{due to } \tilde{b}=\frac{U_1\sum_{i\in\mathcal{N}}\zeta_i^2}{\omega-1} \text{ in (12) and } \sum_{i\in\mathcal{N}}(\beta_i-\hat{\beta}_i)^2=\Delta\Big)\\
&\leqslant\big(1-\frac{\omega}{t}\big)\cdot\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2\big]+\frac{\omega\tilde{b}}{t^2}-\frac{\tilde{b}}{t}+\frac{\tilde{b}}{t+1}+\frac{\rho\Delta}{t}. \qquad (\text{due to } \omega>1,\ \tilde{b}>0)
\end{aligned}
$$

This can be written as $\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t+1)}-\mathbf{p}^*\|_2^2-\frac{\tilde{b}}{t+1}\big]\leqslant(1-\frac{\omega}{t})\cdot\tilde{\mathbb{E}}\big[\|\mathbf{p}^{(t)}-\mathbf{p}^*\|_2^2-\frac{\tilde{b}}{t}\big]+\frac{\rho\Delta}{t}$ for $t\in\{\tau_{\max}+1,\tau_{\max}+2,\ldots,T-1\}$, which is (15).