Proceedings of Machine Learning Research vol XXX:1–32, 2026

Strategy-robust Online Learning in Contextual Pricing

Joon Suk Huh JOON@CS.WISC.EDU
University of Wisconsin–Madison

Kirthevasan Kandasamy KANDASAMY@CS.WISC.EDU
Univeristy of Wisconsin–Madison

Abstract

Learning effective pricing strategies is crucial in digital marketplaces, especially when buyers' valuations are unknown and must be inferred through interaction. We study the online contextual pricing problem, where a seller observes a stream of context–valuation pairs and dynamically sets prices. Moreover, departing from traditional online learning frameworks, we consider a strategic setting in which buyers may misreport valuations to influence future prices—a challenge known as strategic overfitting (Amin et al., 2013).

We introduce a strategy-robust notion of regret for multi-buyer online environments, capturing worst-case strategic behavior in the spirit of the Price of Anarchy. Our first contribution is a polynomial-time approximation scheme (PTAS) for learning linear pricing policies in adversarial, adaptive environments, enabled by a novel online sketching technique. Building on this result, we propose our main construction: the Sparse Update Mechanism (SUM), a simple yet effective sequential mechanism that ensures robustness to *all* Nash equilibria among buyers. Moreover, our construction yields a black-box reduction from online expert algorithms to strategy-robust learners.

Keywords: Online Learning, Dynamic Pricing, Mechanism Design, Algorithmic Game Theory

1. Introduction

The ability to learn how to price effectively is a key component of modern digital markets, enabling the efficient exchange of goods and services. This problem has been studied extensively (Gallego and Van Ryzin, 1994; Kleinberg and Leighton, 2003; Besbes and Zeevi, 2009; Dasu and Tong, 2010; Adamczyk et al., 2015; Besbes and Zeevi, 2015; Den Boer, 2015; Vanunts and Drutsa, 2019; Leme et al., 2021; Romano et al., 2021; Jazi et al., 2025) since the rise of e-commerce. When the seller knows the distribution of buyers' valuations for a good, the revenue-maximizing price can be computed directly from that distribution. In the learning-to-price framework, however, the seller does not know the value distribution and must instead learn it from data on buyers' valuations.

In many practical settings, pricing decisions benefit from leveraging not only valuation data but also contextual features that capture relevant information about the goods, buyers, or the market. This gives rise to the framework of *contextual pricing*, where machine learning models are used to map contexts to prices. In the *online contextual pricing* setting, the seller observes a stream of context–valuation pairs and learns a pricing policy sequentially. In this work, we study the case where the pricing function is linear in the context features—a simple yet expressive model that underlies more complex approaches such as kernel methods and neural networks. The corresponding batch

setting was formalized as Myersonian regression (Mohri and Medina, 2014; Munoz and Vassilvitskii, 2017; Shen et al., 2019; Liu et al., 2020), and our work extends its ideas to the online setting, accounting for both myopic and non-myopic strategic bidders.

Most prior work assumes access to a reliable data source, either in batch (Mohri and Medina, 2014; Munoz and Vassilvitskii, 2017; Shen et al., 2019; Liu et al., 2020) or online form (Javanmard and Nazerzadeh, 2019; Tang et al., 2020; Ban and Keskin, 2021; den Boer and Keskin, 2022; Tullii et al., 2024), that accurately reflects buyers' true valuations. This assumption is reasonable in the batch setting, where data can be collected through market research or other trusted channels. In contrast, the assumption is much more fragile in the online setting, where data is generated through sequential interactions with buyers. Here, the seller typically elicits a buyer's valuation through a bidding mechanism, effectively turning the process into a form of *repeated auction*[1]. This opens the door to strategic behavior. There is no *a priori* reason to believe that buyers will report their valuations truthfully or refrain from manipulating the pricing algorithm to their own benefit. The problem becomes especially pronounced when buyers interact with the seller multiple times, rather than appearing only once.

As a concrete example, consider a buyer, such as a bakery, that purchases milk daily from a producer. The buyer could strategically underbid—offering less than their true valuation—for several days, resulting in no transactions. This behavior may send misleading signals to the producer, potentially causing them to lower the price in subsequent days. As a result, the buyer may make future purchases at a significantly reduced cost, ultimately increasing the buyer's utility at the expense of the producer's revenue, despite earlier utility losses due to declined transactions.

Strategy-robustness. A line of work Amin et al. (2013, 2014); Mohri and Munoz (2015); Deng et al. (2019); Drutsa (2020); Zhiyanov and Drutsa (2020); Liu et al. (2024); Golrezaei et al. (2023); Liu et al. (2018), beginning with Amin et al. (2013), has studied the challenge of learning under strategic behavior, with a focus on preventing the phenomenon of *strategic overfitting*, as illustrated in the example above. These works establish regret bounds under the assumption that buyers act to maximize their own surplus. A natural question in this context is whether a regret guarantee applies to a particular, potentially best-case, utility-maximizing behavior, or whether the guarantee is robust to *all* possible utility-maximizing behaviors. In game-theoretic terms, this distinction corresponds to two different notions: the *Price-of-Stability* (PoS) (Schulz and Moses, 2003), which measures efficiency under best-case strategic behavior, and the *Price-of-Anarchy* (PoA) (Koutsoupias and Papadimitriou, 1999; Chung et al., 2008; Dubey, 1986), which captures efficiency under *worst-case strategic behavior*. Hence, we refer to PoA-type guarantees as *strategy-robust*.

In this work, we introduce a natural multi-buyer generalization of regret that explicitly captures the notion of Price-of-Anarchy, which we refer to as *strategy-robust regret*. That is, our regret guarantee holds even under worst-case utility-maximizing strategic behavior across all buyers. Importantly, because our setting is inherently multi-buyer, the notion of a "utility-maximizing buyer" is replaced by that of a Nash equilibrium, making our formulation more directly aligned with the concept of Price-of-Anarchy. To the best of our knowledge, this is the first result to establish a PoA-style regret bound in the multi-agent contextual pricing setting. Moreover, we achieve this form of strategy robustness in a black-box manner, by treating any no-regret online expert algorithm as an oracle,

1. Often denoted as Online Posted-price Auction (Kleinberg and Leighton, 2003)

without relying on sophisticated primitives such as differential privacy (Dwork et al., 2014; Liu et al., 2018; Abernethy et al., 2019; Huh and Kandasamy, 2024; Zhang et al., 2021).

Contributions and Organization. In §2, we formally define the multi-buyer online contextual pricing problem, which we refer to as *Online Myersonian Regression* (OMR), in an adaptive, adversarial environment. We also introduce the notion of *strategy-robust regret* (4): the difference between the optimal revenue obtained from truthful buyers and the worst-case revenue obtained when buyers play a Nash equilibrium strategy. This notion captures the concept of the Price of Anarchy from game theory. To the best of our knowledge, this is the first time such a regret notion has been proposed in the context of online auctions. We complement this problem setting with a complexity-theoretic barrier: we show that no-regret learning in OMR is computationally hard even when all buyers are truthful. This result is derived by extending the hardness of Myersonian regression (Liu et al., 2020) to the online setting via a standard online-to-batch conversion (Cesa-Bianchi and Lugosi, 2006).

In §3, we address the first challenge: learning the optimal pricing policy in an adaptive environment, assuming buyers are truthful. The adaptive nature of the environment renders conventional non-adaptive sketching methods inapplicable (Hardt and Woodruff, 2013), such as the Johnson–Lindenstrauss dimensionality reduction (Johnson et al., 1984) used in the batch Myersonian regression setting (Liu et al., 2020) to reduce the search (action) space. Moreover, to extend our result to the strategic buyer setting, where buyers may choose their bids adaptively based on past observations (§4), we require an action-space–reduction technique that can handle adaptive, possibly adversarially chosen inputs. To meet this requirement we introduce a simple and constructive online sketching method in Lemma 3, inspired by Zhang (2002); Rakhlin et al. (2015); Ben-David et al. (2009); Hardt and Rothblum (2010), that operates in this setting.

Building on this technique, we provide a black-box reduction from any no-regret online expert algorithm to a polynomial-time approximation scheme (PTAS) for online learning of linear pricing policies against truthful buyers (Alg. 2). The resulting algorithm achieves $\widetilde{\mathcal{O}}(\sqrt{T})$ regret, up to an additive $O(\epsilon T)$ slack for any fixed $\epsilon > 0$ (Theorem 2), and runs in polynomial time for fixed ϵ. This matches the best possible computational efficiency given the aforementioned hardness barrier.

In §4, we tackle the second challenge: ensuring robustness to strategic behavior in the multi-buyer setting. We introduce a surprisingly simple modification to the previous algorithm that guarantees robustness to any Nash equilibrium formed by strategic buyers. Specifically, we present the *Sparse Update Mechanism* (SUM), formally characterized in Lemma 7. SUM relies on a simple idea: the seller privately and independently tosses coins to decide which rounds are used to update the external online expert algorithm. This limits the buyer's influence, as they do not know which rounds affect learning and thus cannot reliably overfit the learning algorithm. We show that the algorithm, when modified with SUM, achieves a *strategy-robust* regret bound (Theorem 5) with a constant additive approximation. To the best of our knowledge, this is the first online contextual pricing algorithm to provide Price-of-Anarchy-style regret guarantees in multi-buyer settings.

Related Work. We provide here a concise overview of the most relevant prior work. For a more comprehensive discussion, we refer the reader to Appendix F.

A foundational contribution to strategy-robust learning in pricing is due to Amin et al. (2013), who introduced the notion of *strategic regret*, the gap between the seller's revenue under truthful buyers

and that under utility-maximizing, strategic buyers. They proposed an algorithm achieving sublinear strategic regret when buyer valuations are drawn *independently* from an unknown distribution. This work was extended by Mohri and Munoz (2015), who provided guarantees against a broader class of strategic buyers that approximately maximize their utility.

The contextual setting, where the seller observes side information (e.g., user features) at the time of sale, was first studied with strategic buyers by Amin et al. (2014), who introduced strategic regret in this setting. Subsequent work (Deng et al., 2019; Drutsa, 2020; Zhiyanov and Drutsa, 2020; Liu et al., 2024; Golrezaei et al., 2023) built on this framework by refining strategic regret guarantees and extending the model to incorporate additional complexities such as noisy feedback. All of these results operate under a stochastic environment, where buyer values and contexts are drawn from a fixed but unknown distribution.

A separate line of work addresses adversarial environments, where buyer valuations are chosen adaptively. The seminal work of Liu et al. (2018) introduced the first algorithm with no strategic regret in this setting, using differential privacy to limit the influence of strategic buyers. In contrast, our approach achieves a similar effect through a simpler mechanism: the seller randomly selects a round from which to learn and feeds its outcome to an external online expert algorithm. Because our method is agnostic to the internal structure of the expert algorithm, it offers simplicity and broad applicability.

It is worth noting that while strategy-robust learning in pricing frameworks has included multi-buyer settings in seminal works such as Amin et al. (2013) and Liu et al. (2018), the regret guarantees in these works are established against individual buyers who act to maximize their own utility. These behaviors may correspond to specific Nash equilibria, but not necessarily to the entire set of equilibria. In particular, it is unclear whether their algorithmic techniques provide guarantees that hold *uniformly* over all possible Nash equilibria.

2. Preliminaries

We begin by formulating the core problem, *Online Myersonian Regression* (OMR), and then define the notion of an online expert algorithm, which will serve as a black-box component in our algorithms.

Notation. We use $[n] \coloneqq \{1, \dots, n\}$ to denote the set of integers from 1 to n. The Euclidean norm is denoted by $\|\cdot\|$, and the inner product by $\langle v, w \rangle$. The d-dimensional Euclidean unit ball is $\mathbb{B}^d \coloneqq \{x \in \mathbb{R}^d : \|x\| \leq 1\}$. The indicator function is written as $\mathbb{1}[\cdot]$. For a sequence $(x_1, \dots, x_t)$, we write $x_{\leq t} \coloneqq (x_1, \dots, x_t)$. The set of probability distributions over the elements in S is denoted by Δ_S, and U_S denotes the uniform distribution over S. The Bernoulli distribution with success probability p is denoted by $\mathsf{Ber}(p)$.The notation $\mathrm{poly}(n)$ refers to a polynomial function in n.

2.1. Problem Setting: Online Myersonian Regression

The main problem we aim to solve is the online (i.e., sequential) variant of Myersonian regression, introduced by Liu et al. (2020), which we refer to as *Online Myersonian Regression* (OMR). In OMR, a seller $\mathcal{A}$ interacts with a sequence of buyers, selling one good per round. In each round $t \in [T]$, a buyer arrives with a private value $\theta_t \in [0, 1]$ for the good. We assume there are n buyers, indexed by $i \in [n]$, who appear over T rounds. Specifically, each buyer i appears in rounds $t \in \mathcal{T}_i$,

where $\mathcal{T}_i \subseteq [T]$ is a time slot such that $\bigcup_{i\in[n]} \mathcal{T}_i = [T]$ and $\mathcal{T}_i \cap \mathcal{T}_j = \emptyset$ for all $i \neq j$; that is, the collection $\mathcal{T} := (\mathcal{T}_i)_{i\in[n]}$ forms a partition of $[T]$.

At the beginning of each round, the seller is provided with a normalized context (i.e., feature) vector $x_t \in \mathbb{R}^d$ with $\|x_t\| = 1$, which contains additional information relevant to pricing—such as demographic attributes of the buyer or characteristics of the good being sold in that round. The seller uses this context to determine a price $p_t \in [0, 1]$ for the good in round t. In this work, we focus on linear pricing policies, where the price is given by a linear function of the context: $p_t(x_t) = \langle w_t, x_t \rangle$.

The key features of OMR are: (1) the context-value pairs are chosen by an *adaptive adversary*, referred to as the environment $\mathcal{E}$, based on past buyer-seller interactions; and (2) the buyers' values $(\theta_t)_{t\in[T]}$ are not known to the seller *a priori*, but are elicited from the buyers, abstracted as adaptive algorithms $\mathcal{B}_i$, through submitted bids b_t which may differ from their true values. The first feature captures the challenge of learning in a dynamic environment, as is standard in online learning, while the second highlights the difficulty of learning from strategic agents who may manipulate the data for their own benefit. A formal overview of the OMR framework is provided in Protocol 1.

Buyer Algorithm. Each buyer $i \in [n]$ employs a bidding algorithm $\mathcal{B}_i$. Formally, $\mathcal{B}_i$ is a sequence of maps $(\mathcal{B}_{i,t})_{t\in[T]}$, where $\mathcal{B}_{i,t} : \mathcal{I}_i^{t-1} \to \Delta([0, 1])$ is a mapping from buyer i's view of the interaction history up to round $t - 1$ in $\mathcal{I}_i^{t-1}$ (the realizations of all random variables observable to the buyer, as specified below) to a distribution over bids in $[0, 1]$ at round t.

Information Structure. We assume that the seller is oblivious to the buyer algorithms $\mathcal{B}$, the environment $\mathcal{E}$, and the round partition $(\mathcal{T}_i)_{i\in[n]}$. Each party observes the variables they generate, as well as any variables explicitly revealed to them by the protocol (e.g., the seller observes the bids submitted to them). In addition, we assume that the context x_t and the posted price p_t (or equivalently, the weight vector w_t) are observable to all parties: the seller $\mathcal{A}$, the buyers $\mathcal{B}$, and the environment $\mathcal{E}$. The environment $\mathcal{E}$ observes the true values $(\theta_t)_{t\in[T]}$, since it selects them, but *does not observe* the bids $(b_t)_{t\in[T]}$. Crucially, the seller does not observe the buyers' true values, as this limitation is central to the learning problem, whereas each buyer may observe the values of other buyers. We emphasize that the aforementioned assumptions are the weakest needed for our results, and that in practical settings stronger informational assumptions may often hold.

Buyer Utility. We assume that each buyer's total utility is a discounted sum of their per-round utilities. Formally, let $\gamma_i \in [0, \overline{\gamma}]$ denote the utility discount factor of buyer i, where $\overline{\gamma} < 1$ is a known constant. Then, buyer i's total expected utility is defined as

$$U_i(\mathcal{B}; \mathcal{A}, \mathcal{E}, \mathcal{T}) := \mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A},\mathcal{B},\mathcal{E},\mathcal{T})} \left[\sum_{t\in\mathcal{T}_i} \gamma_i^{t-1} u(\theta_t, p_t, b_t) \right], \quad u(\theta, p, b) := (\theta - p) \cdot \mathbb{1}[b \geq p]. \quad (1)$$

Here $\mathbb{E}_{\mathsf{OMR}(\mathcal{A},\mathcal{B},\mathcal{E},\mathcal{T})}$ denotes the expectation over all random variables realized during the execution of $\mathsf{OMR}(\mathcal{A}, \mathcal{B}, \mathcal{E}, \mathcal{T})$, as described in Protocol 1.

One may interpret the discount factor γ_i as controlling the effective planning horizon of a buyer's strategy, that is, how much the buyer values future surplus relative to immediate gains. A smaller γ_i implies more myopic behavior, while values closer to 1 reflect greater long-term planning, where the buyer is willing to incur short-term losses for greater future gains. Notably, in the extreme case of

Protocol 1: $\mathsf{OMR}(\mathcal{A}, \mathcal{B}, \mathcal{E}, \mathcal{T})$: (Multi-buyer) Online Myersonian Regression

Players: Seller's algorithm $\mathcal{A}$ and buyers' algorithms $\mathcal{B} \coloneqq (\mathcal{B}_i)_{i\in[n]}$.
Parameters: Environment $\mathcal{E}$; total number of rounds $T \in \mathbb{N}$; round partition $\mathcal{T} \coloneqq (\mathcal{T}_i)_{i\in[n]}$.
Constraints: T is known to both $\mathcal{A}$ and $\mathcal{B}$; $\mathcal{A}$ is oblivious to $\mathcal{B}$, $\mathcal{E}$ and $\mathcal{T}$.

for $t \in [T]$ **do**

Let $i_t \in [n]$ be the buyer index such that $t \in \mathcal{T}_i$.

The environment $\mathcal{E}$ selects a pair (x_t, θ_t), where $x_t \in \mathbb{R}^d$ is a unit vector representing the public context, and $\theta_t \in [0, 1]$ is the private value of buyer i_t for the good. Simultaneously, $\mathcal{A}$ selects $w_t \in \mathbb{B}^d$, which is kept private.

Then, $\mathcal{B}_{i_t}$ submits a sealed bid b_t to $\mathcal{A}$.

$\mathcal{A}$ posts a price $p_t = \langle w_t, x_t\rangle$, and optionally reveals w_t, which is made public to all players.

The buyer obtains the good at price p_t if and only if $p_t \le b_t$, and this allocation outcome is not observed by other buyers.

end

undiscounted utilities, where $\gamma_i = 1$, it is known that no algorithm can guarantee no-regret learning against a utility-maximizing buyer, even in the single-buyer, non-contextual setting (Amin et al., 2013). For this reason, we restrict our attention to the case $\gamma_i \in [0, \overline{\gamma}]$ for some $\overline{\gamma} < 1$.

Nash Equilibrium. Given the definition of buyer utilities, we now formalize the notion of a Nash equilibrium in our setting, which captures a stable outcome where no buyer can unilaterally improve their utility. Let $\mathcal{B}_{-i} \coloneqq (\mathcal{B}_j)_{j\in[n]\setminus\{i\}}$ denote the strategies of all buyers except buyer i. We say that the strategy profile $\mathcal{B} \coloneqq (\mathcal{B}_i)_{i\in[n]}$ is a *Nash equilibrium* if

$$\forall i \in [n], \quad U_i(\mathcal{B}_i, \mathcal{B}_{-i}; \mathcal{A}, \mathcal{E}, \mathcal{T}) = \sup_{\mathcal{B}_i} U_i(\mathcal{B}_i, \mathcal{B}_{-i}; \mathcal{A}, \mathcal{E}, \mathcal{T}).$$

We denote by $\mathsf{NE}(\mathcal{A}, \mathcal{E}, \mathcal{T}, \gamma)$ the set of all Nash equilibria corresponding to a given seller algorithm $\mathcal{A}$, environment $\mathcal{E}$, round partition $\mathcal{T}$ and discount factors $\gamma \coloneqq (\gamma_i)_{i\in[n]}$.

Revenue and Regrets. The seller's total expected revenue is defined as

$$\mathsf{Rev}(\mathcal{A}; \mathcal{B}, \mathcal{E}, \mathcal{T}) \coloneqq \mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A},\mathcal{B},\mathcal{E},\mathcal{T})}\left[\sum_{t=1}^{T} \mathsf{rev}(w_t, x_t, b_t)\right], \quad \mathsf{rev}(w, x, b) \coloneqq [\langle w, x\rangle]_+ \mathbb{1}[b \ge \langle w, x\rangle], \tag{2}$$

where $[x]_+ \coloneqq \max\{0, x\}$ and $\mathsf{rev}(w_t, x_t, b_t)$ is the revenue collected in a single round. We compare this against the optimal revenue achievable by the best fixed linear pricing policy in hindsight, evaluated with respect to the *true values* $(\theta_t)_{t\in[T]}$ of the buyers:

$$\mathsf{Opt}(\mathcal{A}; \mathcal{B}, \mathcal{E}, \mathcal{T}) \coloneqq \mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A},\mathcal{B},\mathcal{E},\mathcal{T})}\left[\sup_{w\in\mathbb{B}^d} \sum_{t=1}^{T} \mathsf{rev}(w, x_t, \theta_t)\right].$$

The difference between the above two quantities, commonly referred to as regret, depends on the buyers' strategies $\mathcal{B}$. We define two refined notions of regret, capturing performance guarantees under different behavioral assumptions on $\mathcal{B}$.

The first is the *standard regret*, where all buyers act truthfully by bidding their true values (i.e., $b_t = \theta_t$ for all t). Let $\mathcal{B}_{\text{tru}}$ denote this truthful bidding strategy profile. Then, the standard regret is

$$\mathsf{Reg}(\mathcal{A}; \mathcal{E}, \mathcal{T}) := \mathsf{Opt}(\mathcal{A}; \mathcal{B}_{\text{tru}}, \mathcal{E}, \mathcal{T}) - \mathsf{Rev}(\mathcal{A}; \mathcal{B}_{\text{tru}}, \mathcal{E}, \mathcal{T}). \tag{3}$$

Our main focus is on the *strategic regret*, which captures the worst-case efficiency loss due to strategic buyer behavior, as formalized by Nash equilibrium. It is defined as

$$\mathsf{SReg}(\mathcal{A}; \mathcal{E}, \mathcal{T}, \gamma) := \sup_{\mathcal{B} \in \mathsf{NE}(\mathcal{A}, \mathcal{E}, \mathcal{T}, \gamma)} \Big[\mathsf{Opt}(\mathcal{A}; \mathcal{B}, \mathcal{E}, \mathcal{T}) - \mathsf{Rev}(\mathcal{A}; \mathcal{B}, \mathcal{E}, \mathcal{T}) \Big]. \tag{4}$$

In contrast to the standard regret (3), the above is a worst-case regret evaluated over *all possible* Nash equilibria. As such, it closely aligns with the notion of *Price-of-Anarchy*, as it quantifies the worst-case degradation in performance (revenue) due to selfish (strategic) behavior at equilibrium.

Hardness of OMR. We conclude this section by discussing the intrinsic computational hardness of OMR, which is inherited from the hardness of Myersonian regression in the batch setting (Liu et al., 2020). This hardness extends to the online setting via a standard online-to-batch conversion (Cesa-Bianchi and Lugosi, 2006), and is captured by the following theorem (for completeness, we include its proof in Appendix A):

Theorem 1 [Adapted from Liu et al. (2018)] *Suppose the Exponential Time Hypothesis (Impagliazzo and Paturi, 2001) holds. Then, no algorithm $\mathcal{A}$ for the seller in OMR can attain sublinear regret, i.e.,* $\sup_{\mathcal{E},\mathcal{T}} \mathsf{Reg}(\mathcal{A}; \mathcal{E}, \mathcal{T}) \in \mathcal{O}(T^z)$ *for some $z \in [0, 1)$, while running in worst-case time* $\mathrm{poly}(d, T)$.

As this complexity-theoretic barrier rules out computationally efficient no-regret algorithms for OMR, we instead aim to design algorithms that satisfy an approximate notion of no-regret:

Definition 1 *A seller's algorithm $\mathcal{A}$ achieves* ϵ-approximate no-regret *if* $\mathsf{Reg}(\mathcal{A}; \mathcal{E}, T) \in \mathcal{O}(T^z + \epsilon T)$ *for some fixed $\epsilon > 0$ and $z \in [0, 1)$. The notion of* ϵ-approximate no-strategic-regret *is defined analogously by replacing* Reg *with* SReg. *Moreover, when an ϵ-approximate no-regret algorithm runs in time polynomial in T and d for fixed ϵ, we refer to it as a* polynomial-time approximation scheme *(PTAS).*

2.2. Online Expert Algorithms

Our algorithms treat any online expert algorithm as a black-box oracle. For completeness, we briefly recall the standard definition of online expert learning and its associated regret (see (Cesa-Bianchi and Lugosi, 2006; Hazan et al., 2016; Bubeck et al., 2015)).

The goal of an online expert algorithm is to compete with the best among K experts, indexed by $z \in [K]$, when reward functions $r_t : [K] \to [0, 1]$ are chosen adaptively by an environment based on past interactions. Formally, an (anytime) expert algorithm is a mapping $\mathcal{L} : \bigcup_{t \in \mathbb{N}} \left([0, 1]^K\right)^t \to \Delta_{[K]}$, where the argument represents the sequence of past reward functions. At each round $t \in [T]$,

the algorithm selects $z_t \sim \mathcal{L}(T, r_{\le t-1})$, where $r_{\le t-1} \coloneqq (r_\tau)_{\tau \in [t-1]}$ denotes the history of rewards up to round $t-1$. Then, the regret of K-expert algorithm $\mathcal{L}$ over T rounds against environment $\mathcal{E}$ is defined as

$$\mathsf{Reg}^{\mathrm{exp}}(\mathcal{L}; \mathcal{E}) \coloneqq \mathbb{E}\left[\sup_{z \in [K]} \sum_{t=1}^{T} \big(r_t(z) - r_t(z_t)\big)\right], \tag{5}$$

which quantifies the expected difference between the cumulative reward of the best fixed expert in hindsight and that obtained by the algorithm. Common examples of online expert algorithms include Hedge (Freund and Schapire, 1997) and Follow-the-Perturbed-Leader (FTPL) (Kalai and Vempala, 2005; Hutter et al., 2005).

3. Online Myersonian Regression with Truthful Buyers

To build toward the goal of strategy-robust learning, we begin by studying the OMR problem under the simplifying assumption that buyers are truthful—that is, they report their true values in every round, so $b_t = v_t$ for all $t \in [T]$. This setting isolates the core learning challenge in OMR, which remains nontrivial even in the absence of strategic behavior. Our main contribution in this section is a black-box reduction from any no-regret online expert algorithm $\mathcal{L}$ to an OMR algorithm $\mathcal{A}^{\mathcal{L}}_{\mathsf{OMR}}$, which translates the expert regret guarantee (5) into a standard regret guarantee (3). We later extend this result to strategic settings in §4.

Key technical challenge. While buyers act truthfully, the environment may still be adversarially and adaptively selecting context–value pairs (x_t, v_t) based on the seller's past actions. This adaptivity poses a significant obstacle to efficient learning, as it renders standard dimensionality reduction techniques, such as Johnson-Lindenstrauss sketching (Johnson et al., 1984) or CountSketch (Charikar et al., 2004), ineffective, since they rely on fixed or non-adaptive input distributions (Hardt and Woodruff, 2013; Cohen et al., 2022). Designing tools that remain effective under such adversarial feedback is therefore a central challenge in our setting.

Algorithm Overview. We begin by describing the algorithm, and then outline the core intuitions in a proof sketch of our main result, Theorem 5. Our algorithm $\mathcal{A}^{\mathcal{L}}_{\mathsf{OMR}}$ (Alg. 2) uses oracle access to an online expert algorithm $\mathcal{L}$. The expert algorithm operates over a carefully constructed set of experts, denoted $\mathcal{Z}_{\epsilon,T}$ and defined in (6), whose size is at most $T^{\mathrm{poly}(1/\epsilon)}$. This is polynomial in T for fixed ϵ, in contrast to the exponential dependence on the weight dimension d that arises when directly searching over a discretization of $\mathbb{B}^d$.

At each round t, $\mathcal{A}^{\mathcal{L}}_{\mathsf{OMR}}$ invokes the expert selected by $\mathcal{L}$, denoted $z_t \in \mathcal{Z}_{\epsilon,T}$, and maps it to a pricing weight vector $w_t = v_t(z_t; x_{\le t}) \in \mathbb{B}^d$ using a deterministic function v_t, defined in (7), that depends only on z_t and the context history $x_{\le t} \coloneqq (x_1, \dots, x_t)$. After the interaction, $\mathcal{A}^{\mathcal{L}}_{\mathsf{OMR}}$ feeds back to $\mathcal{L}$ the reward function $\mathsf{rev}\big(v_t(\,\cdot\,; x_{\le t}), x_t, b_t\big)$. The sketch set $\mathcal{Z}_{\epsilon,T}$ is defined as:

$$\mathcal{Z}_{\epsilon,T} \coloneqq \left\{(\beta_\tau)_{\tau \in S} : S \subseteq [T],\ |S| \le 16\epsilon^{-2},\ \forall \tau \in S,\ \beta_\tau \in [-2, 2] \cap \frac{\epsilon^2}{8}\mathbb{Z}\right\}, \tag{6}$$

That is, an element $z \in \mathcal{Z}_{\epsilon,T}$ is an ordered collection $(\beta_\tau)_{\tau \in S}$ for some $S \subseteq [T]$ where each β_τ lies in a $\epsilon^2/8$-spaced grid in $[-2, 2]$. Enumerating over choices of S and $(\beta_\tau)_{\tau \in S}$, we observe $|\mathcal{Z}_{\epsilon,T}| \le T^{\mathcal{O}(1/\epsilon^2)}$, a polynomial in T for a fixed ϵ.

Algorithm 2: $\mathcal{A}^{\mathcal{L}}_{\mathsf{OMR}}$: Online Myersonian Regression Algorithm for Truthful Buyers

Oracle: Online Expert Algorithm $\mathcal{L}$.

Input: Total rounds T; approximation parameter $\epsilon > 0$.

Let $\mathcal{Z}_{\epsilon,T}$ be the set defined in (6), and set it as the expert set of $\mathcal{L}$.

for $t \in [T]$ **do**

Sample $z_t \sim \mathcal{L}(r_{\le t-1})$.

$w_t \leftarrow v_t(z_t; x_{\le t})$ and post the price $p_t \leftarrow \langle w_t, x_t \rangle$.

Set $r_t(z) \leftarrow \mathsf{rev}(v_t(z; x_{\le t}), x_t, b_t)$, for each $z \in \mathcal{Z}_{\epsilon,T}$

▷ rev and v_t are defined in (2) and (7) respectively.

end

Then, for each $x_{\le t}$, we define $v_t(\,\cdot\,; x_{\le t}) : \mathcal{Z}_{\epsilon,T} \to \mathbb{B}^d$ as follows:

$$v_t(z; x_{\le t}) := \frac{\sum_{\tau \in S \cap [t]} \beta_\tau x_\tau}{\max\left\{1, \left\|\sum_{\tau \in S \cap [t]} \beta_\tau x_\tau\right\|\right\}} \in \mathbb{B}^d. \tag{7}$$

Intuitively, for a given sketch z and past contexts $x_{\le t}$, v_t reconstructs a vector in a low-dimensional subspace of $\mathrm{span}(x_1, \ldots, x_t)$. The following theorem formalizes the regret guarantee of $\mathcal{A}^{\mathcal{L}}_{\mathsf{OMR}}$ in terms of the regret of the expert algorithm $\mathcal{L}$:

Theorem 2 *Let $\mathcal{L}$ be an online expert algorithm whose regret over K experts is bounded by $R(T \log K)$ for some function R. Then, for any approximation parameter $\epsilon \in [0, 1/2]$, time horizon T, and environment $\mathcal{E}$, we have*

$$\mathsf{Reg}(\mathcal{A}^{\mathcal{L}}_{\mathsf{OMR}}; \mathcal{E}, \mathcal{T}) \le R\big(\mathcal{O}(\epsilon^{-2}) \cdot T \log T\big) + 4\epsilon T.$$

In particular, when $\mathcal{L}$ is the Hedge algorithm, the regret is at most $\mathcal{O}\big(\epsilon^{-1}\sqrt{T \log T} + \epsilon T\big)$, and $\mathcal{A}^{\mathcal{L}}_{\mathsf{OMR}}$ runs in time $\mathcal{O}\big(T^{\mathcal{O}(1/\epsilon^2)} \cdot \mathrm{poly}(T)\big)$, hence we obtain a PTAS for OMR with truthful buyers.

Proof of Theorem 2. The core ingredient behind Theorem 2 is an *online sketching technique* that reduces the effective search space over linear pricing weights $w_t \in \mathbb{B}^d$, even when contexts are chosen adversarially. Specifically, we show that any $w \in \mathbb{B}^d$ can be compactly represented by some $z \in \mathcal{Z}_{\epsilon,T}$ such that, for every $t \in [T]$, the vector $v_t(z; x_{\le t})$ closely approximates the inner product $\langle w, x_t \rangle$. This guarantee is formalized in Lemma 3, which is the main technical result in this section:

Lemma 3 [Online Sketch] *For any sequence $(x_t)_{t \in [T]}$ of context vectors in $\mathbb{B}^d$ and any $w \in \mathbb{B}^d$, there exists $z \in \mathcal{Z}_{\epsilon,T}$ such that*

$$\forall t \in [T], \quad \big|\langle v_t(z; x_{\le t}), x_t \rangle - \langle w, x_t \rangle\big| \le \epsilon^2,$$

where $\mathcal{Z}_{\epsilon,T}$ and v_t are defined in (6) and (7), respectively.

Given the above approximation guarantee, a straightforward argument shows that the optimal revenue computed over the sketch set $\mathcal{Z}_{\epsilon,T}$ closely approximates the optimal revenue achievable by linear pricing. This is formalized in the following lemma, which is proved in Appendix C:

Lemma 4 *For any* $(x_t, b_t)_{t\in[T]}$ *and* $\epsilon \in [0, 1/2]$,

$$\sup_{w\in\mathbb{B}^d} \sum_{t=1}^{T} \mathsf{rev}(w, x_t, b_t) \le \sup_{z\in\mathcal{Z}_{\epsilon,T}} \left[\sum_{t=1}^{T} \mathsf{rev}\big(v_t(z; x_{\le t}), x_t, b_t\big) \right] + 4\epsilon T.$$

By the assumed regret guarantee of $\mathcal{L}$, running $\mathcal{L}$ on the sequence of reward functions $r_t(\,\cdot\,) \coloneqq \mathsf{rev}\big(v_t(\,\cdot\,; x_{\le t}), x_t, b_t\big)$ ensures that

$$\mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{OMR}}, \mathcal{B}_{\mathsf{tru}}, \mathcal{E}, \mathcal{T})} \left[\sup_{z\in[K]} \sum_{t=1}^{T} \big(r_t(z) - r_t(z_t)\big) \right] \le R\big(T \log |\mathcal{Z}_{\epsilon,T}|\big).$$

Combining this with Lemma 4, and noting that $|\mathcal{Z}_{\epsilon,T}| \le T^{\mathcal{O}(1/\epsilon^2)}$, yields the regret bound. ■

Proof Sketch of Lemma 3. We conclude this section with a proof sketch of Lemma 3 and defer the full proof to Appendix B.1. Fix $w \in \mathbb{B}^d$ in the lemma statement. For each $t \in [T]$, we explicitly construct v_t such that $\langle v_t, x_t\rangle$ approximates $\langle w, x_t\rangle$ as follows:

1. Let $u, v \in \mathbb{B}^d$ be the zero vector, that is, $u, v \leftarrow 0$.
2. For each $t \in [T]$, do the following: If $|\,\langle v, x_t\rangle - \langle w, x_t\rangle\,| > \epsilon^2/2$, repeatedly update u and v via
$$u \leftarrow u - (\epsilon^2/8)\,\mathsf{sign}(\langle v, x_t\rangle - \langle w, x_t\rangle)x_t, \quad v \leftarrow \frac{u}{\max\{1, \|u\|\}}, \tag{8}$$
until the condition $|\,\langle v, x_t\rangle - \langle w, x_t\rangle\,| \le \epsilon^2/2$ is satisfied. Then, set $v_t \leftarrow v$.

By construction, assuming the loop in Step 2 terminates, we have $|\,\langle v_t, x_t\rangle - \langle w, x_t\rangle\,| \le \epsilon^2/2$ for all t. Hence, once termination is established, this approximation guarantee holds uniformly over all t.

Observe that the update rule in (8) performs a gradient descent step followed by a projection onto $\mathbb{B}^d$. Thus, this loop effectively implements the Lazy Online Gradient Descent (Lazy OGD) (Zinkevich, 2003) with learning rate $\beta \coloneqq \epsilon^2/8$, applied to the convex loss functions $f_t(v) \coloneqq |\,\langle v, x_t\rangle - \langle w, x_t\rangle\,|$. Each time an update is triggered, the incurred loss satisfies $f_t(v) \ge \epsilon^2/2$, since we only update when the approximation error exceeds this threshold. Using the regret bound for Lazy OGD, which scales as $(2\beta)^{-1} + 2\beta M$ over M updates, we obtain:

$$\frac{\epsilon^2 M}{2} \le \frac{4}{\epsilon^2} + \frac{\epsilon^2 M}{4} \implies M \in \mathcal{O}(\epsilon^{-2}).$$

Therefore, we conclude that $|\,\langle v_t, x_t\rangle - \langle w, x_t\rangle\,| \le \epsilon^2/2$ for all t. Furthermore, since the total number of updates is at most $M = \mathcal{O}(\epsilon^{-2})$, each v_t is a linear combination of at most $\mathcal{O}(\epsilon^{-2})$ context vectors from $(x_1, \ldots, x_t)$, and the normalization factor changes at most $\mathcal{O}(\epsilon^{-2})$ times throughout the procedure. With suitable discretization, such linear combinations can be encoded compactly in the form $v_t(z; x_{\le t})$ for some $z \in \mathcal{Z}_{\epsilon,T}$, which completes the proof sketch. ■

4. Strategy-robust Regret Guarantee

In this section, we introduce a simple modification to $\mathcal{A}^{\mathcal{L}}_{\mathsf{OMR}}$, called the *Sparse Update Mechanism (SUM)*, which makes the algorithm strategy-robust. When the oracle $\mathcal{L}$ is no-regret, the resulting algorithm, given in Alg. 3, achieves ϵ-approximately no-strategic-regret (Def. 1).

Algorithm 3: $\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}$: Strategy-robust Online Myersonian Regression via Sparse Update

Oracle: Online Expert Algorithm $\mathcal{L}$.

Input: Total rounds T; approximation parameter $\epsilon > 0$; the upper bound $\overline{\gamma}$ on discount factors.

Let $\mathcal{Z}_{\epsilon,T}$ be the set defined in (6), and set it as the expert set of $\mathcal{L}$.

Sample $\xi_1, \dots, \xi_T \overset{iid}{\sim} \mathsf{Ber}(\rho)$ where $\rho \coloneqq \min\{1, \overline{\gamma}^{-1}(1-\overline{\gamma})\epsilon^5/3\}$.

for $t \in [T]$ **do**

  Sample $\omega \sim \mathsf{Ber}(\epsilon)$.

  if $\omega = 1$ **then**

    Sample $\lambda \sim U_{[0,1]}$ and $w_t \leftarrow \lambda x_t$. ▷ Price uniformly at random.

  end

  else

    Sample $z_t \sim \mathcal{L}\big((\xi_\tau r_\tau)_{\tau\in[t-1]}\big)$ and $w_t \leftarrow v_t(z_t; x_{\le t})$.

  end

  Post the price $p_t \leftarrow \langle w_t, x_t\rangle$.

  Set $r_t(z) \leftarrow \mathsf{rev}(v_t(z; x_{\le t}), x_t, b_t)$, for each $z \in \mathcal{Z}_{\epsilon,T}$

end

Algorithm Overview. The key idea behind SUM is to limit buyer influence by randomly selecting—privately and independently—which rounds are used to update the seller's learning algorithm. Specifically, in each round, the seller flips a biased private coin ξ_t to decide whether to feed the observed interaction back to the expert algorithm $\mathcal{L}$. As a result, only a sparse, random subset of rounds contributes to learning, making it difficult for any buyer to reliably influence future prices.

In addition, the algorithm occasionally selects a price uniformly at random by choosing $w_t = \lambda x_t$ with $\lambda \sim U_{[0,1]}$. As shown in Lemma 6, this random pricing imposes a strictly positive expected cost on any buyer who bids significantly away from their true value. Combined with the limited influence induced by the randomized sparse updating, this ensures that all Nash equilibria are *approximately truthful*, as formalized in the key Lemma 7: In every equilibrium, buyers bid close to true values.

This approach of combining randomized updates and random pricing is conceptually simpler than prior work based on differential privacy (Liu et al., 2018; Abernethy et al., 2019), and it supports a black-box reduction from any no-regret online expert algorithm. The regret guarantee for Alg. 3 is given below:

Theorem 5 *Let $\mathcal{L}$ be an online expert algorithm whose regret over K experts is bounded by $R(T\log K)$ for some concave function R. Then, for any approximation parameter $\epsilon \in [0, 1/4]$, time horizon T, and environment $\mathcal{E}$, we have*

$$\mathsf{SReg}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}; \mathcal{E}, \mathcal{T}, \gamma) \le \mathcal{O}\left(\frac{\overline{\gamma}}{\epsilon^5(1-\overline{\gamma})}\, R\left(\frac{\epsilon^5(1-\overline{\gamma})\,T\log T}{\overline{\gamma}}\right) + \sqrt{\frac{\overline{\gamma}\,T\log T}{\epsilon^7(1-\overline{\gamma})}} + \epsilon T\right).$$

In particular, if $\mathcal{L}$ is the Hedge algorithm, then the strategic regret satisfies

$$\mathsf{SReg}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}; \mathcal{E}, \mathcal{T}, \gamma) \in \mathcal{O}\left(\epsilon^{-3.5}\sqrt{\frac{\overline{\gamma}\, T \log T}{1-\overline{\gamma}}} + \epsilon T\right),$$

and $\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}$ runs in time $\mathcal{O}\big(T^{\mathcal{O}(1/\epsilon^2)} \cdot \mathrm{poly}(T)\big)$, leading to a PTAS for OMR with strategic buyers.

Proof sketch of Theorem 5 (A full proof is given in Appendix D). From the buyer's perspective, their input (i.e., bid) influences future outcomes with probability strictly less than one. By adjusting the update (i.e., feedback) probability, the seller can control the degree of influence a buyer has: the smaller the update probability, the less impact any single bid has on future pricing decisions. When this probability is sufficiently small, the buyer's expected gain from misreporting in any given round becomes small, yet non-zero.

To eliminate even this small expected gain, SUM introduces *random pricing* with a small probability $\mathcal{O}(\epsilon)$, under which the price is occasionally chosen independently of the bid. This imposes a non-zero cost for misreporting, as formally observed in the following lemma proved in Appendix C:

Lemma 6 *Let θ_t and b_t denote the true value and the bid at round t, respectively. Fixing all other random variables, a pricing policy of the form $w_t = \lambda x_t$, where $\lambda \sim U_{[0,1]}$, induces a utility loss, averaged over λ, of at least $\frac{1}{2}(\theta_t - b_t)^2$ compared to the utility from a truthful bid.*

This random pricing mechanism ensures that bidding far from the true value incurs a non-zero expected utility loss. By appropriately calibrating the update probability, we can guarantee a non-zero expected cost for any bid that deviates from the true value by more than some threshold $\delta > 0$. This leads to the conclusion that, in any Nash equilibrium, all buyer strategies must remain within δ of truthful bidding. This insight is formalized in the following key lemma of this section

Lemma 7 [Sparse Update Mechanism] *For any Nash equilibrium $\mathcal{B} \in \mathsf{NE}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{E}, \mathcal{T}, \gamma)$,*

$$\Pr_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}, \mathcal{E}, \mathcal{T})}\left[\, \forall t \in [T],\ |\theta_t - b_t| \le \sqrt{\frac{3\rho\overline{\gamma}}{\epsilon(1-\overline{\gamma})}} \,\right] = 1.$$

The above lemma, proved in Appendix B.2 implies that, with sufficiently small ρ, the bids remain close to the true values. Consequently, the optimal revenue computed from these near-truthful bids is close to the true optimal revenue under truthful bidding, as formalized in the following lemma:

Lemma 8 *For any $(x_t, \theta_t, b_t)_{t\in[T]}$ and $\delta \in [0, 1/4]$ such that $|\theta_t - b_t| \le \delta$ for all $t \in [T]$,*

$$\sup_{w\in\mathbb{B}^d} \sum_{t=1}^{T} \mathsf{rev}(w, x_t, \theta_t) \le \sup_{w\in\mathbb{B}^d} \left[\sum_{t=1}^{T} \mathsf{rev}(w, x_t, b_t)\right] + 2\sqrt{\delta}T.$$

The final ingredient is to observe that the regret increase due to sparse updates is modest, incurring only a constant-factor slowdown, as formally stated in Lemma 10. Specifically, the lemma ensures that the regret remains asymptotically equivalent to the truthful case with respect to T. Combining all the results, we obtain the desired guarantee, completing the proof. ■

5. Conclusion

In this work, we studied online contextual pricing in the presence of multiple strategic buyers, introducing a new notion of *strategy-robust regret* that captures efficiency under worst-case behavior, akin to the *Price-of-Anarchy*. We proposed *Online Myersonian Regression (OMR)*, a generalization of linear contextual pricing in the batch setting, and showed that achieving no-regret learning in this model is computationally hard, even when buyers are truthful. To address this challenge, we developed a *polynomial-time approximation scheme (PTAS)* for truthful buyers, based on a novel online sketching method that enables efficient learning in *adaptive adversarial* environments.

To extend robustness to multi-buyer strategic settings, we introduced the *Sparse Update Mechanism (SUM)*—a simple randomization technique in which the seller updates its pricing policy using feedback from only a sparse subset of rounds. This effectively limits buyer influence and ensures regret guarantees that hold uniformly over *all Nash equilibria*. Our approach is conceptually simple and broadly applicable, as it is based on a black-box reduction from any no-regret expert algorithm. We believe this framework can be further extended to yield no-regret guarantees under more refined benchmarks, such as dynamic or adaptive regret Hazan et al. (2016); Cesa-Bianchi and Lugosi (2006).

References

Jacob D Abernethy, Rachel Cummings, Bhuvesh Kumar, Sam Taggart, and Jamie H Morgenstern. Learning auctions with robust incentive guarantees. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, 2019.

Marek Adamczyk, Allan Borodin, Diodato Ferraioli, Bart de Keijzer, and Stefano Leonardi. Sequential posted price mechanisms with correlated valuations. In *Proceedings of the 11th International Conference on Web and Internet Economics*, pages 1–15. Springer, 2015.

Kareem Amin, Afshin Rostamizadeh, and Umar Syed. Learning prices for repeated auctions with strategic buyers. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, 2013.

Kareem Amin, Afshin Rostamizadeh, and Umar Syed. Repeated contextual auctions with strategic buyers. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, 2014.

Moshe Babaioff, Shaddin Dughmi, Robert Kleinberg, and Aleksandrs Slivkins. Dynamic pricing with limited supply. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 74–91. ACM, 2012.

Gah-Yi Ban and N Bora Keskin. Personalized dynamic pricing with machine learning: High-dimensional features and heterogeneous elasticity. *Management Science*, 67(9):5549–5568, 2021.

Shai Ben-David, Dávid Pál, and Shai Shalev-Shwartz. Agnostic online learning. In *Proceedings of the 22nd Annual Conference on Learning Theory*, volume 3, 2009.

Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420, 2009.

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

Sébastien Bubeck et al. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding frequent items in data streams. *Theoretical Computer Science*, 312(1):3–15, 2004.

Ningyuan Chen and Guillermo Gallego. A primal–dual learning algorithm for personalized dynamic pricing with an inventory constraint. *Mathematics of Operations Research*, 47(4):2585–2613, 2022.

Christine Chung, Katrina Ligett, Kirk Pruhs, and Aaron Roth. The price of stochastic anarchy. In *Proceedings of the 1st International Symposium on Algorithmic Game Theory*, pages 303–314. Springer, 2008.

Edith Cohen, Xin Lyu, Jelani Nelson, Tamás Sarlós, Moshe Shechner, and Uri Stemmer. On the robustness of countsketch to adaptive inputs. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pages 4112–4140. JMLR.org, 2022.

Sriram Dasu and Chunyang Tong. Dynamic pricing when consumers are strategic: Analysis of posted and contingent pricing schemes. *European Journal of Operational Research*, 204(3):662–671, 2010.

Arnoud V Den Boer. Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys in Operations Research and Management Science*, 20(1):1–18, 2015.

Arnoud V den Boer and N Bora Keskin. Dynamic pricing with demand learning and reference effects. *Management Science*, 68(10):7112–7130, 2022.

Yuan Deng, Sébastien Lahaie, and Vahab Mirrokni. A robust non-clairvoyant dynamic mechanism for contextual auctions. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, 2019.

Alexey Drutsa. Optimal non-parametric learning in repeated contextual auctions with strategic buyer. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 2668–2677. JMLR.org, 2020.

Pradeep Dubey. Inefficiency of Nash equilibria. *Mathematics of Operations Research*, 11(1):1–8, 1986.

Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

Guillermo Gallego and Garrett Van Ryzin. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40(8):999–1020, 1994.

Negin Golrezaei, Patrick Jaillet, and Jason Cheuk Nam Liang. Incentive-aware contextual pricing with non-parametric market noise. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130, pages 9331–9361. JMLR.org, 2023.

Moritz Hardt and Guy N Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *Proceedings of 51st Annual IEEE Symposium on Foundations of Computer Science*, pages 61–70. IEEE, 2010.

Moritz Hardt and David P Woodruff. How robust are linear sketches to adaptive inputs? In *Proceedings of the forty-fifth annual ACM Symposium on Theory of Computing*, pages 121–130. ACM, 2013.

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

Joon Suk Huh and Kirthevasan Kandasamy. Nash incentive-compatible online mechanism learning via weakly differentially private online learning. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pages 20643–20659. JMLR.org, 2024.

Marcus Hutter, Jan Poland, and Manfred Warmuth. Adaptive online prediction by following the perturbed leader. *Journal of Machine Learning Research*, 6(4), 2005.

Russell Impagliazzo and Ramamohan Paturi. On the complexity of k-SAT. *Journal of Computer and System Sciences*, 62(2):367–375, 2001.

Adel Javanmard and Hamid Nazerzadeh. Dynamic pricing in high-dimensions. *Journal of Machine Learning Research*, 20(9):1–49, 2019.

Hossein Nekouyan Jazi, Bo Sun, Raouf Boutaba, and Xiaoqi Tan. Posted price mechanisms for online allocation with diseconomies of scale. In *Proceedings of the ACM on Web Conference 2025*, pages 2710–2728. ACM, 2025.

William B Johnson, Joram Lindenstrauss, et al. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, 26(189-206):1, 1984.

Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Proceedings of 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 594–605. IEEE, 2003.

Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Proceedings of the 16th Annual Conference on Theoretical Aspects of Computer Science*, pages 404–413. Springer, 1999.

Renato Paes Leme, Balasubramanian Sivan, Yifeng Teng, and Pratik Worah. Learning to price against a moving target. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 6223–6232. JMLR.org, 2021.

Kyle Y Lin. Dynamic pricing with real-time demand learning. *European Journal of Operational Research*, 174(1):522–538, 2006.

Allen Liu, Renato Leme, and Jon Schneider. Myersonian regression. In *Advances in Neural Information Processing Systems*, volume 33, pages 9135–9144. Curran Associates, 2020.

Jinyan Liu, Zhiyi Huang, and Xiangning Wang. Learning optimal reserve price against non-myopic bidders. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, 2018.

Pangpang Liu, Zhuoran Yang, Zhaoran Wang, and Will Wei Sun. Contextual dynamic pricing with strategic buyers. *Journal of the American Statistical Association*, pages 1–13, 2024.

Yiyun Luo, Will Wei Sun, and Yufeng Liu. Distribution-free contextual dynamic pricing. *Mathematics of Operations Research*, 49(1):599–618, 2024.

Mehryar Mohri and Andres Munoz Medina. Learning theory and algorithms for revenue optimization in second price auctions with reserve. In *Proceedings of the 31st International Conference on International Conference on Machine Learning*, volume 32, pages 262–270. JMLR.org, 2014.

Mehryar Mohri and Andres Munoz. Revenue optimization against strategic buyers. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, 2015.

Andres Munoz and Sergei Vassilvitskii. Revenue optimization with approximate bid predictions. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, 2017.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning via sequential complexities. *Journal of Machine Learning Research*, 16:155–186, 2015.

Giulia Romano, Gianluca Tartaglia, Alberto Marchesi, and Nicola Gatti. Online posted pricing with unknown time-discounted valuations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 5682–5689. AAAI, 2021.

Andreas S Schulz and Nicolás Stier Moses. On the performance of user equilibria in traffic networks. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 86–87. SIAM, 2003.

Weiran Shen, Sébastien Lahaie, and Renato Paes Leme. Learning to clear the market. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 5710–5718. JMLR.org, 2019.

Wei Tang, Chien-Ju Ho, and Yang Liu. Differentially private contextual dynamic pricing. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1368–1376. IFAAMAS, 2020.

Matilde Tullii, Solenne Gaucher, Nadav Merlis, and Vianney Perchet. Improved algorithms for contextual dynamic pricing. In *Advances in Neural Information Processing Systems*, volume 37, pages 126088–126117. Curran Associates, 2024.

Arsenii Vanunts and Alexey Drutsa. Optimal pricing in repeated posted-price auctions with different patience of the seller and the buyer. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, 2019.

Nan Yang and Renyu Zhang. Dynamic pricing and inventory management under inventory-dependent demand. *Operations Research*, 62(5):1077–1094, 2014.

Lefeng Zhang, Tianqing Zhu, Ping Xiong, Wanlei Zhou, and Philip S Yu. More than privacy: Adopting differential privacy in game-theoretic mechanism design. *ACM Computing Surveys (CSUR)*, 54(7):1–37, 2021.

Tong Zhang. Covering number bounds of certain regularized linear function classes. *Journal of Machine Learning Research*, 2:527–550, 2002.

Anton Zhiyanov and Alexey Drutsa. Bisection-based pricing for repeated contextual auctions against strategic buyer. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 11469–11480. JMLR.org, 2020.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning*, pages 928–936. JMLR.org, 2003.

Appendix A. Proof of Hardness for No-Regret Learning in OMR

In this section, we prove the computational hardness of no-regret learning in Online Myersonian Regression (OMR). Our argument builds on the hardness result from Liu et al. (2020), extended to the online setting via the standard online-to-batch conversion (Cesa-Bianchi and Lugosi, 2006).[2]

Proposition 1 (Hardness of Myersonian Regression (Liu et al., 2020)) *Assuming the Exponential Time Hypothesis (Impagliazzo and Paturi, 2001), any (possibly randomized) algorithm that approximates*

$$\mathsf{Opt}\big((x_i,\theta_i)_{i\in[N]}\big) \coloneqq \sup_{w\in\mathbb{B}^d} \frac{1}{N}\sum_{i=1}^{N} \mathsf{rev}(w,x_i,\theta_i),$$

within an additive error of ϵ, for any input sequence $(x_i,\theta_i)_{i\in[N]}$ with $x_i\in\mathbb{B}^d$ and $\theta_i\in[0,1]$, must have worst-case running time at least $\mathcal{O}\big(2^{\Omega(\epsilon^{-1/6})}\cdot\mathrm{poly}(d,N)\big)$.

This theorem rules out a Fully Polynomial-Time Approximation Scheme (FPTAS) for the (batch) Myersonian regression problem. We now extend this hardness result to rule out efficient no-regret algorithms for OMR, as stated below:

2. Note that the Myersonian regression setting in Liu et al. (2020) assumes contexts lie in the unit ball, $x_t\in\mathbb{B}^d$, but their result also applies to normalized contexts with $x_t\in\mathbb{B}^d$.

Theorem 1 [Adapted from Liu et al. (2018)] *Suppose the Exponential Time Hypothesis (Impagliazzo and Paturi, 2001) holds. Then, no algorithm $\mathcal{A}$ for the seller in OMR can attain sublinear regret, i.e.,* $\sup_{\mathcal{E},\mathcal{T}} \mathsf{Reg}(\mathcal{A};\mathcal{E},\mathcal{T}) \in \mathcal{O}(T^z)$ *for some* $z \in [0,1)$*, while running in worst-case time* $\mathrm{poly}(d,T)$.

Proof of Theorem 1. Suppose there exists an algorithm $\mathcal{A}$ such that $\sup_{\mathcal{E},\mathcal{T}} \mathsf{Reg}(\mathcal{A};\mathcal{E},\mathcal{T}) \in \mathcal{O}(T^z)$ for some $z \in [0,1)$, and $\mathcal{A}$ runs in worst-case time $\mathrm{poly}(d,T)$. We show that, using $\mathcal{A}$, one can approximate

$$\mathsf{Opt}\big((x_i,\theta_i)_{i\in[N]}\big) := \sup_{w\in\mathbb{B}^d} \sum_{i=1}^{N} \mathsf{rev}(w,x_i,\theta_i), \tag{9}$$

within an additive error of ϵ, for any input sequence $(x_i,\theta_i)_{i\in[N]}$ with $x_i \in \mathbb{B}^d$ and $\theta_i \in [0,1]$, in total time $\mathrm{poly}(1/\epsilon,d,N)$, thereby contradicting Theorem 1.

The algorithm proceeds as follows: given an input sequence $(x_i,\theta_i)_{i=1}^N$, run $\mathcal{A}$ for T rounds, where in each round the context–value pair (x_t,θ_t) is sampled uniformly at random from $(x_i,\theta_i)_{i=1}^N$. Let $(w_t)_{t=1}^T$ denote the sequence of outputs produced by $\mathcal{A}$. Then, by the high-probability online-to-batch conversion bound (Cesa-Bianchi and Lugosi, 2006), we have

$$\mathsf{Opt}\big((x_t,\theta_t)_{t\in[T]}\big) \le \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\mathsf{rev}(w_t,x_i,\theta_i)\right) + \mathcal{O}(T^{z-1})$$

with probability at least $9/10$. Moreover, since $w_t \in \mathbb{B}^d$, it holds that

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\mathsf{rev}(w_t,x_i,\theta_i)\right) \le \mathsf{Opt}\big((x_t,\theta_t)_{t\in[T]}\big).$$

Together, these bounds imply that the quantity $\frac{1}{TN}\sum_{t,i}\mathsf{rev}(w_t,x_i,\theta_i)$ approximates $\mathsf{Opt}\big((x_t,\theta_t)_t\big)$ within an additive error of $\mathcal{O}(T^{z-1})$, which can be made at most ϵ by choosing $T = \mathrm{poly}(1/\epsilon)$. Since $\mathcal{A}$ runs in time $\mathrm{poly}(d,T)$, the overall procedure runs in time $\mathrm{poly}(1/\epsilon,d,N)$, completing the proof. ■

Appendix B. Proof of Key Lemmas

In this section, we prove two key lemmas: Lemma 3, which analyzes our online sketching technique used to establish the PTAS for OMR with truthful buyers, and Lemma 7, which shows the approximate truthfulness of *any Nash equilibrium* induced by $\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}$ (Alg. 3).

B.1. Proof of Lemma 3

Preliminary. Before proving Lemma 3, we state a technical result that we use: a regret bound for Lazy Online Gradient Descent (Lazy OGD), given in Lemma 9. This bound is a special case of the regret guarantee for Lazy Online Mirror Descent when applied with the Bregman divergence $\frac{1}{2}\|x-y\|^2$. For completeness, we provide a direct proof in Appendix E.

Lemma 9 *Let $M \in \mathbb{N}$, and for each $i \in [M]$, let $f_i : \mathbb{B}^d \to \mathbb{R}$ be any convex loss function with unit-bounded subgradients, i.e., $\|\nabla f_i\| \le 1$. Then, for any $\beta > 0$ and $v^\star \in \mathbb{B}^d$,*

$$\sum_{i\in[M]} \big(f_i(v_i) - f_i(v^\star)\big) \le (2\beta)^{-1} + 2\beta M, \tag{10}$$

where v_i is defined recursively as

$$u_{i+1} \coloneqq u_i - \beta\,\nabla f_i(v_i), \quad u_1 = 0 \quad \text{and} \quad v_i \coloneqq \frac{u_i}{\max\{1, \|u_i\|\}}. \tag{11}$$

With this ingredient, we now prove Lemma 3, restated below for convenience.

Lemma 3 [Online Sketch] *For any sequence $(x_t)_{t\in[T]}$ of context vectors in $\mathbb{B}^d$ and any $w \in \mathbb{B}^d$, there exists $z \in \mathcal{Z}_{\epsilon,T}$ such that*

$$\forall t \in [T], \quad \big|\big\langle v_t(z; x_{\le t}), x_t\big\rangle - \big\langle w, x_t\big\rangle\big| \le \epsilon^2,$$

where $\mathcal{Z}_{\epsilon,T}$ and v_t are defined in (6) and (7), respectively.

Proof of Lemma 3 Fix $(x_t)_{t\in[T]}$ and $w \in \mathbb{B}^d$. For each $t \in [T]$, define $h_t : \mathbb{B}^d \to \mathbb{R}$ as

$$h_t(v) \coloneqq \langle v, x_t\rangle - \langle w, x_t\rangle .$$

We prove the lemma by an explicit construction using the following procedure:

1. Let $u, v \in \mathbb{B}^d$ be the zero vector, that is, $u, v \leftarrow 0$.
2. For each $t \in [T]$, do the following: If $|h_t(v)| > \epsilon^2/2$, repeatedly update u and v as

$$u \leftarrow u - (\epsilon^2/8)\,\mathsf{sign}(h_t(v))x_t, \quad v \leftarrow \frac{u}{\max\{1, \|u\|\}}. \tag{12}$$

   until $|h_t(v)| \le \epsilon^2/2$. Then, let $v_t \leftarrow v$.

First, we show that the above procedure, specifically the repetitive update step until $|h_t(v)| \le \epsilon^2/2$, terminates after finitely many steps. Whenever $|h_t(v)| > \epsilon^2/2$ occurs, we call this event a "mistake." Define f_i as $f_i(v) \coloneqq |h_t(v)|$, where $i \in \mathbb{N}$ indexes the mistakes that occur during the execution of the procedure. Moreover, let v_i and u_i denote the values of v and u at the beginning of the update step corresponding to the i-th mistake (i.e., right before the update is applied).

Then, the update procedure in (12) can be rewritten as $u_{i+1} \leftarrow u_i - (\epsilon^2/8) f_i(v_i)$ and $v_{i+1} \leftarrow u_{i+1}/\max\{1, \|u_{i+1}\|\}$, with initial values $u_1 = v_1 = 0$. This is exactly the lazy OGD update rule given in (11), as stated in Lemma 9, with loss functions $(f_i)_i$ and step size $\beta = \epsilon^2/8$.

Now, suppose there are a total of M mistakes. Since each f_i is convex and has sub-normalized subgradient, as $\|\nabla f_i(v)\| = \|x_t\| \le 1$, we can apply the lazy OGD regret bound (10) from Lemma 9 to obtain the following: For any $v^\star \in \mathbb{B}^d$,

$$\sum_{i\in[M]} \big(f_i(v_i) - f_i(v^\star)\big) \le \frac{4}{\epsilon^2} + \frac{\epsilon^2 M}{4}.$$

In particular, choosing $v^\star = w$ yields $f_i(v^\star) = 0$ for all $i \in [M]$. Also, by definition, $f_i(v_i) > \epsilon^2/2$. Combining these, we obtain

$$\frac{\epsilon^2 M}{2} \leq \frac{4}{\epsilon^2} + \frac{\epsilon^2 M}{4},$$

which implies $M \leq 16/\epsilon^2$.

Let $S \subseteq [T]$ be the rounds when v_t[3] was updated in Step 2. Unrolling the update steps, v_t can be rewritten as

$$\boxed{\forall t \in [T], \quad v_t = \frac{\sum_{\tau \in S \cap [t]} \beta_\tau x_\tau}{\max\left\{1, \left\|\sum_{\tau \in S \cap [t]} \beta_\tau x_\tau\right\|\right\}} \in \mathbb{B}^d}$$

for some fixed $(\beta_\tau)_{\tau \in S}$ with $\beta_\tau \in \frac{\epsilon^2}{8}\mathbb{Z}$ and $|S| = M \leq \epsilon^2/8$. Since the number of updates is at most $16/\epsilon^2$ and the learning rate is $\epsilon^2/8$, we have $|\beta_\tau| \leq M \cdot \epsilon^2/8 \leq 2$, hence $\beta_\tau \in [-2, 2] \cap \frac{\epsilon^2}{8}\mathbb{Z}$. Therefore, the above v_t can be written as $v_t(z; x_{\leq t})$ where $z := (\beta_\tau)_{\tau \in S} \in \mathcal{Z}_{\epsilon,T}$.

Since $|h_t(x_t)| = |\langle v_t, x_t\rangle - \langle w, x_t\rangle| \leq \epsilon^2/2$ for all $t \in [T]$ by the construction at the beginning, this completes the proof of the lemma. ■

B.2. Proof of Lemma 7

In this section, we prove Lemma 7, the core result from §3, which states that buyers report their values *approximately truthfully in any Nash equilibrium*. For convenience, we restate Lemma 7 below.

Lemma 7 [Sparse Update Mechanism] *For any Nash equilibrium* $\mathcal{B} \in \mathsf{NE}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{E}, \mathcal{T}, \gamma)$,

$$\Pr_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}, \mathcal{E}, \mathcal{T})}\left[\forall t \in [T],\ |\theta_t - b_t| \leq \sqrt{\frac{3\rho\overline{\gamma}}{\epsilon(1-\overline{\gamma})}}\right] = 1.$$

Proof of Lemma 7. Let $\mathcal{B} := (\mathcal{B}_i)_{i \in [n]}$ be a Nash equilibrium; that is, $\mathcal{B} \in \mathsf{NE}(\mathcal{A}^{L}_{\mathsf{SUM}}, \mathcal{E}, \mathcal{T}, \gamma)$. Suppose there exists a buyer i who misreports their value by at least $\delta := \sqrt{\frac{3\rho\overline{\gamma}}{\epsilon(1-\overline{\gamma})}}$ at some round $t^\star$, with non-zero probability. That is, there exist $i \in [n]$ and $t^\star \in [T]$ such that

$$\Pr_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}, \mathcal{E}, \mathcal{T})}\Big[\,|\theta_{t^\star} - b_{t^\star}| > \delta\,\Big] > 0.$$

Modified Strategy. Now consider a modified strategy $\mathcal{B}_i'$, which behaves identically to $\mathcal{B}_i$ on any input, including internal randomness, except that it bids truthfully at round $t^\star$—that is, it outputs $b_{t^\star} = \theta_{t^\star}$ whenever $\mathcal{B}_i$ would have produced a bid satisfying $|\theta_{t^\star} - b_{t^\star}| > \delta$. We will show that the new strategy profile $\mathcal{B}'$, obtained by replacing $\mathcal{B}_i$ with $\mathcal{B}_i'$, yields strictly higher utility for buyer i. This contradicts the assumption that $\mathcal{B}$ is a Nash equilibrium.

3. v_t is defined in the second step of the procedure; distinct from v_i defined above.

Coupled Protocols. To see this, define a coupled interaction protocol $\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}; \mathcal{B}', \mathcal{E}, \mathcal{T})$ that simultaneously executes the protocols $\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}, \mathcal{E}, \mathcal{T})$ and $\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}', \mathcal{E}, \mathcal{T})$ using shared randomness. That is, all internal randomness of the seller algorithm and of the buyers is shared across both runs. The probability distribution induced by this coupled protocol, denoted $\Pr_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}; \mathcal{B}', \mathcal{E}, \mathcal{T})}$, forms a coupling of the distributions $\Pr_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}, \mathcal{E}, \mathcal{T})}$ and $\Pr_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}', \mathcal{E}, \mathcal{T})}$; in particular, these two distributions are marginals of the joint distribution defined by the coupled execution. For clarity, we use primed (unprimed) variables to refer to random variables arising from the interaction with $\mathcal{B}'$ (resp. $\mathcal{B}$).

Utility Decomposition. We now systematically compare the utility of buyer i under two strategies: the original strategy $\mathcal{B}_i$ and the modified strategy $\mathcal{B}'_i$. The ex-post utility of buyer i (i.e., before taking the expectation) under strategy $\mathcal{B}_i$ can be decomposed as

$$\sum_{t \in \mathcal{T}_i} \gamma_i^{t-1} u(\theta_t, p_t, b_t) = \sum_{t \in \mathcal{T}_i \cap [t^\star - 1]} \gamma^{t-1} u(\theta_t, p_t, b_t) + \gamma^{t^\star - 1} u(\theta_{t^\star}, p_{t^\star}, b_{t^\star}) + \sum_{t \in \mathcal{T}_i \cap [t^\star + 1 : T]} \gamma^{t-1} u(\theta_t, p_t, b_t),$$

and the utility under $\mathcal{B}'_i$ (evaluated at the primed variables) is decomposed in the same way. Since $\mathcal{B}'_i$ only differs from $\mathcal{B}_i$ at round $t^\star$, we have

$$\sum_{t \in \mathcal{T}_i \cap [t^\star - 1]} \gamma^t u(\theta_t, p_t, b_t) = \sum_{t \in \mathcal{T}_i \cap [t^\star - 1]} \gamma^t u(\theta'_t, p'_t, b'_t).$$

with probability one. Therefore, the difference in utility between the two strategies is

$$\sum_{t \in \mathcal{T}_i} \gamma_i^{t-1} u(\theta_t, p_t, b_t) - \sum_{t \in \mathcal{T}_i} \gamma_i^{t-1} u(\theta'_t, p'_t, b'_t) = \gamma_i^{t^\star - 1} \Big(u(\theta_{t^\star}, p_{t^\star}, b_{t^\star}) - u(\theta'_{t^\star}, p'_{t^\star}, b'_{t^\star}) \Big) + \sum_{t \in \mathcal{T}_i \cap [t^\star + 1 : T]} \gamma_i^{t-1} \Big(u(\theta_t, p_t, b_t) - u(\theta'_t, p'_t, b'_t) \Big).$$

Let $\mathcal{G}$ denote the event that $\mathcal{B}_i$ submits a bid $b_{t^\star}$ such that $|\theta_{t^\star} - b_{t^\star}| > \delta$. By construction, the modified strategy $\mathcal{B}'_i$ differs from $\mathcal{B}_i$ only when this event occurs. Thus, unless $\mathcal{E}$ occurs, the values of the primed and unprimed variables are identical, and there is no utility gap. Therefore, the difference in utilities can further be written as

$$\sum_{t \in \mathcal{T}_i} \gamma_i^{t-1} u(\theta_t, p_t, b_t) - \sum_{t \in \mathcal{T}_i} \gamma_i^{t-1} u(\theta'_t, p'_t, b'_t)$$
$$= \mathbb{1}[\mathcal{G}] \cdot \left(\gamma_i^{t^\star - 1} \Big(u(\theta_{t^\star}, p_{t^\star}, b_{t^\star}) - u(\theta'_{t^\star}, p'_{t^\star}, b'_{t^\star}) \Big) + \sum_{t \in \mathcal{T}_i \cap [t^\star + 1 : T]} \gamma_i^{t-1} \Big(u(\theta_t, p_t, b_t) - u(\theta'_t, p'_t, b'_t) \Big) \right).$$

Effect of Sparse Update. Let $\mathcal{U}$ be the event that $\xi_{t^\star} = 1$, indicating that the online expert algorithm $\mathcal{L}$ receives feedback from round $t^\star$. From the information structure described in §2, we know that future variables (i.e., for $t > t^\star$) can be affected by buyer i's strategy modification only if $\mathcal{U}$ occurs. This is because, when $\mathcal{U}$ does not occur, there is no channel through which the strategy

modification at round $t^\star$ can influence the future: (1) the environment $\mathcal{E}$ does not observe actual bids, so future context–bid pairs remain unchanged by the modification; and (2) other buyers do not observe buyer i's bid, so their decisions are unaffected as well. Therefore, the second term above is non-zero only when the event $\mathcal{G} \cap \mathcal{U}$ occurs, leading to

$$\begin{aligned}
&\sum_{t\in\mathcal{T}_i} \gamma_i^{t-1} u(\theta_t, p_t, b_t) - \sum_{t\in\mathcal{T}_i} \gamma_i^{t-1} u(\theta'_t, p'_t, b'_t) \\
&\qquad = \mathbb{1}[\mathcal{G}] \cdot \gamma_i^{t^\star-1} \Big(u(\theta_{t^\star}, p_{t^\star}, b_{t^\star}) - u(\theta'_{t^\star}, p'_{t^\star}, b'_{t^\star}) \Big) \\
&\qquad\qquad\qquad + \mathbb{1}[\mathcal{G} \cap \mathcal{U}] \cdot \sum_{t\in\mathcal{T}_i\cap[t^\star+1:T]} \gamma_i^{t-1} \Big(u(\theta_t, p_t, b_t) - u(\theta'_t, p'_t, b'_t) \Big) \\
&\qquad \le \mathbb{1}[\mathcal{G}] \cdot \gamma_i^{t^\star-1} \Big(u(\theta_{t^\star}, p_{t^\star}, b_{t^\star}) - u(\theta'_{t^\star}, p'_{t^\star}, b'_{t^\star}) \Big) + \mathbb{1}[\mathcal{G} \cap \mathcal{U}] \cdot \frac{\gamma_i^{t^\star}}{1-\gamma_i},
\end{aligned}$$

where the last line follows since $u(\cdot) \in [0,1]$ and by the geometric series bound.

Variables at $t^\star$. Notice that the buyer's value $(\theta_{t^\star}, \theta'_{t^\star})$ and the prices $(p_{t^\star}, p'_{t^\star})$ at round $t^\star$ are determined by the variables $(\theta_t, \theta'_t, p_t, p'_t, b_t, b'_t)$ from rounds up to $t^\star - 1$, along with the contexts (x_t, x'_t) up to round $t^\star$. Since $\mathcal{B}'_i$ behaves identically to $\mathcal{B}_i$ up to round $t^\star - 1$, and the context–value pair at each round is selected by an adaptive adversary based only on observations from previous rounds, we have $x_{t^\star} = x'_{t^\star}$ and $\theta_{t^\star} = \theta'_{t^\star}$, and thus $p_{t^\star} = p'_{t^\star}$. Finally, by construction, $\mathcal{B}'_i$ bids truthfully at round $t^\star$ when $\mathcal{G}$ occurs, so $b'_{t^\star} = \theta'_{t^\star} = \theta_{t^\star}$. Combining all these facts, we conclude that $p_{t^\star} = p'_{t^\star}$ and $b'_{t^\star} = \theta'_{t^\star} = \theta_{t^\star}$, hence

$$\begin{aligned}
&\sum_{t\in\mathcal{T}_i} \gamma_i^{t-1} u(\theta_t, p_t, b_t) - \sum_{t\in\mathcal{T}_i} \gamma_i^{t-1} u(\theta'_t, p'_t, b'_t) \\
&\qquad\qquad \le \mathbb{1}[\mathcal{G}] \cdot \gamma_i^{t^\star-1} \Big(u(\theta_{t^\star}, p_{t^\star}, b_{t^\star}) - u(\theta_{t^\star}, p_{t^\star}, \theta_{t^\star}) \Big) + \mathbb{1}[\mathcal{G} \cap \mathcal{U}] \cdot \frac{\gamma_i^{t^\star}}{1-\gamma_i}. \qquad (13)
\end{aligned}$$

Conditioning over Random Pricing. Recall that random pricing is invoked when $\omega = 1$, where $\omega \sim \mathsf{Ber}(\epsilon)$ is tossed independently at each round. Let $\mathcal{R}$ denote the event that $\omega = 1$ occurs at round $t^\star$, meaning the seller uses random pricing at that round. First, observe that regardless of whether $\mathcal{G}$ or $\mathcal{R}$ occurs, we always have the following inequality:

$$u(\theta_{t^\star}, p_{t^\star}, \theta_{t^\star}) = (\theta_{t^\star} - p_{t^\star}) \mathbb{1}[\theta_{t^\star} \ge p_{t^\star}] \ge (\theta_{t^\star} - p_{t^\star}) \mathbb{1}[b_{t^\star} \ge p_{t^\star}] = u(\theta_{t^\star}, p_{t^\star}, b_{t^\star}),$$

which follows directly from the dominant-strategy incentive compatibility (DSIC) of the posted-price auction, or through a simple case-by-case verification.

Therefore, we can further bound the first term of (13) by conditioning on $\mathcal{R}$:

$$\begin{aligned}
&\sum_{t\in\mathcal{T}_i} \gamma_i^{t-1} u(\theta_t, p_t, b_t) - \sum_{t\in\mathcal{T}_i} \gamma_i^{t-1} u(\theta'_t, p'_t, b'_t) \\
&\qquad\qquad \le \mathbb{1}[\mathcal{G} \cap \mathcal{R}] \cdot \gamma_i^{t^\star-1} \big(u(\theta_{t^\star}, p_{t^\star}, b_{t^\star}) - u(\theta_{t^\star}, p_{t^\star}, \theta_{t^\star}) \big) + \mathbb{1}[\mathcal{G} \cap \mathcal{U}] \cdot \frac{\gamma_i^{t^\star}}{1-\gamma_i}.
\end{aligned}$$

Taking the Expectation. Now, we take the conditional expectation $\mathbb{E}[\cdot\,|\mathcal{G}]$ over random variables in $\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}; \mathcal{B}', \mathcal{E}, \mathcal{T})$, conditioned on the event $\mathcal{G}$, where $|\theta_{t^\star} - b_{t^\star}| > \delta$. Since (1) $\mathcal{G}$ occurs with non-zero probability, (2) the events $\mathcal{R}$ and $\mathcal{U}$ are independent of $\mathcal{G}$ and occur with probability ϵ and ρ, respectively, and (3) when $\mathcal{R}$ occurs, the price is sampled independently, allowing us to invoke Lemma 6, it follows that

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}},\mathcal{B};\mathcal{B}',\mathcal{E},\mathcal{T})}\left[\sum_{t\in\mathcal{T}_i}\gamma_i^{t-1}u(\theta_t,p_t,b_t)-\sum_{t\in\mathcal{T}_i}\gamma_i^{t-1}u(\theta'_t,p'_t,b'_t)\,\middle|\,\mathcal{G}\right]&\\ \le -\frac{\epsilon\cdot\gamma_i^{t^\star-1}}{2}(\theta_{t^\star}-b_{t^\star})^2+\frac{\rho\cdot\gamma_i^{t^\star}}{1-\gamma_i}&\\ \le -\frac{\epsilon\cdot\delta^2\cdot\gamma_i^{t^\star-1}}{2}+\frac{\rho\cdot\gamma_i^{t^\star}}{1-\gamma_i},&\end{aligned}$$

where the last line follows as $|\theta_{t^\star} - b_{t^\star}| > \delta$ under $\mathcal{E}$. Then, by taking the full expectation over all randomness, the above implies

$$\mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}},\mathcal{B};\mathcal{B}',\mathcal{E},\mathcal{T})}\left[\sum_{t\in\mathcal{T}_i}\gamma_i^{t-1}u(\theta_t,p_t,b_t)-\sum_{t\in\mathcal{T}_i}\gamma_i^{t-1}u(\theta'_t,p'_t,b'_t)\right]\le -\frac{\epsilon\cdot\delta^2\cdot\gamma_i^{t^\star-1}}{2}+\frac{\rho\cdot\gamma_i^{t^\star}}{1-\gamma_i}.$$

Since $\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}; \mathcal{B}', \mathcal{E}, \mathcal{T})$ is a coupling of $\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}, \mathcal{E}, \mathcal{T})$ and $\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}', \mathcal{E}, \mathcal{T})$, we can take marginal expectations and separate the terms, as each involves only primed or unprimed variables, respectively. Therefore, the LHS of the above expression is

$$\underbrace{\mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}},\mathcal{B},\mathcal{E},\mathcal{T})}\left[\sum_{t\in\mathcal{T}_i}\gamma_i^{t-1}u(\theta_t,p_t,b_t)\right]}_{=U_i(\mathcal{B}_i,\mathcal{B}_{-i};\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}},\mathcal{E},\mathcal{T})}-\underbrace{\mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}},\mathcal{B}',\mathcal{E},\mathcal{T})}\left[\sum_{t\in\mathcal{T}_i}\gamma_i^{t-1}u(\theta'_t,p'_t,b'_t)\right]}_{=U_i(\mathcal{B}'_i,\mathcal{B}_{-i};\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}},\mathcal{E},\mathcal{T})}.$$

Hence,

$$U_i(\mathcal{B}_i,\mathcal{B}_{-i};\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}},\mathcal{E},\mathcal{T})-U_i(\mathcal{B}'_i,\mathcal{B}_{-i};\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}},\mathcal{E},\mathcal{T})\le -\frac{\epsilon\cdot\delta^2\cdot\gamma_i^{t^\star-1}}{2}+\frac{\rho\cdot\gamma_i^{t^\star}}{1-\gamma_i}.$$

Checking the Negativity. Finally, we verify that the LHS of the above expression is negative, implying that $\mathcal{B} \coloneqq (\mathcal{B}_i)_{i\in[n]}$ is not a Nash equilibrium, and thus proving the lemma. This follows from the choice of $\delta \coloneqq \sqrt{\frac{3\rho\overline{\gamma}}{\epsilon(1-\overline{\gamma})}}$,

$$\begin{aligned}-\frac{\epsilon\cdot\delta^2\cdot\gamma_i^{t^\star-1}}{2}+\frac{\rho\cdot\gamma_i^{t^\star}}{1-\gamma_i}&=\gamma_i^{t^\star-1}\left(-\frac{\epsilon\cdot\delta^2}{2}+\frac{\rho\cdot\gamma_i}{1-\gamma_i}\right)\\ &\le\gamma_i^{t^\star-1}\left(-\frac{\epsilon\cdot\delta^2}{2}+\frac{\rho\cdot\overline{\gamma}}{1-\overline{\gamma}}\right)\\ &=\gamma_i^{t^\star-1}\left(-\frac{3\rho\cdot\overline{\gamma}}{2(1-\overline{\gamma})}+\frac{\rho\cdot\overline{\gamma}}{1-\overline{\gamma}}\right)<0.\end{aligned}$$

■

Appendix C. Miscellaneous Proofs

In this section, we provide proofs of technical lemmas used in the main results.

Lemma 4 *For any* $(x_t, b_t)_{t\in[T]}$ *and* $\epsilon \in [0, 1/2]$,

$$\sup_{w\in\mathbb{B}^d} \sum_{t=1}^{T} \mathsf{rev}(w, x_t, b_t) \le \sup_{z\in\mathcal{Z}_{\epsilon,T}} \left[\sum_{t=1}^{T} \mathsf{rev}\big(v_t(z; x_{\le t}), x_t, b_t\big)\right] + 4\epsilon T.$$

Proof of Lemma 4. Fix any $w \in \mathbb{B}^d$ and define $w' \coloneqq (1-\epsilon)w \in \mathbb{B}^d$. Then,

$$\begin{aligned}
\sum_{t=1}^{T} \mathsf{rev}(w, x_t, b_t) &= \sum_{t=1}^{T} [\langle w, x_t\rangle]_+ \mathbb{1}[b_t \ge \langle w, x_t\rangle] \\
&= (1-\epsilon)^{-1} \sum_{t=1}^{T} \left[\left\langle w', x_t\right\rangle\right]_+ \mathbb{1}[b_t \ge \langle w, x_t\rangle] \\
&\le \sum_{t=1}^{T} \left[\left\langle w', x_t\right\rangle\right]_+ \mathbb{1}[b_t \ge \langle w, x_t\rangle] + 2\epsilon T,
\end{aligned}$$

since $(1-\epsilon)^{-1} \le 1 + 2\epsilon$ for $\epsilon \in [0, 1/2]$ and $\left|\left[\left\langle w', x_t\right\rangle\right]_+ \mathbb{1}[b_t \ge \langle w, x_t\rangle]\right| \le 1$.

Let $S \subseteq [T]$ be the set of rounds t such that $\langle w', x_t\rangle \ge \epsilon$. Then,

$$\begin{aligned}
\sum_{t=1}^{T} \mathsf{rev}(w, x_t, b_t) &\le \sum_{t\in S} \left[\left\langle w', x_t\right\rangle\right]_+ \mathbb{1}[b_t \ge \langle w, x_t\rangle] + \sum_{t\in S^c} \left[\left\langle w', x_t\right\rangle\right]_+ \mathbb{1}[b_t \ge \langle w, x_t\rangle] + 2\epsilon T \\
&\le \sum_{t\in S} \left[\left\langle w', x_t\right\rangle\right]_+ \mathbb{1}[b_t \ge \langle w, x_t\rangle] + 3\epsilon T, \qquad (14)
\end{aligned}$$

since $\left|\left[\left\langle w', x_t\right\rangle\right]_+ \mathbb{1}[b_t \ge \langle w, x_t\rangle]\right| \le \epsilon$ for all $t \in S^c$.

By Lemma 3, there exists $z \in \mathcal{Z}_{\epsilon,T}$ such that

$$\forall t \in S, \quad \left|\left\langle v_t(z; x_{\le t}), x_t\right\rangle - \left\langle w', x_t\right\rangle\right| \le \epsilon^2 \le \epsilon \left\langle w', x_t\right\rangle, \qquad (15)$$

as for $t \in S$, $\langle w', x_t\rangle \ge \epsilon$. The above implies

$$\forall t \in S, \quad \left\langle v_t(z; x_{\le t}), x_t\right\rangle \le (1+\epsilon)\left\langle w', x_t\right\rangle = (1-\epsilon^2)\left\langle w, x_t\right\rangle \le \left\langle w, x_t\right\rangle,$$

where the last inequality holds as $\langle w, x_t\rangle = (1-\epsilon)^{-1}\langle w', x_t\rangle \ge \epsilon > 0$ for $t \in S$. Therefore,

$$\forall t \in S, \quad \mathbb{1}\left[b_t \ge \left\langle w, x_t\right\rangle\right] \le \mathbb{1}\left[b_t \ge \left\langle v_t(z; x_{\le t}), x_t\right\rangle\right] \qquad (16)$$

Combining these steps, we obtain:

$$\sum_{t=1}^{T} \mathsf{rev}(w, x_t, b_t) \le \sum_{t\in S} \left[\left\langle w', x_t\right\rangle\right]_+ \mathbb{1}[b_t \ge \langle w, x_t\rangle] + 3\epsilon T \qquad \text{by (14)}$$

$$
\begin{aligned}
&\leq \sum_{t\in S} \big[\langle v_t(z; x_{\leq t}), x_t\rangle\big]_+ \mathbb{1}\big[b_t \geq \langle w, x_t\rangle\big] + \epsilon^2 T + 3\epsilon T && \text{by (15)}\\
&\leq \sum_{t\in S} \underbrace{\big[\langle v_t(z; x_{\leq t}), x_t\rangle\big]_+ \mathbb{1}\big[b_t \geq \langle v_t(z; x_{\leq t}), x_t\rangle\big]}_{:= \mathsf{rev}\left(v_t(z; x_{\leq t}), x_t, b_t\right)} + \epsilon^2 T + 3\epsilon T && \text{by (16)}\\
&\leq \sum_{t=1}^{T} \mathsf{rev}\left(v_t(z; x_{\leq t}), x_t, b_t\right) + 4\epsilon T,
\end{aligned}
$$

which proves the lemma. ■

Lemma 8 *For any $(x_t, \theta_t, b_t)_{t\in[T]}$ and $\delta \in [0, 1/4]$ such that $|\theta_t - b_t| \leq \delta$ for all $t \in [T]$,*

$$
\sup_{w\in\mathbb{B}^d} \sum_{t=1}^{T} \mathsf{rev}(w, x_t, \theta_t) \leq \sup_{w\in\mathbb{B}^d} \left[\sum_{t=1}^{T} \mathsf{rev}(w, x_t, b_t)\right] + 2\sqrt{\delta}T.
$$

Proof of Lemma 8. Fix any pricing vector $w \in \mathbb{B}^d$, and define $w' := (1 - \sqrt{\delta})w \in \mathbb{B}^d$. First, consider the case where $\theta_t \geq \sqrt{\delta}$. Then,

$$
\begin{aligned}
\mathsf{rev}(w, x_t, \theta_t) &= \big[\,\langle w, x_t\rangle\,\big]_+ \cdot \mathbb{1}[\theta_t \geq \langle w, x_t\rangle]\\
&= \frac{1}{1-\sqrt{\delta}} \big[\,\langle w', x_t\rangle\,\big]_+ \cdot \mathbb{1}\Big[\big(1-\sqrt{\delta}\big)\theta_t \geq \langle w', x_t\rangle\Big]\\
&\leq \frac{1}{1-\sqrt{\delta}} \big[\,\langle w', x_t\rangle\,\big]_+ \cdot \mathbb{1}[\theta_t - \delta \geq \langle w', x_t\rangle], && \text{(as } \big(1-\sqrt{\delta}\big)\theta_t \leq \theta_t - \delta\text{)}\\
&\leq \frac{1}{1-\sqrt{\delta}} \underbrace{\big[\,\langle w', x_t\rangle\,\big]_+ \cdot \mathbb{1}[b_t \geq \langle w', x_t\rangle]}_{= \mathsf{rev}(w', x_t, b_t)}, && \text{(as } \theta_t - \delta \leq b_t\text{)}\\
&\leq \mathsf{rev}(w', x_t, b_t) + 2\sqrt{\delta}, && \text{(as } (1-x)^{-1} \leq 1 + 2x,\ \forall x \in \big[0, \tfrac{1}{2}\big]\text{)}
\end{aligned}
$$

Now, consider the case where $b_t < \sqrt{\delta}$. In this case, we have

$$
\mathsf{rev}(w, x_t, \theta_t) \leq \sqrt{\delta} \leq \mathsf{rev}(w', x_t, b_t) + 2\sqrt{\delta}.
$$

Putting both cases together, we conclude that for any $w \in \mathbb{B}^d$, there exists $w' \in \mathbb{B}^d$ such that

$$
\sum_{t=1}^{T} \mathsf{rev}(w, x_t, \theta_t) \leq \sum_{t=1}^{T} \mathsf{rev}(w', x_t, b_t) + 2\sqrt{\delta}T,
$$

which proves the lemma. ■

Lemma 6 *Let θ_t and b_t denote the true value and the bid at round t, respectively. Fixing all other random variables, a pricing policy of the form $w_t = \lambda x_t$, where $\lambda \sim U_{[0,1]}$, induces a utility loss, averaged over λ, of at least $\frac{1}{2}(\theta_t - b_t)^2$ compared to the utility from a truthful bid.*

Proof of Lemma 6. Since $\|x_t\| = 1$, for $w_t = \lambda x_t$ with $\lambda \sim U_{[0,1]}$, we have $p_t \sim U_{[0,1]}$, as the price is given by $\langle w_t, x_t \rangle = \lambda \|x_t\|^2 = \lambda$. Then,

$$\int_0^1 (\theta_t - p_t) \mathbb{1}[p_t \le \theta_t] \, dp_t - \int_0^1 (\theta_t - p_t) \mathbb{1}[p_t \le b_t] \, dp = \frac{1}{2}(\theta_t - b_t)^2.$$

The first term corresponds to the utility from a truthful bid, and the second to that from a non-truthful bid, both averaged over the random price p_t. ■

Appendix D. Proof of Theorem 5

In this section, we present the full proof of Theorem 5, restated below for convenience:

Theorem 5 *Let $\mathcal{L}$ be an online expert algorithm whose regret over K experts is bounded by $R(T \log K)$ for some concave function R. Then, for any approximation parameter $\epsilon \in [0, 1/4]$, time horizon T, and environment $\mathcal{E}$, we have*

$$\mathsf{SReg}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}; \mathcal{E}, \mathcal{T}, \gamma) \le \mathcal{O}\left(\frac{\overline{\gamma}}{\epsilon^5(1-\overline{\gamma})} \, R\left(\frac{\epsilon^5(1-\overline{\gamma}) \, T \log T}{\overline{\gamma}} \right) + \sqrt{\frac{\overline{\gamma} \, T \log T}{\epsilon^7(1-\overline{\gamma})}} + \epsilon T \right).$$

In particular, if $\mathcal{L}$ is the Hedge algorithm, then the strategic regret satisfies

$$\mathsf{SReg}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}; \mathcal{E}, \mathcal{T}, \gamma) \in \mathcal{O}\left(\epsilon^{-3.5} \sqrt{\frac{\overline{\gamma} \, T \log T}{1-\overline{\gamma}}} + \epsilon T \right),$$

and $\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}$ runs in time $\mathcal{O}\big(T^{\mathcal{O}(1/\epsilon^2)} \cdot \mathrm{poly}(T)\big)$, leading to a PTAS for OMR with strategic buyers.

Proof of Theorem 5. Recalling the definition of strategic regret, we have

$$\mathsf{SReg}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}; \mathcal{E}, \mathcal{T}, \gamma) \coloneqq \sup_{\mathcal{B} \in \mathsf{NE}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{E}, \mathcal{T}, \gamma)} \Big[\mathsf{Opt}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}; \mathcal{B}, \mathcal{E}, \mathcal{T}) - \mathsf{Rev}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}; \mathcal{B}, \mathcal{E}, \mathcal{T}) \Big],$$

where

$$\mathsf{Opt}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}; \mathcal{B}, \mathcal{E}, \mathcal{T}) \coloneqq \mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}, \mathcal{E}, \mathcal{T})} \left[\sup_{w \in \mathbb{B}^d} \sum_{t=1}^T \mathsf{rev}(w, x_t, \theta_t) \right]$$

$$\mathsf{Rev}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}; \mathcal{B}, \mathcal{E}, \mathcal{T}) \coloneqq \mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}, \mathcal{E}, \mathcal{T})} \left[\sum_{t=1}^T \mathsf{rev}(w_t, x_t, b_t) \right].$$

Expanding $\mathsf{SReg}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}; \mathcal{E}, \mathcal{T}, \gamma)$, we get

$$\begin{aligned} &\mathsf{SReg}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}; \mathcal{E}, \mathcal{T}, \gamma) \\ &\quad = \sup_{\mathcal{B} \in \mathsf{NE}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{E}, \mathcal{T}, \gamma)} \mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}, \mathcal{E}, \mathcal{T})} \left[\sup_{w \in \mathbb{B}^d} \sum_{t=1}^T \mathsf{rev}(w, x_t, \theta_t) - \sum_{t=1}^T \mathsf{rev}(w_t, x_t, b_t) \right] \end{aligned}$$

Let $\delta := \sqrt{\frac{3\rho\overline{\gamma}}{\epsilon(1-\overline{\gamma})}}$. Then, by Lemma 7, we have $|b_t - \theta_t| \le \delta$ for all $t \in [T]$, almost surely. Using Lemma 8 on the stability of revenue with respect to bid perturbations, we obtain

$$\begin{aligned}\mathsf{SReg}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}; \mathcal{E}, \mathcal{T}, \gamma) &\le \mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}, \mathcal{E}, \mathcal{T})} \left[\sup_{w \in \mathbb{B}^d} \sum_{t=1}^{T} \mathsf{rev}(w, x_t, b_t) - \sum_{t=1}^{T} \mathsf{rev}(w_t, x_t, b_t) \right] + 2\sqrt{\delta} T \\ &\le \mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}, \mathcal{E}, \mathcal{T})} \left[\sup_{w \in \mathbb{B}^d} \sum_{t=1}^{T} \mathsf{rev}(w, x_t, b_t) - \sum_{t=1}^{T} \mathsf{rev}(w_t, x_t, b_t) \right] + 2\epsilon T, \end{aligned} \tag{17}$$

for any $\mathcal{B} \in \mathsf{NE}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{E}, \mathcal{T}, \gamma)$, where the last line uses $\rho := \min\{1, \overline{\gamma}^{-1}(1-\overline{\gamma})\epsilon^5/3\}$, which implies $\sqrt{\delta} = \left(\frac{3\rho\overline{\gamma}}{\epsilon(1-\overline{\gamma})}\right)^{1/4} \le \epsilon$.

Finally, in Alg. 3, $\xi_t \cdot \mathsf{rev}(\cdot, x_t, b_t)$ is used as the reward function, where $\xi_t \sim \mathsf{Ber}(\rho)$. The first term in the RHS of the above bound is precisely the non-strategic regret, of Alg. 3 under a fixed buyer algorithm $\mathcal{B}$. Alg. 3's sparse update rule has a bounded effect on the regret, as shown in Lemma 10. Assuming the regret of the base expert algorithm is bounded by a concave function, specifically $\mathsf{Reg}^{\exp}(\mathcal{L}; \mathcal{E}) \le R(T \log K)$ for some concave R, Lemma 10 implies

$$\begin{aligned}&\mathop{\mathbb{E}}_{\mathsf{OMR}(\mathcal{A}^{\mathcal{L}}_{\mathsf{SUM}}, \mathcal{B}, \mathcal{E}, \mathcal{T})} \left[\sup_{w \in \mathbb{B}^d} \sum_{t=1}^{T} \mathsf{rev}(w, x_t, b_t) - \sum_{t=1}^{T} \mathsf{rev}(w_t, x_t, b_t) \right] \\ &\qquad \le \rho^{-1} R\big(\rho T \log |\mathcal{Z}_{\epsilon,T}|\big) + \mathcal{O}\left(\sqrt{\rho^{-1} T \log |\mathcal{Z}_{\epsilon,T}| T}\right), \end{aligned} \tag{18}$$

Substituting $\rho := \min\{1, \overline{\gamma}^{-1}(1-\overline{\gamma})\epsilon^5/3\}$ into the above, and using $|\mathcal{Z}_{\epsilon,T}| \le T^{\mathcal{O}(1/\epsilon^2)}$, and then combining (18) with (17), we obtain the stated bound. ■

D.1. Proof of Regret Stability under Sparse Updates

In this section, we prove the following lemma on the regret overhead introduced by sparse updating. The proof is a direct application of Bernstein's inequality for martingales, stated in Lemma 11 (Lemma A.8 in Cesa-Bianchi and Lugosi (2006)).

Lemma 10 *Let $\mathcal{L} : \mathbb{N} \times \bigcup_{t \in \mathbb{N}} \left([0,1]^K\right)^t \to \Delta_{[K]}$ be an online expert algorithm over K experts. Define $\mathcal{L}_\rho$ to be a modified algorithm where updates are computed using sparsely observed rewards:*

$$z_t \sim \mathcal{L}\big(T,\ (\xi_\tau r_\tau)_{\tau \in [t-1]}\big),$$

where $\xi_1, \ldots, \xi_T \sim \mathsf{Ber}(\rho)$ are i.i.d. Bernoulli random variables. Suppose the regret of $\mathcal{L}$ satisfies $\mathsf{Reg}^{\exp}(\mathcal{L}; \mathcal{E}) \le R(T \log K)$ for some concave function R. Then,

$$\mathsf{Reg}^{\exp}(\mathcal{L}_\rho; \mathcal{E}) \le \rho^{-1} R\big(\rho T \log K\big) + \mathcal{O}\left(\sqrt{\rho^{-1} T \log KT}\right),$$

where the expectation is taken over all sources of randomness, including the sequence $(\xi_t)_{t \in [T]}$ and the choices of r_t made by the environment.

Proof of Lemma 10. Let $(\mathcal{F}_t)_{t\in[T]}$ be a filtration, that is, a sequence of nested σ-algebras satisfying $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$, defined by $\mathcal{F}_t \coloneqq \sigma(\xi_{\leq t}, z_{\leq t})$, the σ-algebra generated by the random variables $\xi_{\leq t} \coloneqq (\xi_1, \ldots, \xi_t)$ and $z_{\leq t} \coloneqq (z_1, \ldots, z_t)$. Then, since the environment selects r_t based on past observations, r_t is $\mathcal{F}_{t-1}$-measurable. Hence, the sequence $(r_t)_{t\in[T]}$ is predictable with respect to the filtration $(\mathcal{F}_t)_{t\in[T]}$.

Fix $z \in [K]$. For each $t \in [T]$, define X_t and Y_t as follows:

$$X_t \coloneqq r_t(z) - r_t(z_t), \quad Y_t \coloneqq \xi_t X_t.$$

Note that X_t is $\mathcal{F}_{t-1}$-measurable (and hence predictable), since both r_t and z_t are $\mathcal{F}_{t-1}$-measurable. Since X_t is $\mathcal{F}_{t-1}$-measurable and ξ_t is independent of $\mathcal{F}_{t-1}$, the conditional expectation of Y_t given $\mathcal{F}_{t-1}$ is

$$\mathbb{E}[Y_t|\mathcal{F}_{t-1}] = X_t\,\mathbb{E}[\xi_t|\mathcal{F}_{t-1}] = X_t\,\mathbb{E}[\xi_t] = \rho X_t. \tag{19}$$

Let $Z_t \coloneqq \mathbb{E}[Y_t|\mathcal{F}_{t-1}] - Y_t$. Then, $\mathbb{E}[Z_t|\mathcal{F}_{t-1}] = 0$, and $|Z_t| \leq 2$, so $(Z_t)_{t\in[T]}$ forms a martingale difference sequence with respect to the filtration $(\mathcal{F}_t)_{t\in[T]}$. Moreover, the sum of conditional variances of Z_t, $\Sigma_T^2 \coloneqq \sum_{t=1}^T \mathbb{E}[Z_t^2|\mathcal{F}_{t-1}]$, is at most ρT, since

$$\mathbb{E}[Z_t^2|\mathcal{F}_{t-1}] \leq \mathbb{E}[Y_t^2|\mathcal{F}_{t-1}] = X_t^2\,\mathbb{E}[\xi_t^2] = X_t^2\,\mathbb{E}[\xi_t] \leq \rho, \tag{20}$$

as $|X_t| = |r_t(z) - r_t(z_t)| \leq 1$.

Next, we apply (19), (20), and Lemma 11 with $K = 2$ and $\nu = \rho T$, yielding

$$\Pr\left[\sum_{t=1}^T (\rho X_t - \xi_t X_t) > \sqrt{2\theta\rho T} + \frac{2\sqrt{2}}{3}\theta\right] \leq e^{-\theta},$$

for all $\theta > 0$. Since $X_t \coloneqq r_t(z) - r_t(z_t)$, this implies

$$\Pr\left[\sum_{t=1}^T \rho\big(r_t(z) - r_t(z_t)\big) > \sum_{t=1}^T \big(\xi_t r_t(z) - \xi_t r_t(z_t)\big) + \sqrt{2\theta\rho T} + \frac{2\sqrt{2}}{3}\theta\right] \leq e^{-\theta},$$

for any fixed $z \in [K]$. Setting $\theta = \log KT$ and applying the union bound over $z \in [K]$, we obtain

$$\Pr\left[\sup_{z\in[K]} \sum_{t=1}^T \rho\big(r_t(z) - r_t(z_t)\big) > \sup_{z\in[K]} \sum_{t=1}^T \big(\xi_t r_t(z) - \xi_t r_t(z_t)\big) + \mathcal{O}\left(\sqrt{\rho T \log KT}\right)\right] \leq T^{-1}.$$

Since $\left|\sum_{t=1}^T \big(r_t(z) - r_t(z_t)\big)\right| \leq T$, this leads to

$$\underbrace{\mathbb{E}\left[\sup_{z\in[K]} \sum_{t=1}^T \big(r_t(z) - r_t(z_t)\big)\right]}_{\coloneqq \mathsf{Reg}^{\mathrm{exp}}(\mathcal{L}_\rho; \mathcal{E})} \leq \rho^{-1}\,\mathbb{E}\left[\sup_{z\in[K]} \sum_{t=1}^T \big(\xi_t r_t(z) - \xi_t r_t(z_t)\big)\right] + \mathcal{O}\left(\sqrt{\rho^{-1} T \log KT}\right). \tag{21}$$

To bound the first term on the RHS, we express it as a nested expectation:

$$\mathbb{E}\left[\sup_{z\in[K]}\sum_{t=1}^{T}\left(\xi_t r_t(z)-\xi_t r_t(z_t)\right)\right]=\mathbb{E}\left[\mathbb{E}\left[\sup_{z\in[K]}\sum_{t=1}^{T}\left(\xi_t r_t(z)-\xi_t r_t(z_t)\right)\,\middle|\,\xi_{\le T}\right]\right].$$

Given fixed $\xi_1,\dots,\xi_T$, the inner conditional expectation corresponds exactly to the regret of $\mathcal{L}$ over the subsequence of rounds t where $\xi_t=1$, with rewards r_t. Hence,

$$\mathbb{E}\left[\sup_{z\in[K]}\sum_{t=1}^{T}\left(\xi_t r_t(z)-\xi_t r_t(z_t)\right)\,\middle|\,\xi_{\le T}\right]\le R\left(\sum_{t=1}^{T}\xi_t\log K\right).$$

Since R is concave, we apply Jensen's inequality:

$$\begin{aligned}\mathbb{E}\left[\sup_{z\in[K]}\sum_{t=1}^{T}\left(\xi_t r_t(z)-\xi_t r_t(z_t)\right)\right]&=\mathbb{E}\left[\mathbb{E}\left[\sup_{z\in[K]}\sum_{t=1}^{T}\left(\xi_t r_t(z)-\xi_t r_t(z_t)\right)\,\middle|\,\xi_{\le T}\right]\right]\\&\le\mathbb{E}\left[R\left(\sum_{t=1}^{T}\xi_t\log K\right)\right]\\&\le R\left(\sum_{t=1}^{T}\mathbb{E}[\xi_t]\log K\right)=R(\rho T\log K).\end{aligned}$$

Combining this with (21), we obtain

$$\mathsf{Reg}^{\exp}(\mathcal{L}_\rho;\mathcal{E})\le\rho^{-1}R\big(\rho T\log K\big)+\mathcal{O}\left(\sqrt{\rho^{-1}T\log KT}\right),$$

which concludes the proof of the stated bound. ■

Lemma 11 [Bernstein's inequality for martingales, Lemma A.8 in Cesa-Bianchi and Lugosi (2006)] *Let $Z_1,\dots,Z_T$ be a bounded martingale difference sequence with respect to the filtration $(\mathcal{F}_t)_{t=1}^T$, and with $|Z_t|\le K$. Then, for all constants $\theta,\nu>0$,*

$$\Pr\left[\sum_{t=1}^{T}Z_t>\sqrt{2\nu\theta}+\big(\sqrt{2}/3\big)K\theta\ \text{ and }\ \Sigma_T^2\le\nu\right]\le e^{-\theta},$$

where $\Sigma_T^2:=\sum_{t=1}^{T}\mathbb{E}[Z_t^2|\mathcal{F}_{t-1}]$ is the sum of the conditional variances.

Appendix E. Proof of Lazy OGD Regret Bound

In this section, we prove Lemma 9, which establishes the regret bound for Lazy Online Gradient Descent. This result follows from the regret bound for the lazy variant of Online Mirror Descent when using the squared Euclidean distance $\frac{1}{2}\|x-y\|^2$ as the Bregman divergence; see, for example, Bubeck et al. (2015); Hazan et al. (2016). Here, we present a more direct proof.

Lemma 9 *Let $M \in \mathbb{N}$, and for each $i \in [M]$, let $f_i : \mathbb{B}^d \to \mathbb{R}$ be any convex loss function with unit-bounded subgradients, i.e., $\|\nabla f_i\| \leq 1$. Then, for any $\beta > 0$ and $v^\star \in \mathbb{B}^d$,*

$$\sum_{i\in[M]} \big(f_i(v_i) - f_i(v^\star)\big) \leq (2\beta)^{-1} + 2\beta M, \tag{10}$$

where v_i is defined recursively as

$$u_{i+1} \coloneqq u_i - \beta\,\nabla f_i(v_i), \quad u_1 = 0 \quad \text{and} \quad v_i \coloneqq \frac{u_i}{\max\{1, \|u_i\|\}}. \tag{11}$$

Proof of Lemma 9 First, unroll the update rule for u_i to write

$$u_i = -\beta \sum_{s=1}^{i-1} \nabla f_s(v_s),$$

since $y_1 = 0$. Let $L_i(x) \coloneqq \frac{1}{2}\|x\|^2 + \beta \sum_{s=1}^{i-1} \big\langle \nabla f_s(v_s), x\big\rangle$. Then, we can rewrite v_i as

$$\begin{aligned} v_i \coloneqq \frac{u_i}{\max(1, \|u_i\|)} = \arg\min_{x\in\mathbb{B}^d} \|x - u_i\|^2 &= \arg\min_{x\in\mathbb{B}^d} \Big(\|x\|^2 - 2\,\langle u_i, x\rangle\Big) \\ &= \arg\min_{x\in\mathbb{B}^d} \underbrace{\Big(\|x\|^2 + 2\beta \sum_{s=1}^{i-1} \langle \nabla f_s(v_s), x\rangle\Big)}_{= 2L_i(x)} \\ &= \arg\min_{x\in\mathbb{B}^d} L_i(x). \end{aligned}$$

Since L_i is 1-strongly convex with respect to $\|\cdot\|$,

$$L_{i+1}(v_{i+1}) - L_{i+1}(v_i) \leq \big\langle \nabla L_{i+1}(v_{i+1}), v_{i+1} - v_i\big\rangle - \frac{1}{2}\|v_{i+1} - v_i\|^2 = -\frac{1}{2}\|v_{i+1} - v_i\|^2,$$

where the last equality follows from the optimality of $v_{i+1} = \arg\min_{x\in\mathbb{B}^d} L_{i+1}(x)$ and the convexity of L_{i+1}. On the other hand,

$$L_{i+1}(v_{i+1}) - L_{i+1}(v_i) \geq L_i(v_{i+1}) - L_i(v_i) + \beta\big\langle \nabla f_i(v_i), v_{i+1} - v_i\big\rangle \geq \beta\big\langle \nabla f_i(v_i), v_{i+1} - v_i\big\rangle,$$

as $v_i = \arg\min_{x\in\mathbb{B}^d} L_i(x)$. Combining the above inequalities,

$$\frac{1}{2}\|v_{i+1} - v_i\|^2 \leq -\beta\big\langle \nabla f_i(v_i), v_{i+1} - v_i\big\rangle \leq \beta\|\nabla f_i(v_i)\|\|v_{i+1} - v_i\|,$$

hence $\|v_{i+1} - v_i\| \leq 2\beta\|\nabla f_i(v_i)\|$ and

$$\big\langle \nabla f_i(v_i), v_i - v_{i+1}\big\rangle \leq \|\nabla f_i(v_i)\|\,\|v_{i+1} - v_i\| \leq 2\beta\|\nabla f_i(v_i)\|^2. \tag{22}$$

Next, we claim the following by induction: For all $i \leq M$, we have

$$\forall x \in \mathbb{B}^d, \quad \sum_{s=1}^{i} \big\langle \nabla f_s(v_s), v_{s+1} - x\big\rangle \leq \frac{\|x\|^2}{2\beta}. \tag{23}$$

For $t = 0$, the above holds as $0 \le \|x\|^2/2\beta$. Suppose (23) holds for some $i \le M-1$. Then, it holds especially for $x = v_{i+2} \coloneqq \arg\min_{x\in\mathbb{B}^d} L_{i+2}(x)$ and

$$\begin{aligned}
\sum_{s=1}^{i+1} \langle \nabla f_s(v_s), v_{s+1}\rangle &= \langle \nabla f_{i+1}(v_{i+1}), v_{i+2}\rangle + \sum_{s=1}^{i} \langle \nabla f_s(v_s), v_{s+1}\rangle \\
&\le \langle \nabla f_{i+1}(v_{i+1}), v_{i+2}\rangle + \sum_{s=1}^{i} \langle \nabla f_s(v_s), v_{i+2}\rangle + \frac{\|v_{i+2}\|^2}{2\beta} \\
&\qquad\qquad \text{(by the induction hypothesis)} \\
&= \underbrace{\sum_{s=1}^{i+1} \langle \nabla f_s(v_s), v_{i+2}\rangle + \frac{\|v_{i+2}\|^2}{2\beta}}_{= L_{i+2}(v_{i+2})} \\
&\le \underbrace{\sum_{s=1}^{i+1} \langle \nabla f_s(v_s), x\rangle + \frac{\|x\|^2}{2\beta}}_{= L_{i+2}(x)}, \quad \forall x \in \mathbb{B}^d. \qquad \text{(as } L_{i+2}(v_{i+2}) \le L_{i+2}(x)\text{)}
\end{aligned}$$

Hence, (23) holds for $i+1$.

Finally, combining (22) and (23) with $t = T$ and $x = v^\star$, we obtain

$$\begin{aligned}
\sum_{i=1}^{M} \big[f_i(v_i) - f_i(v^\star)\big] &\le \sum_{i=1}^{M} \langle \nabla f_i(v_i), v_i - v^\star\rangle && \text{(as } f_i \text{ is convex.)} \\
&= \sum_{i=1}^{M} \langle \nabla f_i(v_i), v_{i+1} - v^\star\rangle + \sum_{i=1}^{M} \langle \nabla f_i(v_i), v_i - v_{i+1}\rangle \\
&\le \frac{\|v^\star\|^2}{2\beta} + 2\beta \sum_{i=1}^{M} \|\nabla f_i(v_i)\|^2 && \text{(by (22) and (23))} \\
&\le \frac{1}{2\beta} + 2\beta T && \text{(as } \|v^\star\|, \|\nabla f_i(v_i)\| \le 1\text{)}
\end{aligned}$$

■

Appendix F. Other Related Work

In this section, we expand on the related work to more clearly situate our contributions within the existing literature.

Dynamic Pricing. The seminal work on online learning for pricing is by Kleinberg and Leighton (2003), who studied repeated posted-price auctions under both stochastic and adversarial environments. In the stochastic setting, buyer valuations are drawn from an unknown distribution, while in the adversarial case, they may be chosen adaptively and adversarially. This work laid the foundation for a broad literature on dynamic pricing, often with bandit feedback (observing only whether

the buyer purchases the good), including variants such as demand learning (Lin, 2006; Besbes and Zeevi, 2009, 2015). Subsequent research has extended these models to settings with limited supply (Babaioff et al., 2012; Yang and Zhang, 2014), strategic buyer behavior (Amin et al., 2013; Liu et al., 2018), and contextual information (Chen and Gallego, 2022; Luo et al., 2024). For a comprehensive overview of this literature, see the survey by Den Boer (2015).

Contextual Pricing. Learning an optimal pricing policy in the presence of contextual information has been studied in the batch learning setting by Mohri and Medina (2014); Munoz and Vassilvitskii (2017); Shen et al. (2019). These works provide algorithms with PAC-style guarantees under the assumption that context–value pairs are drawn i.i.d. from an unknown distribution. A more recent contribution by Liu et al. (2020) investigates the computational and sample complexity of learning optimal linear pricing policies, a task they refer to as Myersonian regression. They show that obtaining a fully polynomial-time approximation scheme (FPTAS) is impossible under standard complexity-theoretic assumptions.