Contextual Dynamic Pricing with Heterogeneous Buyers

Anonymous Author(s)
Affiliation
Address
email

Abstract

We initiate the study of contextual dynamic pricing with a heterogeneous population of buyers, where a seller repeatedly (over T rounds) posts prices that depend on the observable d dimensional context and receives binary purchase feedback. Unlike prior work assuming homogeneous buyer types, in our setting the buyer's valuation type is drawn from an unknown distribution with finite support $K_\star$. We develop a contextual pricing algorithm based on Optimistic Posterior Sampling with regret $\widetilde{O}(K_\star\sqrt{dT})$, which we prove to be tight in d and T up to logarithmic terms. Finally, we refine our analysis for the non-contextual pricing case, proposing a variance-aware Zooming algorithm that achieves the optimal dependence on $K_\star$.

1 Introduction

In online learning for contextual pricing, a learner (aka seller) repeatedly sets prices for different products with the goal of maximizing revenue through interactions with agents (aka buyers or customers). Concretely, in each round $t = 1, \ldots, T$, nature selects a product with a d-dimensional feature representation u_t (context) and the seller selects a price $p_t \geq 0$. In the simplest variant, the *linear valuation model*, customers have a fixed intrinsic valuation model (type) that is unknown to the learner; this has a d-dimensional representation $\theta^\star$ whose coordinates reflect the valuation that each product feature adds, i.e., the customer's valuation is $v_t = \langle \theta^\star, u_t \rangle + \varepsilon_t$ where ε_t is a noise term. The customer makes a purchase only when their valuation is higher than the price, i.e., $v_t \geq p_t$. The learner's goal is to maximize revenue, i.e., the sum of the prices in rounds when purchases occur. An equivalent objective is to minimize *regret*, which is measured against a benchmark that always selects the customer's valuation as the price for the given round.

One key difficulty is that the learner faces both an infinite action space (i.e., all possible prices) and a discontinuous revenue function, hence causing sharp revenue loss for the learner. However, the problem offers a richer feedback structure than classical multi-armed bandits: a non-purchase indicates that all higher prices would also be rejected by the buyer, while a purchase confirms that all lower prices would be accepted too. The two primary approaches from the literature to tackle this problem involve estimating the unknown parameter $\theta^\star$ through online regression or multi-dimensional binary search (see Section 1.1 for further discussion).

A crucial limitation for both approaches is that they require all customers to behave *homogeneously* according to a single type $\theta^\star$ (see related work for results robust to small deviations from this assumption). Moving beyond this homogeneity assumption, we pose the following question:

How can one design contextual pricing algorithms with a heterogeneous population of customers?

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

1.1 Our contribution

Our setting (Section 2). To study contextual pricing with a heterogeneous buyer population, we assume that the type θ_t in round t is drawn from a fixed, unknown distribution $D_\star$. When $D_\star$ is supported on a single type $\theta_\star$, we recover the homogeneous setting. In our model, the number of distinct buyer types $K_\star = |\operatorname{supp} D_\star| (> 1)$ reflects the *degree of heterogeneity*.

There are several obstacles to applying existing algorithms from the literature. First, canonical contextual pricing algorithms based on linear regression either compete against (simple) linear policies or assume context-independent and identically distributed (i.i.d.) valuation noise. In contrast, the optimal policy in our setting may best respond based on a *context-dependent* type rather than a *fixed* type, and the stochasticity due to heterogeneity is inherently context-dependent and thus non-i.i.d. Second, given that *the buyer types are not observable*, one cannot connect the observed feedback to shrinkage of type-dependent uncertainty sets; this rules out running canonical multi-dimensional binary search / contextual pricing algorithms for each buyer type in parallel. Third, since in our setting there is a *continuum of actions*, any canonical contextual bandits algorithm whose regret scales with the discretized action count (e.g., EXP4) will suffer suboptimal performance.

Our contextual pricing algorithm (Section 3). To tackle the above challenges, we employ recent advances in the contextual bandit literature that attain a better scaling with the number of actions, thus evading the shortcomings of EXP4 with naïve discretization. In particular, we build on the *Optimistic Posterior Sampling* (OPS) approach [29] which, in our setting, maintains a posterior μ_t over all candidate type distributions. We call these candidate type distributions *models* and refer to their (possibly infinite) family as $\mathcal{D}$. At a high level, in every round, OPS best responds to a model sampled from μ_t. As typical in online learning, the posterior update penalizes models that *disagree* with the observed feedback (*model mismatch*) aiming to converge to the model $D_\star$. To encourage exploration in the absence of full information, this penalty is reduced by an *optimism bias* term that rewards models with the highest potential to positively contribute to the revenue. The OPS approach enables regret bounds of $\sqrt{T \cdot c \cdot \log |\mathcal{D}|}$, scaling with a *disagreement coefficient* c that measures the structural complexity of the reward functions and captures the tension between exploration and exploitation. Note that c can be significantly smaller than the number of actions.

Our main technical contributions in adapting OPS to heterogeneous contextual pricing are twofold. First, to bound the disagreement coefficient c, we observe that, for any fixed context, the aggregate demand function induced by $D_\star$ has at most $K_\star$ "jumps",[1] thus creating $K_\star + 1$ intervals. Over each interval, we bound the disagreement coefficient by a factor of 2. Combining these arguments with a novel decomposition lemma for the disagreement coefficient of functions with $K_\star$ breakpoints, we show that $c \le 2(K_\star + 1)$. When $K_\star$ is known, we apply a variant of OPS over a finite covering of the class $\mathcal{D}$ containing *all* possible distributions over $K_\star$ types, of log cardinality $dK_\star \log T$. Second, to extend our sublinear regret guarantee to the infinite model class $\mathcal{D}$, we modify OPS to conservatively perturb its recommended prices (which cannot overly impact regret due to one-sided Lipschitzness of the revenue function). We then construct a coupling between the actual trajectory of OPS and one where $D_\star$ belongs to the finite cover, allowing us to transfer regret bounds. Finally, we adapt to unknown $K_\star$ by initializing OPS with a non-uniform prior over models. These technical contributions enable us to show a regret guarantee of $\widetilde{O}(K_\star \sqrt{dT})$. Finally, we show that this guarantee is optimal (up to logarithmic terms) with respect to the dependence on both the contextual dimension d and the time horizon T, establishing a lower bound of $\Omega(\sqrt{K_\star dT})$ for sufficiently large $T = \Omega(dK_\star^3)$.

Non-contextual improvements (Section 4). The above upper and lower bounds raise a natural question on the optimal dependence on the number of buyer types $K_\star$; we resolve this question in the non-contextual version of the problem ($d = 1$) by providing an algorithm with an upper bound of $\widetilde{O}(\sqrt{K_\star T})$. Our algorithm, `ZoomV`, combines zooming (i.e., adaptive discretization) methods from Lipschitz bandits [13] with variance-aware confidence intervals [2]. We show that the regret of `ZoomV` scales with a novel variance-aware zooming dimension that can be significantly smaller than the standard measure of complexity for Lipschitz bandits. For pricing, this variance adaptation unlocks our $\widetilde{O}(\min\{\sqrt{K_\star T}, T^{2/3}\})$ bound (versus $O(T^{2/3})$, obtained via the standard zooming analysis).

We note that the non-contextual version of pricing for heterogeneous buyers has been previously studied by [4], who establish a matching upper and lower bound if all types are "well separated" from

[1] Each "jump" corresponds to a change of type from (say) type i to type $i + 1$.

each other. Although our algorithmic approach towards this guarantee is very different, our result can be viewed as a strengthening of theirs by removing this separation assumption.

Stronger type observability (Section 5). Finally, we consider contextual pricing under the assumption that the learner can *identify* each arriving type, i.e., where the learner observes ex-post information about the sampled type θ_t. We analyze two observability models: one where the learner receives a discrete identifier $z_t \in [K_\star]$ —under which a computationally efficient pricing algorithm matches the $\tilde{O}(K_\star\sqrt{dT})$ bound of OPS— and another where the full type vector $\theta_t \in \mathbb{B}^d$ is observed — for which we reduce the dependence on $K_\star$ and d to achieve regret $\tilde{O}(\sqrt{\min\{K_\star, d\}T})$. These results demonstrate how richer feedback reduces complexity in dynamic pricing.

Related work. We briefly review three existing lines of work, deferring full discussion to Appendix A. The closest to ours is *contextual pricing/search*, where a learner interacts with nature to learn a hidden vector $\theta^\star \in \mathbb{R}^d$ while receiving single-bit feedback [5, 18, 21, 17, 12, 11, 6, 19]. Under an appropriate pricing loss, this setting reduces to ours with a *homogeneous* buyer population, i.e., $K_\star = 1$. However, the deterministic nature of the feedback leads to aggressive search policies via cutting planes that do not lend themselves to the heterogeneous case. Although some works tolerate i.i.d., context-independent valuation noise [12, 11, 6, 19], their methods do not treat our non-i.i.d., context-dependent noise due to heterogeneity. Next is the case of *non-contextual dynamic pricing* where $d = 1$, originally treated with multi-armed bandits methods by [14]. The closest non-contextual work is [4], whose finite-types model reduces to ours with $d = 1$. We improve upon their regret bound, but neither our improvement nor the existing methods generalize readily to the contextual setting. Finally, our setting relates to *Lipschitz bandits*. Although revenue is *not* fully Lipschitz, it satisfies a *one-sided Lipschitzness*, enabling the use of zooming [13] when $d = 1$ (see, e.g., [24]). We successfully refine these for the non-contextual case, but contextual variants of zooming [26, 15] scale with complexity parameters which admit no direct bounds for heterogeneous pricing.

2 Setup and Preliminaries

Notation. Let $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ denote the Euclidean norm and inner product on $\mathbb{R}^d$. Let $\mathbb{S}^{d-1}, \mathbb{B}^d \subseteq \mathbb{R}^d$ denote the unit sphere and ball, respectively. Let $\Delta(S)$ denote the set of all probability measures on a measurable set $S \subseteq \mathbb{R}^d$, and let $\mathrm{supp}(D)$ denote the support of $D \in \Delta(\mathbb{R}^d)$. We use $\Delta_k(S)$ for those $D \in \Delta(S)$ with $|\mathrm{supp}(D)| \leq k$. For a positive integer m, let $[m] \coloneqq \{1, 2, \ldots, m\}$.

Setup. We consider T rounds of repeated interaction between a seller, a population of buyers, and an adversary. At each round $t \in [T]$, the seller posts a price $p_t \in [0, 1]$ for an item to be sold and a buyer, sampled from the population, decides whether or not to buy the item based on their valuation $v_t \in [0, 1]$. We denote the indicator of their purchase by $y_t = \mathbb{1}\{v_t \geq p_t\}$. The valuation of the buyer is determined by two factors: their *type* θ_t (encoding their intrinsic preferences), and an external *context* u_t, which describes the current item to be sold and any relevant environmental factors. The learner does *not* know θ_t, but they *do* know u_t. We use a linear valuation model: i.e., θ_t and u_t lie in d-dimensional spaces $\Theta \subseteq [0, 1]^d$ and $\mathcal{U} \subseteq \mathbb{S}^{d-1}$, respectively, and take $v_t = \langle\theta_t, u_t\rangle$. We assume that $\langle\theta, u\rangle \in [0, 1]$ for all $\theta \in \Theta$ and $u \in \mathcal{U}$. We impose no further assumptions on the contexts, allowing them to be generated (potentially adaptively) by the adversary. On the other hand, we assume that each θ_t is sampled independently from a fixed distribution $D_\star \in \Delta(\Theta)$ that describes the buyer population, unknown to the seller. All together, the following occur at each round $t \in [T]$:

1. the adversary selects a context $u_t \in \mathcal{U}$;
2. a buyer arrives with type $\theta_t \in \Theta$ sampled independently from $D_\star$, with valuation $v_t = \langle u_t, \theta_t\rangle$;
3. the seller observes u_t and posts price $p_t \in [0, 1]$ for the item;
4. the seller observes the purchase decision $y_t = \mathbb{1}\{v_t \geq p_t\}$ and receives revenue $p_t y_t$.

Benchmark. The seller's goal is to maximize their cumulative revenue compared to that which they could have achieved with knowledge of $D_\star$. To express this concisely, we introduce some additional notation. Each distribution Q over valuations in $[0, 1]$ induces the following:

- a demand function $\mathsf{dem}_Q(p) \coloneqq \mathbb{P}_{v\sim Q}[v \geq p]$,
- an expected revenue function $\mathsf{rev}_Q(p) \coloneqq p \cdot \mathsf{dem}_Q(p)$,

- a revenue-maximizing best response $\mathsf{br}_Q := \arg\max_{p\in[0,1]} \mathsf{rev}_Q(p)$ (breaking ties arbitrarily),
- and a gap function $\mathsf{gap}_Q(p) := \mathsf{rev}_Q(\mathsf{br}_Q) - \mathsf{rev}_Q(p)$.

Note that each type distribution $D \in \Delta(\Theta)$ and context $u \in \mathcal{U}$ induce a value distribution $Q = \mathsf{proj}(D, u)$, where $\mathsf{proj}(D, u)$ is defined as the probability law of $\langle \theta, u\rangle$ when $\theta \sim D$. We then set $\mathsf{dem}_D(p, u) := \mathsf{dem}_Q(p)$, $\mathsf{rev}_D(p, u) := \mathsf{rev}_Q(p)$, $\mathsf{br}_D(u) := \mathsf{br}_Q$, and $\mathsf{gap}_D(p, u) := \mathsf{gap}_Q(p)$, accordingly. We abbreviate a subscript of $D_\star$ by "$\star$" alone, writing, e.g., $\mathsf{dem}_\star(p, u)$ and $\mathsf{br}_\star(u)$.

A seller policy $\mathcal{A}$ is a (potentially randomized) map from a history $\{u_\tau, p_\tau, y_\tau\}_{\tau=1}^{t-1}$ and the current context u_t to a posted price p_t. An adversary policy $\mathcal{B}$ is a (potentially randomized) map from a history $\{u_\tau, \theta_\tau, p_\tau, y_\tau\}_{\tau\in[t-1]}$ to the next context u_t. We then define the seller's *pricing regret* by

$$R_{\mathcal{A},\mathcal{B}}(T) = \textstyle\sum_{t\in[T]} \mathsf{gap}_\star(p_t, u_t) = \sum_{t\in[T]} \big(\mathsf{rev}_\star(\mathsf{br}_\star(u_t), u_t) - \mathsf{rev}_\star(p_t, u_t)\big),$$

where $\{u_t, p_t\}_{t\in[T]}$ are selected according to $\mathcal{A}$ and $\mathcal{B}$. We will omit the policies from the subscript when clear from context. We focus on controlling the pricing regret in expectation, and will say that $\mathcal{A}$ satisfies a regret bound $f(T)$ if $\mathbb{E}[R_{\mathcal{A},\mathcal{B}}(T)] \le f(T)$ for all $\mathcal{B}$.

Our guarantees will scale with context dimension d and the *degree of heterogeneity*, which we quantify via the support size $K_\star := |\operatorname{supp}(D_\star)|$ (that may be infinite). We do *not* assume that $K_\star$ is known to the seller. Designing an effective seller policy is challenging because $D_\star$, $K_\star$, and the realized buyer types are unknown to the seller, who must carefully balance exploration and exploitation given only the current context and the history of purchase outcomes.

Basic pricing facts. Finally, we provide some basic properties of the pricing problem, with proofs in Appendix B. Essential for this work is the one-sided Lipschitzness of the expected revenue function. This is an immediate consequence of the monotonicity of demand functions, and it has previously been used to apply techniques from Lipschitz bandits to non-contextual pricing [24].

Lemma 2.1 (One-sided Lipschitzness)**.** *Fix any distribution $Q \in \Delta([0, 1])$ and let $0 \le p < p' \le 1$. We then have $\mathsf{rev}_Q(p') - \mathsf{rev}_Q(p) \le \mathsf{dem}_Q(p)(p' - p) \le p' - p$.*

Throughout this work, we must handle distributional uncertainty over value distributions. To compare two distributions $P, Q \in \mathcal{P}([0, 1])$, we employ the *Levy metric* defined by

$$d_{\mathrm{L}}(P, Q) := \inf\{\varepsilon > 0 : \mathsf{dem}_P(x - \varepsilon) - \varepsilon \le \mathsf{dem}_Q(x) \le \mathsf{dem}_P(x + \varepsilon) + \varepsilon \;\forall x \in \mathbb{R}\}. \quad (1)$$

This quantity is at most 1 and equals the side length of the largest square which can be inscribed between the graphs of dem_P and dem_Q (equivalently, the CDFs of P and Q). For type distributions $D, D' \in \mathcal{P}(\Theta)$, we use the Levy distance between their projected value distributions, taking $d_{\mathrm{L}}(D, D') := \sup_{u\in\mathbb{S}^{d-1}} d_{\mathrm{L}}(\mathsf{proj}(D, u), \mathsf{proj}(D', u))$. We use this metric because, if D and D' are close under d_{L}, then there exists a policy which performs well on both of them; this property motivates the use of the Lévy metric throughout the dynamic pricing literature (see, e.g., [23]).

Lemma 2.2 (Pricing implication of Lévy metric bound)**.** *Suppose that $D, D' \in \Delta(\Theta)$ satisfy $d_{\mathrm{L}}(D, D') < \varepsilon$. Then the conservative best-response policy $\pi(u) = \max\{\mathsf{br}_D(u) - \varepsilon, 0\}$ satisfies $\mathsf{rev}_D(\pi(u), u) \ge \mathsf{rev}_D(\mathsf{br}_D(u), u) - \varepsilon$ and $\mathsf{rev}_{D'}(\pi(u), u) \ge \mathsf{rev}_{D'}(\mathsf{br}_{D'}(u), u) - 3\varepsilon$ for all $u \in \mathcal{U}$.*

3 Contextual Algorithm with Optimal Dependence on d and T

In this section, we develop statistically efficient algorithms for contextual pricing. In Section 3.1, we treat the simpler setting where $D_\star$ belongs to a finite model class $\mathcal{D}$ and $K_\star$ is known. In Section 3.2, we remove these two assumptions and prove a lower bound with matching dependence on d and T.

3.1 Warm-Up: Heterogeneous Contextual Pricing with a Finite Model Class

As a warm-up, we consider pricing when $D_\star$ belongs to a known, finite model class.

Assumption 3.1. Assume that $D_\star \in \mathcal{D}$, where $\mathcal{D} \subseteq \Delta(\Theta)$ is finite and known to the seller.

This realizability assumption simplifies our analysis, and the resulting algorithm extends naturally to infinite classes. We employ optimistic posterior sampling (`OPS`), originally studied for contextual bandits under the name "Feel-Good Thompson Sampling" [29]. In our instantiation for contextual pricing, OPS (Algorithm 1) maintains a posterior distribution over models, initialized at prior $\mu_1 \in \Delta(\mathcal{D})$. At round t with context $u_t \in \mathcal{U}$, we sample a model $D_t \sim \mu_t$, play the best-response price $p_t = \mathsf{br}_{D_t}(u_t)$, and observe purchase feedback y_t. Then, for each candidate model $D \in \mathcal{D}$, we update its posterior weight $\mu_{t+1}(D)$ according to the loss $\ell_\lambda(\mathsf{proj}(D, u_t), p_t, y_t)$, defined by

$$\ell_\lambda(Q, p, y) \coloneqq \underbrace{(y - \mathsf{dem}_Q(p))^2}_{\text{model mismatch}} - \underbrace{\lambda \mathsf{rev}_Q(\mathsf{br}_Q)}_{\text{optimism bias}}.$$

The "model mismatch" penalty captures the extent to which the observed demand y_t differs from that predicted by D when p_t is played. In particular, as a function of $D \in \mathcal{D}$, the expected model mismatch $\mathbb{E}_{y_t}[(y_t - \mathsf{dem}_D(p_t, u_t))^2 \mid p_t]$ is minimized only by models D which make the same prediction as $D_\star$, i.e., those for which $\mathsf{dem}_D(p_t, u_t) = \mathsf{dem}_\star(p_t, u_t)$. The "optimism bias" reduces the loss for models which have the potential to provide large revenue, ensuring that we still explore.

ALGORITHM 1: OPS: Contextual Pricing with a Finite Model Class

---

Input: finite model class $\mathcal{D} \subseteq \Delta(\Theta)$, support size $K \geq 1$;
initialize uniform prior $\mu_1 = \mathrm{Unif}(\mathcal{D})$ and optimism strength $\lambda = \sqrt{\log(|\mathcal{D}|)/KT}$;
for *each round* $t \in [T]$ **do**
  observe context u_t;
  sample model $D_t \sim \mu_t$;
  play $p_t = \mathsf{br}_{D_t}(u_t)$ and observe y_t;
  update $\mu_{t+1}(D) \propto \mu_t(D) \exp\big(-\ell_\lambda(\mathsf{proj}(D, u_t), p_t, y_t)\big)$ for each $D \in \mathcal{D}$;

Theorem 3.2. *Under Assumption 3.1,* OPS *with $K = K_\star$ achieves regret $\widetilde{O}(\sqrt{K_\star T \log|\mathcal{D}|})$.*

The requirement of known $K_\star$ is imposed for simplicity and will be removed in Section 3.2. To prove Theorem 3.2, we employ a disagreement coefficient that controls the per-context complexity of balancing exploration and exploitation. In general, for an arbitrary measurable space $\mathcal{X}$ and function class $\mathcal{F} : \mathcal{X} \to \mathbb{R}$, we define the *disagreement coefficient* of $\mathcal{F}$ by

$$\mathsf{dis}(\mathcal{F}) \coloneqq \sup_{\varepsilon, \delta > 0} \sup_{\nu \in \Delta(\mathcal{X})} \frac{\delta^2}{\varepsilon^2} \mathbb{P}_{p \sim \nu}\Big(\exists f \in \mathcal{F} : \mathop{\mathbb{E}}_{q \sim \nu}[f(q)^2] \leq \varepsilon^2 \wedge |f(p)| > \delta\Big). \tag{2}$$

Variants of this quantity have previously been used to analyze a variety of structured bandits and active learning problems (see Remark 3.7). The δ^2/ε^2 scaling was historically chosen so that $\mathsf{dis}(\mathcal{F})$ can be directly bounded by the domain size $|\mathcal{X}|$. For our application, $\mathcal{X} = [0, 1]$ is the (infinite) price set and each function $f \in \mathcal{F}$, induced by a model $D \in \mathcal{D}$ and context $u \in \mathcal{U}$, measures the discrepancy between the demand functions of D and $D_\star$ after projection onto u, i.e., $f(p) = \mathsf{dem}_D(p, u) - \mathsf{dem}_\star(p, u)$. In particular, we set

$$\mathsf{dis}(\mathcal{D}, D_\star) \coloneqq \sup_{u \in \mathcal{U}} \mathsf{dis}\big(\{\mathsf{dem}_D(\cdot, u) - \mathsf{dem}_\star(\cdot, u) : D \in \mathcal{D}\}\big).$$

By this definition, if $\mathsf{dis}(\mathcal{D}, D_\star)$ is small and the seller plays price $p \sim \nu$ when faced with context u, it is unlikely for a model D to disagree with $D_\star$ at p if it is close to $D_\star$ under the $L^2(\nu)$ norm, i.e., if $\mathbb{E}_{q \sim \nu}[(\mathsf{dem}_D(q, u) - \mathsf{dem}_\star(q, u))^2]$ is small. The regret of OPS is bounded by $\mathsf{dis}(\mathcal{D}, D_\star)$ as follows.

Lemma 3.3. *Under Assumption 3.1 with optimism strength $\lambda > 0$,* OPS *(Algorithm 1) achieves regret $\widetilde{O}\big(\lambda\, \mathsf{dis}(\mathcal{D}, D_\star) T + \log(|\mathcal{D}|)/\lambda\big)$.*

The proof in Appendix C.3 combines the OPS analysis of [29] with a decoupling lemma due to [8]. To control $\mathsf{dis}(\mathcal{D}, D_\star)$, we show that each function of the form $\mathsf{dem}_D(\cdot, u) - \mathsf{dem}_\star(\cdot, u)$ can be decomposed into $K_\star + 1$ non-increasing pieces. In Appendix C.4, we prove the following disagreement coefficient bound for non-increasing functions.

Lemma 3.4. *Let $\mathcal{F} : [0, 1] \to \mathbb{R}$ be the set of nonincreasing functions. Then* $\mathsf{dis}(\mathcal{F}) \leq 2$.

Next, we examine a useful notion of *composite* function classes. A function class $\mathcal{G} : \mathcal{Z} \to \mathbb{R}$ is called an N-composite of $\mathcal{F} : \mathcal{X} \to \mathbb{R}$ if there exists a disjoint partition $\mathcal{Z} = \mathcal{Z}_1 \cup \cdots \cup \mathcal{Z}_N$ and mappings $\{h_i : \mathcal{Z}_i \to \mathcal{X}\}_{i \in [N]}$ such that each $g \in \mathcal{G}$ can be decomposed as $g(x) = f_i(h_i(x))$ for all $x \in \mathcal{Z}_i$ and $i \in [N]$, for some choice of $\{f_i : \mathcal{X} \to \mathbb{R}\}_{i \in [N]}$. We show the following in Appendix C.5.

Lemma 3.5. *If $\mathcal{G}$ is an N-composite of $\mathcal{F}$, then* $\mathsf{dis}(\mathcal{G}) \le N\mathsf{dis}(\mathcal{F})$.

Proof of Theorem 3.2. For each $u \in \mathcal{U}$, the function $\mathsf{dem}_\star(\cdot, u)$ is piecewise constant with $K_\star + 1$ sections, since jumps can only occur at the projections of the $K_\star$ types. For any $D \in \mathcal{D}$, the demand $\mathsf{dem}_D(\cdot, u)$ is monotonic, since increasing price always reduces demand. Thus, $\mathsf{dem}_D(\cdot, u) - \mathsf{dem}_\star(\cdot, u)$ is non-increasing on each of the $K_\star + 1$ sections, and so the function classes defining $\mathsf{dis}(\mathcal{D}, D_\star)$ are $(K_\star + 1)$-composites of the non-increasing function class. Applying Lemmas 3.4 and 3.5 then implies that $\mathsf{dis}(\mathcal{D}, D_\star) \le 2(K_\star + 1)$. The theorem then follows by the regret bound of Lemma 3.3 and our choice of λ. $\square$

Remark 3.6 (Comparison to Thompson sampling)**.** *Standard Thompson sampling corresponds to the alternative choice of log losses: $\ell(Q, p, y) = \log \mathbb{P}_{z \sim \mathrm{Ber}(\mathsf{dem}_Q(p))}[z = y] = y \log \mathsf{dem}_Q(p) + (1 - y)\log(1 - \mathsf{dem}_Q(p))$. In comparison,* `OPS` *uses the squared loss (this is not essential but simplifies analysis) and an optimism bias towards models under which the seller can attain large revenue. This is crucial for obtaining frequentist (rather than Bayesian) regret bounds, as outlined in [29].*

Remark 3.7 (Relation with existing results)**.** *Variants of the disagreement coefficient and the related Alexander capacity are well-studied in the active learning and empirical process theory literature [9]. The version above was first considered by [7]. [8] proved a regret bound which translates to $\widetilde{O}(\sqrt{\mathsf{dis}(\mathcal{D}) T \log |\mathcal{D}|})$ in our setting, matching Theorem 3.2. However, the estimation-to-decisions (E2D) meta-algorithm which they employ is non-constructive, hence we apply OPS instead. In [29], the regret of OPS is controlled by a distinct "decoupling coefficient." Our proof of Lemma 3.3 shows that a (slightly modified) decoupling coefficient is bounded by the disagreement coefficient.*

3.2 The General Case

We now seek to eliminate the assumptions that $D_\star$ belongs to a finite class $\mathcal{D}$ and that the support size $K_\star$ is known to the seller. For the first point, we loosen the requirement that $D_\star$ belongs to $\mathcal{D}$ and take $\mathcal{D}$ to be a large, finite cover of the full distribution space $\Delta(\Theta)$. Then, we replace the uniform prior μ_1 with a non-uniform prior that places less weight on models with large support sizes. Ultimately, this will enable a choice of optimism strength λ that is independent of $K_\star$, achieving our second goal. Unfortunately, if $D_\star$ is close but not equal to a model in $\mathcal{D}$, our analysis of `OPS` fails.

To remedy this, we employ perturbed OPS (`POPS`, Algorithm 2), an OPS variant with conservatively perturbed and discretized prices. This modified algorithm and its analysis require some new notation. Given a value distribution $Q \in \Delta([0, 1])$, define the ε-smoothed demand function $\mathsf{dem}^\varepsilon_Q$ by

$$\mathsf{dem}^\varepsilon_Q(p) := \mathop{\mathbb{E}}_{\delta \sim \mathrm{Unif}([0,\varepsilon])}[\mathsf{dem}_Q(p - \delta)].$$

Similarly, we let $\mathsf{rev}^\varepsilon_Q(p) := p\,\mathsf{dem}^\varepsilon_Q(p)$. Define contextual extensions $\mathsf{dem}^\varepsilon_D(p, u)$ and $\mathsf{dem}^\varepsilon_D(p, u)$ as in the non-smoothed case. To incorporate discretization, write $\mathcal{P}_\varepsilon := \varepsilon\mathbb{N} \cap [0, 1]$ for prices which are multiples of ε and define $\mathsf{br}^\varepsilon_Q := \arg\max_{p \in \mathcal{P}_\varepsilon} \mathsf{rev}^\varepsilon(p)$ (lifting to $\mathsf{br}^\varepsilon_D(u)$ as before).

Now, at each round $t \in [T]$, `POPS` samples a model $D_t \sim \mu_t$ from the current posterior $\mu_t \in \Delta(\mathcal{D})$ and computes its (discretized) best response $\hat{p}_t = \mathsf{br}^\varepsilon_{D_t}(u_t)$. Instead of posting price $\hat{p}_t$ directly, POPS posts $p_t = \max\{\hat{p}_t - \delta_t, 0\}$, where $\delta_t \sim \mathrm{Unif}([0, \varepsilon])$ is a small random perturbation. Due to this perturbation and discretization, we employ the modified loss $\ell^\varepsilon_\lambda(\mathsf{proj}(D_t, u_t), \hat{p}_t, y_t)$, where

$$\ell^\varepsilon_\lambda(Q, \hat{p}, y) := \left(\mathsf{dem}^\varepsilon_Q(\hat{p}) - y\right)^2 - \lambda \mathsf{rev}^\varepsilon_Q(\mathsf{br}^\varepsilon_Q).$$

The perturbations allow us, in the analysis of `POPS`, to couple its trajectory when run with $D_\star \notin \mathcal{D}$ to a trajectory where $D_\star \in \mathcal{D}$. The discretization is needed to bound a modified disagreement coefficient which appears in the analysis. All together, we obtain the following.

Theorem 3.8. `POPS` *(Algorithm 2) achieves regret $\widetilde{O}(K_\star \sqrt{dT})$ with appropriately tuned parameters and without prior knowledge of $K_\star$. Moreover, even for known $K_\star \ge 2$ and stochastic contexts, no contextual pricing policy can achieve expected regret $o(\sqrt{K_\star d T})$ for all instances if $T \ge d \cdot K_\star^3$.*

For the upper bound, our analysis views the perturbation at Step 5 as performed by nature, rather than the seller. Treating the seller's action as $\hat{p}_t$, they then observe a purchase ($y_t = 1$) with probability

$$\mathop{\mathbb{E}}_{\delta_t}[\mathsf{dem}_\star(\max\{\hat{p}_t - \delta_t, 0\}, u_t) \mid \hat{p}_t, u_t] = \mathop{\mathbb{E}}_{\delta_t}[\mathsf{dem}_\star(\hat{p}_t - \delta_t, u_t) \mid \hat{p}_t, u_t] = \mathsf{dem}^\varepsilon_\star(\hat{p}_t, u_t),$$

ALGORITHM 2: Perturbed OPS (`POPS`) for Contextual Pricing with Infinite Model Class

Input: discretization error $\varepsilon \in [0, 1)$, finite model cover $\mathcal{D} \subseteq \Delta(\Theta)$,
model prior $\mu_1 \in \Delta(\mathcal{D})$, optimism strength $\lambda > 0$;
for *each round* $t \in [T]$ **do**
  observe context u_t;
  sample model $D_t \sim \mu_t$ and perturbation strength $\delta_t \sim \mathrm{Unif}([0, \varepsilon])$;
  play $p_t = \max\{\hat{p}_t - \delta_t, 0\}$, where $\hat{p}_t = \mathsf{br}^{\varepsilon}_{D_t}(u_t)$ and observe y_t;
  update $\mu_{t+1}(D) \propto \mu_t(D) \exp\big(-\ell^{\varepsilon}_{\lambda}(\mathsf{proj}(D, u_t), \hat{p}_t, y_t)\big)$ for each $D \in \mathcal{D}$;

justifying the definitions above. Through this lens, `POPS` can viewed as OPS for an alternative, smoothed demand model. To bound regret, we apply a three-step argument. First, we show that `POPS` maintains our OPS regret bound when $D_\star \in \mathcal{D}$. This requires bounding a modified decoupling coefficient and is the only step where discretization is used. A direct application of the previous OPS analysis provides a regret bound with respect to a smoothed and discretized benchmark. Fortunately, one-sided Lipschitzness of revenue (Lemma 2.1) ensures that this modified regret benchmark is within $O(\varepsilon T)$ of the original benchmark, as we prove in Appendix C.6.

Lemma 3.9. *Under Assumption 3.1, using prior $\mu_1 \in \Delta(\mathcal{D})$, discretization error $\varepsilon \in [0, 1)$, and optimism strength $\lambda > 0$, `POPS` (Algorithm 2) achieves regret $\widetilde{O}\big(\lambda K_\star T + \log(1/\mu_1(D_\star))/\lambda + \varepsilon T\big)$.*

Next, we show that, if there exists $D \in \mathcal{D}$ whose smoothed demand function uniformly approximates that of $D_\star$, then the trajectory of `POPS` under $D_\star$ can be coupled with that under D, such that the trajectories coincide with high probability. See Appendix C.7 for the proof.

Lemma 3.10. *If there exists $D \in \mathcal{D}$ for which $\|\mathsf{dem}^{\varepsilon}_D - \mathsf{dem}^{\varepsilon}_{D_\star}\|_\infty \leq \varepsilon$, then the trajectory $\{u_t, \hat{p}_t, y_t\}_{t=1}^T$ of `POPS` with type distribution $D^\star$ can be coupled with that $\{u'_t, \hat{p}'_t, y'_t\}_{t=1}^T$ of `POPS` with type distribution D, such that the two trajectories are identical with probability $1 - \varepsilon T$.*

Finally, we show that, to obtain a uniform ε-cover of the smoothed demand functions, it suffices to find a $O(\varepsilon^2)$-cover of the type distributions under the Lévy metric d_{L}, as defined in (1). Moreover, we show that the family of all type distributions with support size at most K, $\Delta_K(\Theta)$, admits an appropriately small Lévy cover. For notation, we write $N(X, \mathrm{d}, \tau)$ for the size of the smallest subset $X' \subseteq X$ which covers set X under metric d up to accuracy τ (i.e., for each $x \in X$, there exists $x' \in X'$ such that $\mathrm{d}(x, x') \leq \tau$). A proof of the following appears in Appendix C.8.

Lemma 3.11. *If $D, D' \in \Delta(\Theta)$ satisfy $d_{\mathrm{L}}(D, D') \leq \varepsilon^2/2$, then $\|\mathsf{dem}^{\varepsilon}_D - \mathsf{dem}^{\varepsilon}_{D'}\|_\infty \leq \varepsilon$. Moreover, we have $\log N\big(\Delta_K(\Theta), d_{\mathrm{L}}, \varepsilon\big) = \widetilde{O}(dK \log 1/\varepsilon)$.*

In Appendix C.9, we combine these lemmas to obtain the upper bound. For the lower bound, we modify a construction of [4] for the non-contextual case so that it can be tensored into d-dimensions without leaking information between orthogonal contexts.

4 Non-Contextual Refinements via Zooming

Our results from Section 3 leave a key open question: what is the optimal regret dependence on $K_\star$? We resolve this question for the non-contextual setting, where $d = 1$ and, without loss of generality, $u_t \equiv 1$ for all t. To do so, we employ adaptive discretization (aka *zooming*) methods from Lipschitz bandits [13] with novel variance-aware confidence intervals and achieve a regret bound of $\widetilde{O}(\sqrt{K_\star T})$. Throughout, we label the $K_\star$ types $\mathrm{supp}(D_\star) = \{\theta^{(1)} < \theta^{(2)} < \cdots < \theta^{(K_\star)}\}$ and set $\theta^{(0)} = 0$.

Our algorithm, `ZoomV`, mirrors standard zooming [13] with two key adjustments. First, since revenue is only one-sided Lipschitz, we use a dyadic price selection rule inspired by [24]. Second, our confidence intervals incorporate empirical variance, a method previously used for variance-aware K-armed bandits [2]. In more detail, `ZoomV` maintains a set S of active prices in $[0, 1]$ and a variance-aware confidence interval for the expected revenue at each $p \in S$. Each active price "covers" an interval of neighboring, larger prices, with the width of this covering interval scaling proportionally to that of the confidence interval. The intuition is that a small increase in price can only marginally increase expected revenue, so it is not worth exploring such covered prices. Initially, every price in

$[0,1]$ is covered by some point in S. At each round, ZoomV optimistically chooses a price in S and updates its confidence and covering intervals. If after the update there exists an uncovered price, then we add a new point to S which covers it, maintaining the invariant that every price is covered.

Theorem 4.1. ZoomV *(Algorithm 4) achieves regret* $\widetilde{O}\big(\min\{\sqrt{K_\star T}, T^{2/3}\}\big)$ *for non-contextual pricing with heterogeneous buyers, without knowledge of* $K_\star$.

To bound regret, we employ a variance-aware zooming dimension which controls its performance. For comparison, we first recall the definition of the standard zooming dimension, which characterizes a certain complexity of the expected reward function $\mathsf{rev}(p)$. For each $\delta > 0$, write $X_\delta := \{p \in [0,1] : \mathsf{gap}_\star(p) \le \delta\}$ for the set of δ-approximate revenue maximizers. Write $N(X,\delta) := N(X, |\cdot|, \delta)$ for the smallest δ-covering of a set $X \subseteq \mathbb{R}$. Then, for each $c > 0$, define the *zooming dimension*

$$\texttt{ZoomDim}(c) := \inf\{z \ge 0 : N(X_\delta, \delta/10) \le c\delta^{-z}\ \forall \delta > 0\}.$$

To address variance, write $\sigma^2(p) = p^2 \mathsf{dem}_\star(p)(1 - \mathsf{dem}_\star(p))$ for the revenue variance when p is played. We then define the variance-weighted covering number $N_{\mathrm{v}}(X,\delta) := \min\big\{\sum_{x \in X'} \sigma^2(x) : X' \text{ is a } \delta\text{-cover of } X\big\}$. Finally, for each $c > 0$, we define the *variance-aware zooming dimension*

$$\texttt{ZoomDimV}(c) := \inf\{z \ge 0 : N_{\mathrm{v}}(X_\delta, \delta/10) \le c\delta^{-z}\ \forall \delta > 0\}.$$

Note that $\texttt{ZoomDimV}(c) \le \texttt{ZoomDim}(c)$, since $\sigma^2(p) \le 1$.

We prove the following regret bound, mirroring standard zooming guarantees (see Theorem 1.3 of 13), but with ZoomDim improved to ZoomDimV. The proof can be found in Appendix D.1. Then in the lemma afterwards, we bound our modified zooming dimension for the pricing problem.

Lemma 4.2. *For* $c > 0$, ZoomV *achieves regret* $\widetilde{O}\big(c^{1/(2+z)}\, T^{1-1/(2+z)}\big)$, *where* $z = \texttt{ZoomDimV}(c)$.

Lemma 4.3. *We have* $\texttt{ZoomDimV}(10K_\star) = 0$ *and* $\texttt{ZoomDimV}(10) \le \texttt{ZoomDim}(10) \le 1$.

Proof. For each $\delta > 0$ and type $i \in [K_\star]$, let $X_\delta^{(i)}$ denote the set of activated arms p with $\mathsf{gap}_\star(p) \le \delta$ that lie in the interval $(\theta^{(i-1)}, \theta^{(i)}]$, to the left of type i. Since revenue is linearly increasing within each such interval, with slope $d_i = \mathsf{dem}(\theta^{(i)})$, the gap condition requires that each $p \in X_\delta^{(i)}$ also satisfies $p \ge \theta^{(i)} - \delta d_i^{-1}$. Moreover, for $p \in X_\delta^{(i)}$, we have $\sigma^2(p) \le d_i$. Thus, we obtain

$$N_{\mathrm{v}}(X_\delta, \delta/20) \le \sum_{i \in [K_\star]} N_{\mathrm{v}}(X_\delta^{(i)}, \delta/20) \le \sum_{i \in [K_\star]} 10 d_i^{-1} \cdot d_i \le 20 K_\star,$$

implying the first bound. For the second, we note that $N(X_\delta, \delta/20) \le N([0,1], \delta/20) \le 20\delta^{-1}$. $\square$

Proof of Theorem 4.1. The $\sqrt{K_\star T}$ bound follows by Lemma 4.2 with $c = 4K_\star$ and $z = 0$, using Lemma 4.3. Similarly, the $T^{2/3}$ bound follows by taking $c = 2$ and $z = 1$. $\square$

Remark 4.4 (Comparison to [4])**.** *The non-contextual setting was previously studied by [4], whose Algorithm 1 achieves regret* $\widetilde{O}(\sqrt{K_\star T}) + V(V+1)$, *where* $V = \max_{i \in [K_\star]} \big(\theta^{(K_\star)}\big)^4 \big(\theta^{(i)} - \theta^{(i-1)}\big)^{-5}$. *They maintain a set of intervals which contain all types with substantial probability mass, gradually refining these intervals until they are all of width* $O(T^{-1/2})$, *at which point they employ UCB over the intervals' left endpoints. Unfortunately, the instance-dependent term* $V(V+1)$ *can blow up to infinity for worst-case realizations of* $D_\star \in \Delta_{K_\star}([0,1])$, *in contrast to our guarantee.*

Remark 4.5 (Performance of standard zooming)**.** *Although original zooming bounds were for two-sided Lipschitz rewards, [24] noted that these extend to the one-sided case. In particular, Corollary 3.1 of [24] implies the regret bound of Lemma 4.2 with* $z = \texttt{ZoomDim}(c)$. *Inspecting the proof of Lemma 4.3, we have* $\texttt{ZoomDim}(10K_\star d_{\min}^{-1}) = 0$, *where* $d_{\min}$ *is the smallest demand at any of the* $K_\star$ *types. This gives regret* $\widetilde{O}(\sqrt{K_\star T/d_{\min}})$, *which can be arbitrarily larger than* $\sqrt{K_\star T}$. *One could hope that there is still some room for improvement, since small demands for near-optimal prices imply that the best achievable revenue is also small. However, even when* $K_\star = 1$ *and the analysis is the simplest, this leads to an unsatisfactory regret bound of* $\widetilde{O}(\min\{\sqrt{T/d_{\min}}, d_{\min} T\}) = \widetilde{O}(T^{2/3})$.

5 Improved Performance with Ex-Post Type Observability

We study dynamic pricing with heterogeneous buyers, assuming the learner can *identify* each arriving type. After setting a price p_t, the learner observes the pricing feedback $\mathbb{1}\{\langle u_t, \theta_t \rangle \geq p_t\}$ and some information about the sampled type θ_t. We consider the two models of observability. In the first model, the learner only observes an identifier $z_t \in [K_\star]$ that specifies which of the $K_\star$ candidate types was drawn. In practice, the learner need not know $K_\star$ *a priori*. Here, we design an algorithm that matches the $\widetilde{O}(K_\star \sqrt{dT})$ regret bound of `POPS` and can be implemented efficiently, using a contextual search algorithm for $K_\star = 1$ as a subroutine. In the second model, the learner observes the full type embedding $\theta_t \in \Theta$. Here, we show that best-responding to a simple plug-in estimate for $D_\star$ achieves an improved regret bound of $\widetilde{O}(\sqrt{\min\{K_\star, d\}T})$.

Observed type identifiers. Our algorithm for the model where the learner only observes the identifier uses a $K_\star = 1$ contextual search algorithm, `ProjectedVolume` [18], as a subroutine. We maintain a separate instance of this `ProjectedVolume` algorithm for each observed type and keep track of the empirical type frequencies along with the number of times we've explored each type. It then adaptively chooses which types to explore (or exploit) based on confidence estimates for the type distribution. We present the full algorithm and prove the following regret bound in Appendix E.1.

Theorem 5.1. *Consider contextual dynamic pricing with ex-post type observability where the learner observes which type $z_t \in [K_\star]$ arrived. Then, Algorithm 5 achieves regret $\widetilde{O}(K_\star \sqrt{dT})$ and takes no more than time $\mathrm{poly}(K_\star, d, T)$ per round.*

Observed type vectors. If the full type vector θ_t is revealed at the end of each round, we can achieve improved regret with a simpler algorithm. Indeed, writing $\hat{D}_t = \frac{1}{t} \sum_{\tau=1}^{t} \delta_t$ for the empirical type distribution after round t, we take each p_t as the best response to $\hat{D}_{t-1}$ along the current context.

Theorem 5.2. *Consider contextual dynamic pricing with ex-post type observability where the learner observes $\theta_t \in \Theta$ at the end of each round. Then the algorithm which plays $p_1 = 1/2$ and best response $p_t = \mathsf{br}_{\hat{D}_{t-1}}(u_t)$ for remaining rounds achieves regret $\widetilde{O}(\sqrt{\min\{K_\star, d\}T})$. Each price can be computed in time $\mathrm{poly}(K_\star, d)$.*

The proof in Appendix E.2 observes that the empirical revenue function $\mathsf{rev}_{\hat{D}_t}$ converges uniformly in both arguments (price and context) to the true revenue function $\mathsf{rev}_\star$. This occurs at rate $\sqrt{V/t}$, where V is the VC dimension of the function class $\{\theta \in \mathrm{supp}(D_\star) \mapsto \mathbb{1}\{\langle u, \theta \rangle \geq p\} : u \in \Theta, p \in [0, 1]\}$. This is a family of linear thresholds in $\mathbb{R}^d$, so $V \leq d + 1$. Moreover, the domain has size $K_\star$, so $V \leq K_\star$. We thus obtain the claimed regret bound of $\widetilde{O}\big(\sum_{t=1}^{T} \sqrt{\min\{K_\star, d\}/t}\big) = \widetilde{O}\big(\sqrt{\min\{K_\star, d\}T}\big)$.

6 Discussion

In this work, we have introduced contextual dynamic pricing with heterogeneous buyers. Our main algorithm achieves a regret bound of $\tilde{O}(K^\star \sqrt{dT})$, which is optimal up to $\sqrt{K^\star}$. Our analysis bounds the disagreement coefficient by leveraging a novel decomposition lemma for aggregate demand functions with $K_\star$ breakpoints. Additionally, we propose a variance-aware zooming algorithm for the non-contextual pricing case and improve regret dependence on $K_\star$ by incorporating adaptive discretization methods from the Lipschitz bandits literature. Finally, under stronger observability assumptions on the buyers' types, we develop efficient algorithms that significantly reduce regret to $\tilde{O}(\sqrt{\min\{K_\star, d\}T})$, demonstrating the benefits of richer feedback in dynamic pricing settings.

There are several research questions that stem from our work. The first question revolves around *computation*. Our algorithms are dependent on the size of the model class, which is exponential in $K_\star, d, T$. It would be interesting to know if polynomial dependence can be achieved at the same regret rate. The second question is around the optimal dependence on $K_\star$ for the general, contextual case. While in Section 4 we showed how to optimize the dependence of our bounds on $K_\star$ it is unclear how to scale this approach for the contextual version of the problem. One starting point could be the results on Zooming techniques for contextual bandits (see e.g., [26]). Finally, it would be interesting to see if our results can be applied to broader settings where a learner tries to learn from heterogeneous agents while receiving only binary feedback: for example, it is unknown if the approach presented in this work generalizes to general contextual search settings (i.e., with ε-ball or symmetric loss) or if it generalizes for settings that share some core properties with pricing, but differ in the fundamental techniques used to address them (see e.g., [10]).

References

[1] K. Amin, A. Rostamizadeh, and U. Syed. Repeated contextual auctions with strategic buyers. *Advances in Neural Information Processing Systems*, 27, 2014.

[2] J.-Y. Audibert, R. Munos, and C. Szepesvári. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19):1876–1902, 2009.

[3] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013.

[4] N. Cesa-Bianchi, T. Cesari, and V. Perchet. Dynamic pricing with finitely many unknown valuations. In A. Garivier and S. Kale, editors, *Algorithmic Learning Theory (ALT)*, 2019.

[5] M. C. Cohen, I. Lobel, and R. Paes Leme. Feature-based dynamic pricing. *Management Science*, 66(11):4921–4943, 2020.

[6] J. Fan, Y. Guo, and M. Yu. Policy optimization using semiparametric models for dynamic pricing. *Journal of the American Statistical Association*, 119(545):552–564, 2024.

[7] D. Foster, A. Rakhlin, D. Simchi-Levi, and Y. Xu. Instance-dependent complexity of contextual bandits and reinforcement learning: A disagreement-based perspective. In *Proceedings of Thirty Fourth Conference on Learning Theory (COLT)*, 2021.

[8] D. J. Foster, S. M. Kakade, J. Qian, and A. Rakhlin. The statistical complexity of interactive decision making. *arXiv preprint arXiv:2112.13487*, 2021.

[9] S. Hanneke et al. Theory of disagreement-based active learning. *Foundations and Trends® in Machine Learning*, 7(2-3):131–309, 2014.

[10] C.-J. Ho, A. Slivkins, and J. W. Vaughan. Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 359–376, 2014.

[11] A. Javanmard. Perishability of data: dynamic pricing under varying-coefficient models. *Journal of Machine Learning Research*, 18(53):1–31, 2017.

[12] A. Javanmard and H. Nazerzadeh. Dynamic pricing in high-dimensions. *Journal of Machine Learning Research*, 20(9):1–49, 2019.

[13] R. Kleinberg, A. Slivkins, and E. Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, page 681–690, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605580470. doi: 10.1145/1374376.1374475. URL https://doi.org/10.1145/1374376.1374475.

[14] R. D. Kleinberg and F. T. Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Symposium on Foundations of Computer Science (FOCS)*, pages 594–605, 2003.

[15] A. Krishnamurthy, J. Langford, A. Slivkins, and C. Zhang. Contextual bandits with continuous actions: Smoothing, zooming, and adapting. *Journal of Machine Learning Research*, 21(137): 1–45, 2020.

[16] A. Krishnamurthy, T. Lykouris, C. Podimata, and R. Schapire. Contextual search in the presence of irrational agents. In *ACM SIGACT Symposium on Theory of Computing (STOC)*, 2021.

[17] A. Liu, R. P. Leme, and J. Schneider. *Optimal Contextual Pricing and Extensions*. 2021.

[18] I. Lobel, R. Paes Leme, and A. Vladu. Multidimensional binary search for contextual decision-making. *Operations Research*, 66(5):1346–1361, 2018.

[19] Y. Luo, W. W. Sun, and Y. Liu. Distribution-free contextual dynamic pricing. *Mathematics of Operations Research*, 49(1):599–618, 2024.

[20] A. Maurer and M. Pontil. Empirical bernstein bounds and sample variance penalization. *Conference on Learning Theory (COLT)*, 2009.

[21] R. Paes Leme and J. Schneider. Contextual search via intrinsic volumes. *SIAM Journal on Computing*, 51(4):1096–1125, 2022.

[22] R. Paes Leme, C. Podimata, and J. Schneider. Corruption-robust contextual search through density updates. In *Conference on Learning Theory (COLT)*, volume 178, 2022.

[23] R. Paes Leme, B. Sivan, Y. Teng, and P. Worah. Pricing query complexity of revenue maximization. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 399–415. SIAM, 2023.

[24] C. Podimata and A. Slivkins. Adaptive discretization for adversarial lipschitz bandits. In *Conference on Learning Theory*, pages 3788–3805. PMLR, 2021.

[25] S. Shalev-Shwartz and S. Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[26] A. Slivkins. Contextual bandits with similarity information. In *Proceedings of the 24th annual Conference On Learning Theory*, pages 679–702. JMLR Workshop and Conference Proceedings, 2011.

[27] A. Slivkins et al. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.

[28] J. Xu and Y.-X. Wang. Towards agnostic feature-based dynamic pricing: Linear policies vs linear valuation with unknown noise. In *International Conference on Artificial Intelligence and Statistics*, pages 9643–9662. PMLR, 2022.

[29] T. Zhang. Feel-good thompson sampling for contextual bandits and reinforcement learning. *SIAM Journal on Mathematics of Data Science*, 4(2):834–857, 2022.

[30] S. Zuo. Corruption-robust lipschitz contextual search. In C. Vernade and D. Hsu, editors, *Proceedings of The 35th International Conference on Algorithmic Learning Theory*, volume 237 of *Proceedings of Machine Learning Research*, pages 1234–1254. PMLR, 25–28 Feb 2024.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We summarize our key contributions and results, which all have proofs.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [NA]

   Justification: This is a theoretical work and we point out all necessary assumptions for the applicability of the results

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We provide the full set of assumptions and a complete proof.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [NA]

Justification: The paper does not contain any experimental work.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
    (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
    (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
    (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
    (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: The paper does not contain any experimental work.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: The paper does not contain any experimental work.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: The paper does not contain any experimental work.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: The paper does not contain any experimental work.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: We abide by the Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: This is a theoretical work without any clear and immediate social impact.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to

generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not release new assets

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.

Supplementary material for *Contextual Dynamic Pricing with Heterogeneous Buyers*

A Extended Discussion on Related Work

Our work relates closely to three lines of work outlined below.

(I) Contextual pricing/search. The closest line to our work is *contextual pricing/search*. In contextual search, there is a repeated interaction between a learner and nature, where the learner is trying to learn a hidden vector $\theta^\star \in \mathbb{R}^d$ over time while receiving only single-bit feedback. Mathematically, at each round $t \in [T]$, the learner receives a (potentially adversarially chosen) context $u_t \in \mathbb{R}^d$ and decides to query $y_t \in \mathbb{R}$. The learner receives feedback $\sigma_t = \mathrm{sgn}(\langle u_t, \theta^\star \rangle - y_t) \in \{-1, +1\}$ and incurs loss $\ell_t(y_t, \langle u_t, \theta^\star \rangle)$ [5]. The learner does *not* observe $\ell_t(y_t, \langle u_t, \theta^\star \rangle)$; only the binary feedback σ_t. Under this framework, contextual *pricing* corresponds to the case where the buyer population is *homogeneous* and the loss function ℓ_t corresponds to the lost revenue as a result of posting price y_t (aka the "pricing loss"). Contextual search works have also considered two other loss functions: the symmetric/absolute loss $\ell_t(y_t, \langle u_t, \theta^\star \rangle) := |y_t - \langle u_t, \theta^\star \rangle|$; and the ε-ball loss $\ell_t(y_t, \langle u_t, \theta^\star \rangle) = \mathbb{1}\{|y_t - \langle u_t, \theta^\star \rangle| > \varepsilon\}$, which are motivated by settings other than the pricing one.

There have been two approaches in the literature for learning in contextual pricing and contextual search (for homogeneous agents/buyers). The first approach (e.g., 5, 18, 21, 17) employs a version of multidimensional binary search: specifically, the algorithms maintain a "knowledge set" with all the possible values of $\theta^\star$ which are "consistent" with the feedback that nature has given thus far. Similar to traditional binary search, the query point is chosen to be the point that (given the nature's feedback) will eliminate roughly half of the current knowledge set. As the knowledge set shrinks, the learner ends up with a small knowledge set for the possible values of $\theta^\star$; this is enough to guarantee sublinear regret. The series of works in [5, 18, 21, 17] optimized regret bounds for the three different loss functions (i.e., symmetric, ε-ball, and pricing). The specific algorithms were different at each paper, but they all maintained this "binary search" flavor. Most of the algorithms employing a multidimensional binary search approach can be "robustified" to very little noise in the agents' responses; since the learner will irrevocably shrink the knowledge set according to the feedback received from nature, they can only afford very few mistakes.

The second approach (e.g., 12, 11, 6, 19) focuses exclusively on pricing settings. This approach uses regression-based algorithms for learning the correct price and *needs to assume* stochastic noise in the buyers' responses. There have also been works studying other aspects of contextual pricing (e.g., strategic agents [1] and unknown noise distribution [28]). Apart from the methodological differences with our work, both streams of literature focus on a *homogeneous* agent population and cannot be readily adapted for a *heterogeneous* population setting.

Moving closer to the heterogeneous agents problem, Krishnamurthy et al. [16] studied "corruption-robust" contextual search, where the agent population is mostly homogeneous, except for $C = o(T)$ corrupted agent responses. Their regret bounds were subsequently strengthened in [22], but the latter only works for contextual search with absolute and ε-ball loss (i.e., does *not* cover the pricing loss). This model has been also studied with a Lipschitz target function [30]. Learning with corruptions can be seen as a first step towards learning from heterogeneous agents, but the approaches above do not scale appropriately for truly heterogeneous agent populations. In contrast, we focus on *fully heterogeneous* settings, where we do not constrain the number or the size of the different buyer types.

(II) Non-contextual pricing. The special case $d = 1$ was introduced by [14], who studied non-contextual dynamic pricing for a homogeneous, stochastic, and advetrsarial buyer population. For the adversarial buyer population, the authors assumed that there can be T different valuations and showed tight regret bounds of $\tilde{O}(T^{2/3})$. In contrast, in our setting, we assume that users are "clustered" in $K_\star$ types), and so the lower bound of $\Omega(T^{2/3})$ of Kleinberg and Leighton [14] does not apply.

The closest to our work is the work of [4], who consider pricing a heterogeneous agent population with an unknown number of types, but the types are still limited to be less than $o(T)$. Throughout the paper, we discuss how their bounds relate to ours for the special case of $d = 1$. None of the aforementioned techniques readily generalize to contextual pricing settings.

(III) Lipschitz bandits. Finally, our work is related to the literature on Lipschitz bandits. Although the pricing loss is *not* fully Lipschitz, it has recently been observed that it satisfies a *one-sided Lipschitzness*. This allows us to leverage techniques from adaptive discretization [13] to obtain improved bounds for $d = 1$. Zooming had previously been applied to pricing (see, e.g., [24]), but these algorithms are insufficient for the $K_\star$-types setting. Indeed, their performance scales with a zooming dimension `ZoomDim` that is too large here. Our `ZoomV` uses variance-aware confidence intervals so that its performance scales with a smaller, variance-aware zooming dimension `ZoomDimV`.

Finally, [15] consider contextual bandits with continuous action spaces, which encompasses the setting of this work. Their regret bounds cover the setting where, for a fixed context, the expected reward is Lipschitz in the learner's action. Although their analysis can be adapted to the one-sided Lipschitz setting of contextual pricing, their results either require stochastic contexts or incur large regret due to naïve discretization. Even in the stochastic case, their regret bound scales with a policy zooming coefficient which does not appear to admit a useful bound in terms of $K_\star$.

B Proofs for Section 2

B.1 Proof of Lemma 2.1

We simply bound $\mathsf{rev}_Q(p') - \mathsf{rev}_Q(p) = p'\mathsf{dem}_Q(p') - p\mathsf{dem}_Q(p) \leq \mathsf{dem}_Q(p)(p' - p) \leq p' - p$, using monotonicity of demand functions. $\square$

B.2 Proof of Lemma 2.2

By the definition of d_{L}, it suffices to prove the lemma when $d = 1$. The first revenue lower bound holds by Lemma 2.1. For the second, omitting the context u and writing $[x]_+ = \max\{x, 0\}$, we use the Lévy metric guarantee to bound

$$\begin{aligned}
\mathsf{rev}_{D'}(\pi) &= [\mathsf{br}_D - \varepsilon]_+ \,\mathsf{dem}_{D'}([\mathsf{br}_D - \varepsilon]_+) \\
&\geq [\mathsf{br}_D - \varepsilon]_+ \,\mathsf{dem}_{D'}(\mathsf{br}_D - \varepsilon) && \text{(demand equals 1 for } p \leq 0) \\
&\geq [\mathsf{br}_D - \varepsilon]_+ \,(\mathsf{dem}_D(\mathsf{br}_D - 2\varepsilon) - \varepsilon) && (d_{\mathrm{L}} \text{ bound}) \\
&\geq [\mathsf{br}_D - \varepsilon]_+ \,(\mathsf{dem}_D(\mathsf{br}_D) - \varepsilon) && \text{(monotonicity of } \mathsf{dem}_D) \\
&\geq \mathsf{rev}_D(\mathsf{br}_D) - 2\varepsilon \\
&\geq \mathsf{rev}_D(\mathsf{br}_{D'}) - 2\varepsilon && (\mathsf{br}_D \text{ maximizes } \mathsf{rev}_D) \\
&= \mathsf{br}_{D'} \,\mathsf{dem}_D(\mathsf{br}_{D'}) - 2\varepsilon \\
&\geq \mathsf{br}_{D'} \,(\mathsf{dem}_{D'}(\mathsf{br}_{D'} - \varepsilon) - \varepsilon) - 2\varepsilon && (d_{\mathrm{L}} \text{ bound}) \\
&\geq \mathsf{br}_{D'} \,\mathsf{dem}_{D'}(\mathsf{br}_{D'}) - 3\varepsilon \\
&= \mathsf{rev}_{D'}(\mathsf{br}_{D'}) - 3\varepsilon,
\end{aligned}$$

as desired. $\square$

C Proofs for Section 3

To unify analysis of Sections 3.1 and 3.2, we introduce a more general problem setup and algorithm.

C.1 Generalized Problem Setup

To start, we replace the price set $[0, 1]$ with a subset $\mathcal{P} \subseteq [0, 1]$, which will remain $[0, 1]$ in Section 3.1 but will be restricted to a finite set for Section 3.2. Then, instead of selecting a type distribution $D_\star$, we have the adversary choose a demand function $f^\star$ which maps price $p \in \mathcal{P}$ and context $u \in \mathcal{U}$ to a purchase probability $f^\star(p, u) \in [0, 1]$. Then, at round t, if the adversary selects context $u_t \in \mathcal{U}$ and the learner posts price $p_t \in \mathcal{P}$, purchase decision $y_t \in \{0, 1\}$ is sampled independently from $\mathrm{Ber}(f^\star(p_t, u_t))$. This abstracts away our previous notions of buyer types and values and will also model the smoothed environment of Section 3.2. We impose the corresponding notion of realizability.

Setting C.1 (realizability, general)**.** Under the setup described above, the demand function $f^\star$ belongs to a known, finite class $\mathcal{F}$ of measurable functions from $\mathcal{P} \times \mathcal{U}$ to $[0, 1]$.

Often, we shall fix a context and consider univariate (non-contextual) demand functions. Given a univariate demand function $g : [0,1] \to [0,1]$, we define the corresponding revenue function $\mathsf{rev}_g(p) = p \cdot g(p)$, best-response $\mathsf{br}_g = \arg\max_{p \in \mathcal{P}} \mathsf{rev}_g$ (breaking ties arbitrarily), and gap $\mathsf{gap}_g(p) = \mathsf{rev}_g(\mathsf{br}_g) - \mathsf{rev}_g(p)$. For a contextual demand function $f : [0,1] \times \mathcal{U} \to [0,1]$ and a context $u \in \mathcal{U}$, write $\mathsf{proj}(f, u)$ for the induced univariate demand function $p \mapsto f(p, u)$. We define $\mathsf{rev}_f(p, u) = \mathsf{rev}_{\mathsf{proj}(f,u)}(p)$, $\mathsf{br}_f(u) = \mathsf{br}_{\mathsf{proj}(f,u)}$, $\mathsf{gap}_f(p, u) = \mathsf{gap}_{\mathsf{proj}(f,u)}(p)$, and $\mathsf{proj}(\mathcal{F}, u) := \{\mathsf{proj}(f, u) : f \in \mathcal{F}\}$, along with $\mathsf{rev}_\star := \mathsf{rev}_{f^\star}$, $\mathsf{br}_\star := \mathsf{br}_{f^\star}$, and $\mathsf{gap}_\star := \mathsf{gap}_{f^\star}$. Finally, we define

$$\mathsf{dis}(\mathcal{F}, f_\star) := \sup_{u \in \mathcal{U}} \mathsf{dis}\big(\{f(\cdot, u) - f_\star(\cdot, u) : f \in \mathcal{F}\}\big),$$

generalizing the definition in Section 3.1. Regret is defined as the sum of gaps $\sum_{t=1}^T \mathsf{gap}_\star(p_t, u_t)$.

C.2 Generalized OPS and its Regret Guarantee

We now present a extension of `OPS` and `POPS` to the generalized setup of Appendix C.1. Both can be recovered for appropriate choices of $\mathcal{F}$ and $\mathcal{P}$, which we will discuss later. First, for a univariate demand function $g : \mathcal{P} \to [0,1]$, price p, and purchase decision y, we define loss

$$\ell_\lambda(g, p, y) := (g(p) - y)^2 - \lambda \mathsf{rev}_g(\mathsf{br}_g). \tag{3}$$

We now adapt OPS to this setting, introducing `GOPS` (Algorithm 3).

ALGORITHM 3: `GOPS`: Generalized OPS for Contextual Pricing with Finite Model Class

Input: finite demand function class $\mathcal{F}$, model prior $\mu_1 \in \Delta(\mathcal{F})$, optimism strength $\lambda > 0$;
for *each round* $t \in [T]$ **do**
  observe context u_t;
  sample demand function $f_t \sim \mu_t$;
  play $p_t = \mathsf{br}_{f_t}(u_t)$ and observe y_t;
  update $\mu_{t+1}(f) \propto \mu_t(f) \exp\big(-\ell_\lambda(\mathsf{proj}(f, u_t), p_t, y_t)\big)$ for each $f \in \mathcal{F}$;

We prove the following regret bound.

Lemma C.2. *Under Setting C.1, with model prior $\mu_1 \in \Delta(\mathcal{F})$ and optimism strength $\lambda \geq 4/T$,* `GOPS` *(Algorithm 3) achieves regret $25\lambda(\mathsf{dis}(\mathcal{F}, f_\star) \vee 1) T \log^2(T) + \log(1/\mu_1(f_\star))/\lambda$.*

Our proof employs the following decoupling lemma.

Lemma C.3 (8)**.** *Let $\mathcal{G}$ be a finite family of univariate demand functions and fix $g^\star \in \mathcal{G}$. Then, for any $\nu \in \Delta(\mathcal{G})$ and $\gamma > 0$, we have*

$$\mathop{\mathbb{E}}_{g \sim \nu}\left[|g(\mathsf{br}_g) - g^\star(\mathsf{br}_g)|\right] \leq 6 \frac{\mathsf{dis}(\mathcal{G}) \log^2(\gamma \vee e)}{\gamma} + \gamma \mathop{\mathbb{E}}_{\tilde{g}, g \sim \nu}\left[(\tilde{g}(\mathsf{br}_g) - g^\star(\mathsf{br}_g))^2\right].$$

This is simply Lemma E.2 of [8] with function class $\{g - g^\star : g \in \mathcal{G}\}$ and $\Delta \to 0$. In our proof, $\mathcal{G}$ and $g^\star$ will be the projections of $\mathcal{F}$ and $f^\star$ onto a fixed context $u \in \mathcal{U}$.

The remainder of our analysis is a slight modification to that of [29], which we provide for completeness. For each round $t \in [T]$ of OPS, we adopt the following notation:

- history up to round t: $S_t := \{u_\tau, f_\tau, p_\tau, y_\tau\}_{\tau=1}^t$,
- true univariate demand function: $g_t^\star := \mathsf{proj}(f^\star, u_t)$,
- univariate demand function posterior: $\nu_t := \mathsf{proj}(\mu_t, u_t) := \mathrm{Law}_{f \sim \mu_t}(\mathsf{proj}(f, u_t))$,
- sampled univariate demand function: $g_t := \mathsf{proj}(f_t, u_t)$, so that $p_t = \mathsf{br}_{g_t}$,
- independently sampled univariate demand function (for analysis): $\tilde{g}_t \sim \nu_t$,
- regret: $\mathrm{REG}_t := \mathsf{rev}_\star(\mathsf{br}_\star(u_t), u_t) - \mathsf{rev}_\star(p_t, u_t) = \mathsf{rev}_{g_t^\star}(p_t) - \mathsf{rev}_{g_t^\star}(p_t)$,
- least-squares errors: $\mathrm{LS}_t(g) := (g(p_t) - g^\star(p_t))^2$,
- "feel-good" (optimism) bonuses: $\mathrm{FG}_t(g) := \mathsf{rev}_g(\mathsf{br}_g) - \mathsf{rev}_{g_t^\star}(\mathsf{br}_{g_t^\star})$,

- loss discrepancies: $\Delta L_t(g) := \ell_\lambda(g, p_t, y_t) - \ell_\lambda(g_t^\star, p_t, y_t)$,
- potential function: $Z_t := \mathbb{E}_{S_t} \log \mathbb{E}_{f\sim\mathrm{Unif}(\mathcal{F})} \exp\big(-\sum_{\tau=1}^t \Delta L_t(\mathsf{proj}(f, u_t))\big)$.

Our proof requires several supporting lemmas. The first is a basic concentration result.

Lemma C.4. *For $c \geq 0$ and a random variable X supported on $[0,1]$, we have $\log \mathbb{E} \exp(-cX) \leq (\frac{1}{2}c^2 - c)\,\mathbb{E}\,X$. For X supported on $[a,b]$, we have $\log \mathbb{E}\exp(cX) \leq c\,\mathbb{E}\,X + \frac{1}{8}(b-a)^2 c^2$.*

Proof. For the first inequality, we bound

$$\log \mathbb{E}\exp(-cX) \leq \mathbb{E}[\exp(-cX) - 1] \leq \mathbb{E}[-cX + \tfrac{1}{2}c^2X^2] \leq \mathbb{E}[-cX + \tfrac{1}{2}c^2 X] = (\tfrac{1}{2}c^2 - c)\,\mathbb{E}\,X.$$

The second inequality is exactly Hoeffding's lemma. ☐

The next lemma mirrors Lemma 4 of [29]. This is a consequence of the definition of ΔL_t and the sub-Gaussianity of its components.

Lemma C.5. *For round t of `GOPS` (Algorithm 3), we have*

$$\log \mathop{\mathbb{E}}_{g\sim\nu_t} \mathop{\mathbb{E}}_{y_t|u_t,p_t} \exp\big(-\Delta L_t(g)\big) \leq -\frac{1}{4} \mathop{\mathbb{E}}_{g\sim\nu_t} \mathrm{LS}_t(g) + \lambda \mathop{\mathbb{E}}_{g\sim\nu_t} \mathrm{FG}_t(g) + \frac{3}{2}\lambda^2.$$

We note that this lemma does not rely on how p_t is selected.

Proof. Let $g \sim \nu_t$ and $y \sim \mathrm{Law}(y_t|u_t, p_t) = \mathrm{Ber}(g_t(p_t))$ be independent. Let $\varepsilon = y - g_t^\star(p_t)$ denote the discrepancy between the observed and expected demand. Since demands lie in $[0,1]$, Lemma C.4 with $X = \varepsilon$ gives

$$\mathop{\mathbb{E}}_{y} \exp\big(-2\varepsilon\big(g_t^\star(p_t) - g(p_t)\big)\big) \leq \exp\big(\tfrac{1}{2}\big(g_t^\star(p_t) - g(p_t)\big)^2\big) = \exp\big(\tfrac{1}{2}\mathrm{LS}_t(g)\big). \tag{4}$$

Moreover, we have

$$\begin{aligned} -\Delta L_t(g) &= -\big(\varepsilon + g_t^\star(p_t) - g(p_t)\big)^2 + \varepsilon^2 + \lambda \mathrm{FG}_t(g) \\ &= -2\varepsilon\big(g_t^\star(p_t) - g(p_t)\big) - \mathrm{LS}_t(g) + \lambda\mathrm{FG}_t(g). \end{aligned}$$

Combining with (4) gives

$$\mathop{\mathbb{E}}_{y} \exp\big(-\Delta L_t(g)\big) \leq \exp\big(-\tfrac{1}{2}\mathrm{LS}_t(g) + \lambda \mathrm{FG}_t(g)\big).$$

Therefore, we have

$$\begin{aligned} \log \mathop{\mathbb{E}}_{g,y} \exp\big(-\Delta L_t(g)\big) &\leq \log \mathop{\mathbb{E}}_{g} \exp\big(-\tfrac{1}{2}\mathrm{LS}_t(g) + \lambda\mathrm{FG}_t(g)\big) \\ &\leq \frac{2}{3} \log \mathop{\mathbb{E}}_{g} \exp\big(-\tfrac{3}{4}\mathrm{LS}_t(g)\big) + \frac{1}{3}\log \mathop{\mathbb{E}}_{g} \exp\big(3\lambda \mathrm{FG}_t(g)\big), \end{aligned}$$

where the last inequality follows by Hölder's inequality. For the first term, we use Lemma C.4 with $X = \mathrm{LS}_t(g)$ and $c = \frac{3}{4}$ to bound

$$\frac{2}{3}\log\Big(\mathop{\mathbb{E}}_{g} \exp\big(-\tfrac{3}{4}\mathrm{LS}_t(g)\big)\Big) \leq \frac{2}{3}\Big(\frac{1}{2}c^2 - c\Big) \mathop{\mathbb{E}}_{g} \mathrm{LS}_t(g) = -\frac{5}{16} \mathop{\mathbb{E}}_{g} \mathrm{LS}_t(g).$$

For the second term, we apply the lemma with $X = \mathrm{FG}_t(g)$ and $c = 3\lambda$ to obtain

$$\frac{1}{3}\log \mathop{\mathbb{E}}_{g} \exp\big(3\lambda\mathrm{FG}_t(g)\big) \leq \lambda \mathop{\mathbb{E}}_{g} \mathrm{FG}_t(g) + \frac{3}{2}\lambda^2.$$

Combining, we have

$$\log \mathop{\mathbb{E}}_{g,y} \exp\big(-\Delta L_t(g)\big) \leq -\frac{5}{16} \mathop{\mathbb{E}}_{g} \mathrm{LS}_t(g) + \lambda \mathop{\mathbb{E}}_{g} \mathrm{FG}_t(g) + \frac{3}{2}\lambda^2,$$

implying the lemma. ☐

Our last helper lemma mirrors Lemma 5 of [29].

Lemma C.6. *For round t of* `GOPS` *(Algorithm 3), we have*

$$\frac{1}{4\lambda}\,\mathbb{E}\,\mathrm{LS}_t(\tilde{g}_t) - \mathbb{E}\,\mathrm{FG}_t(g_t) \le \frac{3}{2}\lambda + \frac{1}{\lambda}(Z_{t-1} - Z_t).$$

This lemma also does not rely on how prices are chosen.

Proof. Recall that $\mu_1 = \mathrm{Unif}(\mathcal{F})$. Defining $W_t(f \mid S_t) := \exp\big(-\sum_{\tau=1}^{t} \Delta L_\tau(\mathsf{proj}(f, u_\tau))\big)$, we have $Z_t = \mathbb{E}_{S_t} \log \mathbb{E}_{f\sim\mu_1} W_t(f \mid S_t)$. Note that

$$\mu_t(f) = \frac{W_{t-1}(f \mid S_{t-1})}{\mathbb{E}_{f\sim\mu_1} W_{t-1}(f \mid S_{t-1})}\mu_1(f).$$

We then bound

$$\begin{aligned}
Z_t &= Z_{t-1} + \mathop{\mathbb{E}}_{S_t} \log \frac{\mathbb{E}_{f\sim\mu_1} W_t(f \mid S_t)}{\mathbb{E}_{f\sim\mu_1} W_{t-1}(f \mid S_{t-1})} \\
&= Z_{t-1} + \mathop{\mathbb{E}}_{S_t} \log \frac{\mathbb{E}_{f\sim\mu_1} W_{t-1}(f \mid S_{t-1}) \exp\big(-\Delta L_t(\mathsf{proj}(f, u_t))\big)}{\mathbb{E}_{f\sim\mu_1} W_{t-1}(f \mid S_{t-1})} \\
&= Z_{t-1} + \mathop{\mathbb{E}}_{S_t} \log \mathop{\mathbb{E}}_{f\sim\mu_t} \exp\big(-\Delta L_t(\mathsf{proj}(f, u_t))\big) \\
&\overset{(a)}{\le} Z_{t-1} + \mathop{\mathbb{E}}_{S_{t-1}, u_t, p_t} \log \mathop{\mathbb{E}}_{y_t \mid u_t, p_t} \mathop{\mathbb{E}}_{g\sim\nu_t} \exp\big(-\Delta L_t(g)\big) \\
&\overset{(b)}{\le} Z_{t-1} - \frac{1}{4}\,\mathbb{E}\,\mathrm{LS}_t(\tilde{g}_t) + \lambda\,\mathbb{E}\,\mathrm{FG}_t(g_t) + \frac{3}{2}\lambda^2,
\end{aligned}$$

where (a) uses Jensen's inequality and (b) uses Lemma C.4. Rearranging gives the lemma. □

Now, we return to Lemma C.2, where we will finally incorporate our price selection rule and the decoupling lemma.

Proof of Lemma C.2 For each round $t \in [T]$, we recall that $p_t = \mathsf{br}_{g_t}$ and decompose

$$\begin{aligned}
\mathrm{REG}_t &= \mathsf{rev}_{g_t^\star}(\mathsf{br}_{g_t^\star}) - \mathsf{rev}_{g_t^\star}(p_t) \\
&= \big[\mathsf{rev}_{g_t}(p_t) - \mathsf{rev}_{g_t^\star}(p_t)\big] - \big[\mathsf{rev}_{g_t}(\mathsf{br}_{g_t}) - \mathsf{rev}_{g_t^\star}(\mathsf{br}_{g_t^\star})\big] \\
&= \big[\mathsf{rev}_{g_t}(p_t) - \mathsf{rev}_{g_t^\star}(p_t)\big] - \mathrm{FG}_t(g_t).
\end{aligned}$$

Conditioning on S_{t-1} and u_t, we bound with Lemma C.3 using $\gamma = \frac{1}{4\lambda}$

$$\begin{aligned}
\mathbb{E}\big[\mathsf{rev}_{g_t}(p_t) - \mathsf{rev}_{g_t^\star}(p_t) \,\big|\, S_{t-1}, u_t\big] &= \mathop{\mathbb{E}}_{g\sim\nu_t}\big[\mathsf{rev}_g(\mathsf{br}_g) - \mathsf{rev}_{g_t^\star}(\mathsf{br}_g)\big] \\
&\le \mathop{\mathbb{E}}_{g\sim\nu_t}\big[|g(\mathsf{br}_g) - g_t^\star(\mathsf{br}_g)|\big] \\
&\le 24\lambda\,\mathsf{dis}(\mathcal{F}, f_\star)\log^2(4\lambda^{-1} \vee e) + \frac{1}{4\lambda}\mathop{\mathbb{E}}_{\tilde{g}, g\sim\nu}\big[(\tilde{g}(\mathsf{br}_g) - g_t^\star(\mathsf{br}_g))^2\big] \\
&= 24\lambda\,\mathsf{dis}(\mathcal{F}, f_\star)\log^2(4\lambda^{-1} \vee e) + \frac{1}{4\lambda}\,\mathbb{E}\big[\mathrm{LS}_t(\tilde{g}_t) \,\big|\, S_{t-1}, u_t\big].
\end{aligned}$$

Taking expectations over S_{t-1} and u_t, we bound

$$\begin{aligned}
\mathbb{E}\,\mathrm{REG}_t &\le 24\lambda\,\mathsf{dis}(\mathcal{F}, f_\star)\log^2(4\lambda^{-1} \vee e) + \frac{1}{4\lambda}\,\mathbb{E}\,\mathrm{LS}_t(\tilde{g}_t) - \mathbb{E}\,\mathrm{FG}_t(g_t) \\
&\le 24\lambda\,\mathsf{dis}(\mathcal{F}, f_\star)\log^2(4\lambda^{-1} \vee e) + \frac{3}{2}\lambda + \frac{1}{\lambda}(Z_{t-1} - Z_t),
\end{aligned}$$

using Lemma C.6. Summing over $t \in [T]$ and noting that $Z_0 = 0$, we bound

$$\mathbb{E}[R(T)] \le T\left(24\,\lambda\mathsf{dis}(\mathcal{F}, f_\star)\log^2(4\lambda^{-1} \vee e) + \frac{3}{2}\lambda\right) - \frac{1}{\lambda}Z_T.$$

Moreover, by realizability, we have

$$Z_T \geq \mathop{\mathbb{E}}_{S_t} \log(\mu_1(f^\star) W_T(f^\star \mid S_t)) = \log(\mu_1(f_\star)).$$

Combining, we obtain

$$\begin{aligned}
\mathbb{E}[R(T)] &\leq T\left(24\lambda\, \mathsf{dis}(\mathcal{F}, f_\star) \log^2(4\lambda^{-1} \vee e) + \frac{3}{2}\lambda\right) - \frac{\log(\mu_1(f_\star))}{\lambda} \\
&\leq T\big(25\lambda(\mathsf{dis}(\mathcal{F}, f_\star) \vee 1) \log^2(4\lambda^{-1} \vee e)\big) - \frac{\log(\mu_1(f_\star))}{\lambda}, \\
&\leq T\big(25\lambda(\mathsf{dis}(\mathcal{F}, f_\star) \vee 1) \log^2 T\big) - \frac{\log(\mu_1(f_\star))}{\lambda},
\end{aligned}$$

as desired. □

C.3 Proof of Lemma 3.3

Under the general setup of Appendix C.1, we take $\mathcal{F}$ to be the class of demand functions induced by $\mathcal{D}$, set $\mathcal{P} = [0, 1]$, and fix $\mu_1 = \mathrm{Unif}(\mathcal{F})$. By these choices, `GOPS` coincides exactly with `OPS`, as does our notion of regret. Thus, Lemma C.2 gives the desired regret bound of

$$T\big(25\lambda(\mathsf{dis}(\mathcal{D}, D_\star) \vee 1) \log^2 T\big) - \frac{\log(\mu_1(D_\star))}{\lambda} = \widetilde{O}\big(\lambda T \mathsf{dis}(\mathcal{D}, D_\star) + \log(|\mathcal{D}|)/\lambda\big). \quad \square$$

C.4 Proof of Lemma 3.4

Fixing $f \in \mathcal{F}$, $\nu \in \Delta([0, 1])$, and $p \in [0, 1]$, suppose that $\mathbb{E}_{q\sim\nu}[f(q)^2] \leq \varepsilon^2$ and $|f(p)| > \delta$. If $f(p) > \delta$, then $f(q) > \delta$ for all $q \leq p$ by monotonicity. Thus, $\mathbb{P}_{q\sim\nu}(q \leq p)\delta^2 \leq \mathbb{E}_{q\sim\nu}[f(q)^2] \leq \varepsilon^2$. Otherwise, if $f(p) < -\delta$, we analogously have $\mathbb{P}_{q\sim\nu}(q \geq p)\delta^2 \leq \mathbb{E}_{q\sim\nu}[f(q)^2] \leq \varepsilon^2$. Thus, for $\nu \in \Delta(\mathcal{X})$, we have

$$\begin{aligned}
&\mathbb{P}_{p\sim\nu}\big(\exists f \in \mathcal{F} : \mathop{\mathbb{E}}_{q\sim\nu}[f(q)^2] \leq \varepsilon^2 \wedge |f(p)| > \delta\big) \\
\leq &\mathbb{P}_{p\sim\nu}\left(\mathbb{P}_{q\sim\nu}(q \leq p) \leq \frac{\varepsilon^2}{\delta^2} \vee \mathbb{P}_{q\sim\nu}(q \geq p) \leq \frac{\varepsilon^2}{\delta^2}\right) \\
\leq &\mathbb{P}_{p\sim\nu}\left(\mathbb{P}_{q\sim\nu}(q \leq p) \leq \frac{\varepsilon^2}{\delta^2}\right) + \mathbb{P}_{p\sim\nu}\left(\mathbb{P}_{q\sim\nu}(q \geq p) \leq \frac{\varepsilon^2}{\delta^2}\right) \\
\leq &2\frac{\varepsilon^2}{\delta^2}\,.
\end{aligned}$$

Plugging this into the definition of dis finishes the proof. □

C.5 Proof of Lemma 3.5

For any distribution $\nu \in \Delta(\mathcal{Z})$, write $\nu_i = h_i \circ \nu|_{\mathcal{Z}_i}$ for law of $h_i(p)$ when $p \sim \nu$, conditioned on $p \in \mathcal{Z}_i$, and let $\mu_\nu(i) = \mathbb{P}_{p\sim\nu}(p \in \mathcal{Z}_i)$. We then bound

$$\begin{aligned}
\mathsf{dis}(\mathcal{G}) &= \sup_{\varepsilon,\delta>0} \sup_{\nu\in\mathcal{P}(\mathcal{Z})} \mathop{\mathbb{E}}_{i\sim\mu_\nu} \frac{\delta^2}{\varepsilon^2} \mathbb{P}_{p\sim\nu|_{\mathcal{Z}_i}}\left(\exists g \in \mathcal{G} : \mathop{\mathbb{E}}_{q\sim\nu}[g(q)^2] \leq \varepsilon^2 \wedge |g(p)| > \delta\right) \\
&\leq \sup_{\varepsilon,\delta>0} \sup_{\nu\in\mathcal{P}(\mathcal{Z})} \mathop{\mathbb{E}}_{i\sim\mu_\nu} \frac{\delta^2}{\varepsilon^2} \mathbb{P}_{p\sim\nu_i}\left(\exists f \in \mathcal{F} : \mu_\nu(i) \mathop{\mathbb{E}}_{q\sim\nu_i}[f(q)^2] \leq \varepsilon^2 \wedge |f(p)| > \delta\right) \\
&\leq \sup_{\varepsilon,\delta>0} \sup_{\nu\in\mathcal{P}(\mathcal{Z})} \frac{\delta^2}{\varepsilon^2} \mathop{\mathbb{E}}_{i\sim\mu_\nu}\left[\frac{\varepsilon^2}{\delta^2\mu_\nu(i)} \mathsf{dis}(\mathcal{F})\right] \\
&= N \mathsf{dis}(\mathcal{F}),
\end{aligned}$$

as desired. Here, the first inequality uses that $\mathbb{E}_{q\sim\nu}[g(q)^2] \geq \mu_\nu(i)\, \mathbb{E}_{q\sim\nu_i}[f(h_i(x)]$ for some $f \in \mathcal{F}$, and the second uses that $\nu_i \in \Delta(\mathcal{X})$ and the definition of dis. □

C.6 Proof of Lemma 3.9

POPS as generalized OPS. We observe that `POPS` is an instance of `GOPS` (Algorithm 3), with discretized price set $\mathcal{P} = \mathcal{P}_\varepsilon$ and smoothed demand function class $\mathcal{F} = \{\mathsf{dem}_D^\varepsilon : D \in \mathcal{D}\}$ (where each $f \in \mathcal{F}$ is viewed as a function on $\mathcal{P}_\varepsilon \times \mathcal{U}$ rather than $[0,1] \times \mathcal{U}$). Here, we view each $\hat{p}_t$ as the posted price instead of p_t. Indeed, taking $f^\star = \mathsf{dem}_\star^\varepsilon$, we have

$$\mathbb{E}[y_t | \hat{p}_t, u_t] = f^\star(\hat{p}_t, u_t)$$

and, for value distribution $Q \in \Delta([0,1])$ with smoothed demand function $g = \mathsf{dem}_Q^\varepsilon$, we have

$$\ell_\lambda^\varepsilon(Q, \hat{p}_t, u_t) = \ell_\lambda(g, \hat{p}_t, u_t),$$

where ℓ_λ on the right hand size is defined in (3). We do note, however, that the regret benchmark with smoothed demands and discretized prices differs slightly from the original benchmark.

Fixing the regret benchmark. Applying Lemma C.2 for this choice of $\mathcal{P}$ and $\mathcal{F}$, we have that

$$\mathbb{E}\left[\sum_{t=1}^T \mathsf{rev}_\star^\varepsilon(\mathsf{br}_\star^\varepsilon(u_t), u_t) - \mathsf{rev}_\star^\varepsilon(\hat{p}_t, u_t)\right] \le T\big(25\lambda(\mathsf{dis}(\mathcal{F}, f_\star) \vee 1) \log^2 T\big) - \frac{\log(\mu_1(D_\star))}{\lambda}.$$

Note that

$$\mathbb{E}\left[\sum_{t=1}^T \mathsf{rev}_\star^\varepsilon(\hat{p}_t, u_t)\right] = \mathbb{E}\left[\sum_{t=1}^T \mathsf{rev}_\star(p_t, u_t)\right],$$

so the regret bound above measures cumulative revenue in line with our original regret definition. Although the benchmark does not match that in the original definition, we have $|\mathsf{rev}_\star^\varepsilon(\mathsf{br}_\star^\varepsilon(u_t), u_t) - \mathsf{rev}_\star(\mathsf{br}_\star(u_t), u_t)| = O(\varepsilon)$ for all $t \in [T]$ due to one-sided Lipschitzness (Lemma 2.1). Consequently, the regret of POPS is bounded by

$$\mathbb{E}[R(T)] = \widetilde{O}\left(T\lambda(\mathsf{dis}(\mathcal{F}, f_\star) \vee 1) - \frac{\log(\mu_1(D_\star))}{\lambda} + \varepsilon T\right).$$

It remains to bound the disagreement coefficient by $O(K_\star)$, giving the lemma. Our argument below mirrors that in the proof of Theorem 3.2 but takes into account the smoothing and discretization.

Bounded disagreement coefficient. For each $u \in \mathcal{U}$, the function $\mathsf{dem}_\star^\varepsilon(\cdot, u)$ with domain $\mathcal{P}_\varepsilon$ is piecewise constant with $O(K_\star)$ sections. Indeed, the unsmoothed demand function $\mathsf{dem}_\star(\cdot, u)$ is piecewise constant with $O(K_\star)$ sections, and smoothing can only introduce new sections at the $O(K_\star)$ prices in $\mathcal{P}_\varepsilon$ that are within distance ε of a previous section boundary. Moreover, smoothing preserves monotonicity of demand functions. Hence, the function classes defining $\mathsf{dis}(\mathcal{F}, f_\star)$ are $O(K_\star)$-composites of the nonincreasing function class. Applying Lemmas 3.4 and 3.5 then implies that $\mathsf{dis}(\mathcal{F}, f_\star) = O(K_\star)$, giving the lemma. $\square$

C.7 Proof of Lemma 3.10

Fix any round t with context u_t and best-response price $\hat{p}_t$. Since $\|f_D^\varepsilon - f_{D_\star}^\varepsilon\|_\infty \le \varepsilon$, feedback y_t coincides with that which would have been obtained if $D_\star = D$ with probability at least $1 - \varepsilon$, conditioned on u_t and $\hat{p}_t$. Since the update to μ_t is only a function of u_t, $\hat{p}_t$, and y_t (notably, not the realized price p_t), we can iterate through all rounds and apply a union bound to obtain the lemma. $\square$

C.8 Proof of Lemma 3.11

For part one, fix $D, D' \in \Delta(\Theta)$ with $d_{\mathrm{L}}(D, D') \le \varepsilon^2/2$. Then, for all $u \in \mathcal{U}$ and $\hat{p} \in [0,1]$, we have

$$f_D^\varepsilon(\hat{p}, u) = \mathop{\mathbb{E}}_{\delta \sim \mathrm{Unif}([0,\varepsilon])}[\mathsf{dem}_D(\max\{\hat{p} - \delta, 0\}, u)]$$

Note that the maximum is unneeded since $\mathsf{proj}(D, u)$ is supported on $[0,1]$ and places no mass on negative values. Writing $Q = \mathsf{proj}(D, u)$ and $Q' = \mathsf{proj}(D', u)$, we then have

$$|f_D^\varepsilon(\hat{p}, u) - f_{D'}^\varepsilon(\hat{p}, u)| = \left|\mathop{\mathbb{E}}_{\delta \sim \mathrm{Unif}([0,\varepsilon])}[\mathsf{dem}_Q(\hat{p} - \delta) - \mathsf{dem}_{Q'}(\hat{p} - \delta)]\right|$$

$$
\begin{aligned}
&\le \frac{1}{\varepsilon}\left|\int_0^1 \mathsf{dem}_Q(\hat{p}-\delta) - \mathsf{dem}_{Q'}(\hat{p}-\delta)\,\mathrm{d}\delta\right| \\
&\le \frac{1}{\varepsilon}\left|\int_0^1 \mathsf{dem}_Q(t) - \mathsf{dem}_{Q'}(t)\,\mathrm{d}t\right| \\
&= \frac{1}{\varepsilon}\left|\int_0^1 \mathsf{dem}_Q(t) - \mathsf{dem}_{Q'}(t)\,\mathrm{d}t\right|.
\end{aligned}
$$

Writing $\tau = \varepsilon^2/2$, we further have

$$
\begin{aligned}
\int_0^1 \mathsf{dem}_Q(t) - \mathsf{dem}_{Q'}(t)\,\mathrm{d}t &= \int_\tau^{1+\tau} \mathsf{dem}_Q(t-\tau) - \int_0^1 \mathsf{dem}_{Q'}(t)\,\mathrm{d}t \\
&\le \int_\tau^1 \mathsf{dem}_Q(t-\tau) - \mathsf{dem}_{Q'}(t)\,\mathrm{d}t + \tau \\
&\le 2\tau,
\end{aligned}
$$

where the last step uses the fact that $d_{\mathrm{L}}(Q, Q') \le \tau$. A symmetric argument gives the reverse bound. Consequently, we have $|f^\varepsilon_D(\hat{p}, u) - f^\varepsilon_{D'}(\hat{p}, u)| \le \frac{1}{\varepsilon} \cdot 2\tau = \varepsilon$, as desired.

For part two, let $\mathcal{C} = \{C_1, C_2, \dots, C_n\}$ denote the intersection of the standard partition of $\mathbb{R}^d$ into cubes of side length $\varepsilon/\sqrt{d}$ with $\Theta \subseteq \mathbb{B}^d$. Denote the lexicographically smallest vertex of each C_i by c_i, and note that $\log(n) = O(d \log(d/\varepsilon))$. Given any $D \in \Delta_K(\Theta)$, we define the initial discretization $\hat{D}_0 = \sum_{i=1}^n D(C_i)\delta_{c_i}$. We obtain the final discretized measure $\hat{D}$ by rounding each weight to a neighboring multiple of ε/K (choice doesn't matter so long as we maintain unit mass, this is always possible). Then, for any context $u \in \mathbb{S}^{d-1}$ and price $p \in [0, 1]$, we have

$$
\begin{aligned}
\mathbb{P}_{\theta\sim D}(\langle u, \theta\rangle \le p) - \mathbb{P}_{\theta\sim \hat{D}}(\langle u, \theta\rangle \le p + \varepsilon) \le \varepsilon \\
\mathbb{P}_{\theta\sim \hat{D}}(\langle u, \theta\rangle \le p - \varepsilon) - \mathbb{P}_{\theta\sim D}(\langle u, \theta\rangle \le p) \le \varepsilon,
\end{aligned}
$$

using that each cube in $\mathcal{C}$ has diameter ε and that the mass in each cube was perturbed by at most ε/K (so with K cubes the probability of any event is shifted by at most ε). Thus, $d_{\mathrm{L}}(D, \hat{D}) \le \varepsilon$. Using balls and bins, we can thus bound

$$
N(\Delta_K(\Theta), d_{\mathrm{L}}, \varepsilon) \le n^K \binom{K + K/\varepsilon}{K} \le n^K (2K/\varepsilon)^K = \exp(O(Kd \log(d/\varepsilon) + K \log(K/\varepsilon)),
$$

as desired. $\square$

C.9 Proof of Theorem 3.8

Fix $\varepsilon = T^{-2}$ and $\lambda = \sqrt{d/T}$. We now construct $\mathcal{D}$ and μ_1. Write $M = \lceil \log T \rceil$, and, for $i = 1, \dots, M$, take $\mathcal{D}_i$ to be a minimal $(\varepsilon^2/2)$-cover of $\Delta_{2^i}(\Theta)$ under the Lévy metric. By Lemma 3.11, we have $\log|\mathcal{D}_i| = \widetilde{O}(2^i d \log 1/\varepsilon)$. Now set $\mathcal{D} = \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_M$ and take $\mu_1(D) \propto (2^i |\mathcal{D}_i|)^{-1}$ for $D \in \mathcal{D}_i$. This ensures that $\log(1/\mu_1(D)) = \widetilde{O}(2^i d \log T)$ for $D \in \mathcal{D}_i$.

Assume without loss of generality that $M \ge \log K_\star$; otherwise, the regret bound is vacuous. Then, there exists $\hat{D} \in \mathcal{D}_{\lceil \log K_\star \rceil} \subseteq \mathcal{D}$ such that $d_{\mathrm{L}}(\hat{D}, D_\star) \le \varepsilon^2/2$ and $\|\mathsf{dem}^\varepsilon_{\hat{D}} - \mathsf{dem}^\varepsilon_{D_\star}\|_\infty \le \varepsilon$, again using Lemma 3.11. Thus, by Lemma 3.10, the realized trajectory of `POPS` $\{u_t, \hat{p}_t, p_t, y_t\}_{t=1}^T$ can be coupled with an alternative trajectory $\{u'_t, \hat{p}'_t, p'_t, y'_t\}_{t=1}^T$ of `POPS` with type distribution $\hat{D}$, such that $\{u_t, \hat{p}_t, y_t\}_{t=1}^T = \{u'_t, \hat{p}'_t, y'_t\}_{t=1}^T$ with probability at least $1 - \varepsilon T$. By Lemma 3.9, we have

$$
\mathbb{E}\left[\sum_{t\in[T]} \mathsf{rev}_{\hat{D}}(\mathsf{br}_{\hat{D}}(u'_t), u'_t) - \mathsf{rev}_{\hat{D}}(p'_t, u'_t)\right] = \widetilde{O}\big(K\sqrt{dT} - \sqrt{T/d} \cdot \log \mu_1(\hat{D})\big) = \widetilde{O}(K_\star \sqrt{dT}).
$$

At this point, we can use the coupling guarantee and the bound $d_{\mathrm{L}}(\hat{D}, D_\star) \le \varepsilon^2$ to show that the left hand side above and the true expected regret differ by $O(\varepsilon^2 T) = O(1)$, giving the theorem.

The sketch in Section 3.2 was precise until the last paragraph. To formalize the remainder, we note that, by the coupling guarantee,

$$\begin{aligned}
\mathbb{E}[R(T)] &\leq \mathbb{E}\left[\sum_{t=1}^{T} \mathsf{rev}_\star(\mathsf{br}_\star(u_t), u_t) - \mathsf{rev}_\star(p_t, u_t) \,\middle|\, \mathcal{E}\right] + \varepsilon T^2 \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \mathsf{rev}_\star(\mathsf{br}_\star(u'_t), u'_t) - \mathsf{rev}_\star(p_t, u'_t) \,\middle|\, \mathcal{E}\right] + 1
\end{aligned}$$

Since $d_{\mathrm{L}}(\hat{D}, D_\star) \leq \varepsilon^2$, Lemma 2.2 implies that $\mathsf{rev}_\star(\mathsf{br}_\star(u'_t), u'_t) \leq \mathsf{rev}_{\hat{D}}(\mathsf{br}_{\hat{D}}(u'_t), u'_t) + O(\varepsilon^2)$ for all rounds t. We further have

$$\begin{aligned}
\mathbb{E}[\mathsf{rev}_\star(p_t, u'_t) \mid u'_t, \hat{p}_t] &= \mathsf{rev}^{\varepsilon}_{D_\star}(\hat{p}_t, u'_t) \\
&\leq \mathsf{rev}^{\varepsilon}_{\hat{D}}(\hat{p}_t, u'_t) + \varepsilon \\
&= \mathbb{E}[\mathsf{rev}_{\hat{D}}(p'_t, u'_t) \mid u'_t, \hat{p}'_t = \hat{p}_t] + \varepsilon \\
&\leq \mathbb{E}[\mathsf{rev}_{\hat{D}}(p'_t, u'_t) \mid u'_t, \hat{p}'_t, \mathcal{E}] + \varepsilon + \varepsilon T.
\end{aligned}$$

Combining the above, we obtain

$$\begin{aligned}
\mathbb{E}[R(T)] &\leq \mathbb{E}\left[\sum_{t=1}^{T} \mathsf{rev}_\star(\mathsf{br}_\star(u'_t), u'_t) - \mathsf{rev}_\star(p_t, u'_t) \,\middle|\, \mathcal{E}\right] + \varepsilon T^2 \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \mathsf{rev}_{\hat{D}}(\mathsf{br}_{\hat{D}}(u'_t), u'_t) - \mathsf{rev}_\star(p_t, u'_t) \,\middle|\, \mathcal{E}\right] + O(\varepsilon T^2) \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \mathsf{rev}_{\hat{D}}(\mathsf{br}_{\hat{D}}(u'_t), u'_t) \,\middle|\, \mathcal{E}\right] - \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}[\mathsf{rev}_\star(p_t, u'_t) \mid u'_t, \hat{p}_t] \,\middle|\, \mathcal{E}\right] + O(1) \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \mathsf{rev}_{\hat{D}}(\mathsf{br}_{\hat{D}}(u'_t), u'_t) \,\middle|\, \mathcal{E}\right] - \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\big[\mathsf{rev}_{\hat{D}}(p'_t, u'_t) \mid u'_t, \hat{p}'_t, \mathcal{E}\big] \,\middle|\, \mathcal{E}\right] + O(1) \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \mathsf{rev}_{\hat{D}}(\mathsf{br}_{\hat{D}}(u'_t), u'_t) - \sum_{t=1}^{T} \mathsf{rev}_{\hat{D}}(p'_t, u'_t) \,\middle|\, \mathcal{E}\right] + O(1) \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \mathsf{rev}_{\hat{D}}(\mathsf{br}_{\hat{D}}(u'_t), u'_t) - \sum_{t=1}^{T} \mathsf{rev}_{\hat{D}}(p'_t, u'_t)\right] + O(1) \\
&= \widetilde{O}(K_\star \sqrt{dT}),
\end{aligned}$$

as desired. □

Finally, we seek to understand the fundamental barriers to effective pricing. The following lower bound shows that our $\widetilde{O}(K_\star\sqrt{dT})$ guarantee for `POPS` has optimal dependence on d and T.

Theorem C.7. *Fix $K_\star \geq 2$ and $T \geq d \cdot K_\star^3$. For any d-dimensional contextual pricing algorithm $\mathcal{A}$, there exists a problem instance with $K_\star$ types and stochastic contexts such that $\mathbb{E}[R_{\mathcal{A}}(T)] = \Omega(\sqrt{K_\star dT})$.*

This result holds even when the seller knows $K_\star$. Previously, [4] proved a lower bound of $\Omega(\sqrt{K_\star T})$ for the non-contextual case. Our proof modifies their one-dimensional construction so that it can be cleanly tensored into d dimensions. We defer the proofs of some minor claims to Appendix C.10.

Proof of Theorem C.7 Starting in 1D, we define valuations $\frac{1}{2} \leq v_1, \ldots, v_{K_\star} \leq 1$ by $v_i := \frac{1}{2} + \frac{i-1}{4K_\star - 2i - 2}$. Define the base distribution Q_0 on $\{v_1, \ldots, v_{K_\star}\}$ as follows: $\forall i \in [K_\star - 1], Q_0(v_i) := 1/(2K_\star - 2)$, and for $i = K_\star, Q_0(v_i) := 1/2$. Observe that each valuation v_i has the same expected revenue of $1/2$. That is, $\mathsf{rev}_{Q_0}(v_i) = v_i \mathbb{P}_{V \sim Q_0}(V \geq v_i) = 1/2$ for all $i \in [K_\star]$. Now, for $j \in \{2, \ldots, K_\star - 1\}$, we define distribution Q_j by slightly lowering the probability of v_{j-1} and increasing the probability of v_j by the same amount. That is, $Q_j(v_i) := Q_0(v_i) - \varepsilon$ for $i = j - 1$

and $Q_j(v_i) := Q_0(v_i) - \varepsilon$ for $i = j$, where $\varepsilon > 0$ is a small constant to be determined. In contrast to the family of instances specified in [4], our Q_j distributions share the same multiset of probability weights, differing only in the locations of the perturbed valuations. This difference allows us to tensor these problem instances into a d-dimensional instance without leaking information between instances.

Tensoring one-dimensional instances. To extend these one-dimensional distributions into d-dimensional space, define the base distribution $D_0 \in \Delta([0,1]^d)$ as follows. For each $i \in [K_\star]$, let $\theta_i = [v_i, \dots, v_i] \in \mathbb{R}^d$, $w_i = Q_0(v_i)$, and take $D_0 = \sum_{i=1}^{K_\star} w_i \delta_{\theta_i}$. Thus, θ_i is a d-dimensional vector with all entries equal to v_i, and w_i is the probability D_0 places on θ_i, taken to match Q_0 at v_i. Now, for a selection $j = (j_1, \dots, j_d) \in \{2, \dots K_\star - 1\}^d$, define the perturbed instance D_j by starting from D_0 and modifying it as follows:

- Adjust the probabilities w_1 and w_2 by $w_1 \leftarrow w_1 - \varepsilon$ and $w_2 \leftarrow w_2 + \varepsilon$.
- For each dimension $\ell \in [d]$, adjust the θ_i vectors so that the marginal distribution of their ℓth coordinates coincides with Q_{j_ℓ}. Specifically, swap $\theta_{j_\ell - 1}[\ell]$ with $\theta_1[\ell]$ and $\theta_{j_\ell}[\ell]$ with $\theta_2[\ell]$.

This construction ensures that for each dimension i, the marginal distribution of D_j is Q_{j_i} (see Lemma C.8).

Lower bounding the regret. First, we show that no one-dimensional pricing algorithm can achieve regret $o(\sqrt{KT})$ for $T \geq K_\star^3$ when $D_\star$ is sampled uniformly at random from the Q_j distributions in a similar fashion to [4]. For completeness, the full argument (including the tuning of ε) appears in Lemma C.9. For the contextual setting, we sample a selection $j = (j_1, \dots, j_d) \in [K]^d$ uniformly at random and set $D_\star$ to the perturbed instance D_j. Then we split the time horizon T into d contiguous epochs of length T/d. During epoch l, we choose the contexts to be the standard basis vector for dimension l. This means the algorithm interacts with a "new" one-dimensional problem instance during that epoch and thus incurs regret $\Omega(\sqrt{KT/d})$ when $T \geq d \cdot K_\star^3$. Summing this over the d epochs gives the desired result. To achieve this lower bound with stochastic contexts, we forgo epochs and instead sample each u_t uniformly at random from the d standard basis vectors; a similar argument still applies (see Lemma C.10 for details). $\square$

C.10 Missing Lemmas from Proof of Theorem C.7

Lemma C.8. *For each coordinate $\ell \in [d]$, in the instance D_j, the marginal distribution of the ℓth coordinate is precisely Q_{j_ℓ}. That is, for every valuation v_k,*

$$\mathbb{P}_{\theta \sim D_j}\big(\theta[\ell] = v_k\big) \;=\; Q_{j_\ell}(v_k).$$

Proof. Recall that in the base instance $D_0 = \sum_{i=1}^{K_\star} w_i \, \delta_{\theta_i}$, each type i occurs with probability $w_i = Q_0(v_i)$ and, in every coordinate, its value is v_i. Thus, the marginal distribution on any coordinate is Q_0. For a fixed coordinate ℓ, the construction perturbs D_0 as follows:

1. In the ℓth coordinate, we swap the entries of the following types:
   - Replace the ℓth coordinate of type 1 by $v_{j_\ell - 1}$ and that of type $j_\ell - 1$ by v_1.
   - Replace the ℓth coordinate of type 2 by v_{j_ℓ} and that of type j_ℓ by v_2.
2. The probability weight of type 1 is decreased by ε and that of type 2 is increased by ε. All other weights remain unchanged.

After these modifications, the marginal on coordinate ℓ is obtained by "reassigning" the masses as follows:

- The mass originally at v_1 (from type 1) is now associated with $v_{j_\ell - 1}$ (since type 1 now has coordinate $v_{j_\ell - 1}$); however, due to the probability adjustment, the effective mass at $v_{j_\ell - 1}$ becomes
$$Q_0(v_{j_\ell - 1}) - \varepsilon.$$

- Similarly, the mass originally at v_2 (from type 2) is now associated with v_{j_ℓ}, yielding an effective mass
$$Q_0(v_{j_\ell}) + \varepsilon.$$
- For all other values v_i (with $i \notin \{1, 2, j_\ell - 1, j_\ell\}$), the coordinate remains unchanged and the mass is $Q_0(v_i)$.

This exactly corresponds to the definition of the perturbed one-dimensional distribution Q_{j_ℓ}, which is given by:
$$Q_{j_\ell}(v_{j_\ell - 1}) = Q_0(v_{j_\ell - 1}) - \varepsilon, \quad Q_{j_\ell}(v_{j_\ell}) = Q_0(v_{j_\ell}) + \varepsilon,$$
with $Q_{j_\ell}(v_i) = Q_0(v_i)$ for all $i \notin \{j_\ell - 1, j_\ell\}$. Thus, for any coordinate ℓ, the marginal distribution of the ℓth coordinate of D_j is exactly Q_{j_ℓ}, as required. ☐

Lemma C.9 (Regret of one dimensional family)**.** *Fix any number of types $K_\star \geq 3$ and $T \geq K_\star^3$, the distribution over problem instances described in Appendix C.9 ensures that the expected regret of any pricing strategy is $\Omega(\sqrt{K_\star T})$.*

Proof. This proof follows the same steps in the lower bound proof of [4], albeit with a different set of problem instances. First we recall the construction: we define valuations $\frac{1}{2} \leq v_1, \ldots, v_{K_\star} \leq 1$ by
$$v_i = \frac{1}{2} + \frac{i - 1}{4K_\star - 2i - 2}.$$
Define the base distribution Q_0 on $\{v_1, \ldots, v_{K_\star}\}$ as follows:
$$Q_0(v_i) = \begin{cases} \frac{1}{2K_\star - 2} & \forall i \in [K_\star - 1] \\ \frac{1}{2} & i = K_\star \end{cases}.$$
Observe that each valuation v_i has the same expected revenue of $1/2$. That is, $\mathsf{rev}_{Q_0}(v_i) = v_i \mathbb{P}_{V \sim Q_0}(V \geq v_i) = 1/2$ for all $i \in [K_\star]$. Now, for $j \in \{2, \ldots, K_\star - 1\}$, we define distribution Q_j by slightly lowering the probability of v_{j-1} and increasing the probability of v_j by the same amount:
$$Q_j(v_i) = \begin{cases} Q_0(v_i) & i \in [K_\star] \backslash \{j - 1, j\} \\ Q_0(v_i) - \varepsilon & i = j - 1 \\ Q_0(v_i) + \varepsilon & i = j \end{cases}$$
Here $\varepsilon > 0$ will be determined later. Observe that if the buyer valuations are drawn from Q_j, the valuation v_j achieves expected revenue at least $1/2 + \varepsilon$, whereas all other $v \neq v_j$ achieve about $1/2$.

Let J be uniformly distributed over $[K_\star]$, and let the valuations $V_1, \ldots, V_T$ be i.i.d. from Q_J. This means that we first select $J \in [K_\star]$ uniformly at random (the "good" valuation index), and then, given $J = j$, each V_t is drawn independently from Q_j. A pricing strategy $\mathcal{A}$ is a sequence of prices $(X_1, \ldots, X_T)$ that may depend on past observations. The regret of $\mathcal{A}$ over T rounds is
$$R_{\mathcal{A}}(T) \;=\; \max_{k \in [K_\star]} \sum_{t=1}^{T} r_t(v_k) \;-\; \sum_{t=1}^{T} r_t(X_t),$$
where $r_t(\cdot)$ is the realized revenue in round t.

We first argue that whenever the algorithm $\mathcal{A}$ does *not* pick the "good" valuation v_J, it loses at least ε in expected revenue at that round. Concretely, define
$$N_j \;=\; \sum_{t=1}^{T} \mathbb{1}\{X_t = v_j\}$$
as the number of times the algorithm posts price v_j. Because picking v_J yields zero regret on those rounds and any other price yields at least ε of regret, we obtain
$$\mathbb{E}_{J,V}\Big[\sum_{t=1}^{T} r_t(v_k) - \sum_{t=1}^{T} r_t(X_t)\Big] \;\geq\; \varepsilon\Big(T - \mathbb{E}_{J,V}[\,N_J\,]\Big). \tag{5}$$

(The expectation is over both the random choice of J and the random draws of valuations $V_1, \ldots, V_T$.)

Next, we control $\mathbb{E}_{J,V}[\, N_j \,]$ by coupling the distributions Q_j and Q_0 and applying Pinsker's inequality. Let $Y_t = \mathbf{1}\{V_t \geq X_t\}$, and write $q_j(\cdot)$ and $q_0(\cdot)$ for the induced distributions on the entire sequence $Y^T = (Y_1, \ldots, Y_T)$ under Q_j and Q_0, respectively. For any deterministic function $f : \{0,1\}^T \to [0, M]$,

$$\mathbb{E}_{V\sim Q_j}[f(Y)] - \mathbb{E}_{V\sim Q_0}[f(Y)] \;\leq\; M\,\sqrt{\tfrac{1}{2}\,\mathrm{KL}\big(q_0 \,\|\, q_j\big)}.$$

By a chain rule argument, whenever $X_t \neq v_j$, the local conditional distributions for Y_t coincide under Q_0 and Q_j, thus incurring zero KL divergence. The only deviation comes from the rounds t where $X_t = v_j$. For all $v_j \geq 3/4$ and $\varepsilon < 1/6$, using the fact that $\mathrm{KL}(x||x+\alpha) \leq \alpha^2(x+\alpha)^{-1}(1-x-\alpha)^{-1}$

$$\mathrm{KL}\left(\frac{1}{2v_j} \,\|\, \frac{1}{2v_j} + \varepsilon\right) \leq 18\varepsilon^2$$

Therefore,

$$\mathrm{KL}\big(q_0 \,\|\, q_j\big) \;\leq\; 18\,\varepsilon^2 \sum_{t=1}^{T} Q_0(X_t = v_j) \;=\; 18\,\varepsilon^2\, T\, Q_0(v_j) \;=\; 18\,\varepsilon^2\, \mathbb{E}_{V\sim Q_0}[N_j]$$

This allows us to bound the difference in expectations of any deterministic statistic of Y. In particular, since N_j itself is a deterministic function of Y under a fixed strategy,

$$\mathbb{E}_{V\sim Q_j}[N_j] \;\leq\; \mathbb{E}_{V\sim Q_0}[N_j] + \varepsilon\, T\,\sqrt{18\,\mathbb{E}_{V\sim Q_0}[N_j]}.$$

Taking expectation also over J, one gets

$$\mathbb{E}_{J,V}[\, N_J\,] \;\leq\; \mathbb{E}_J\Big[\mathbb{E}_{V\sim Q_0}[\, N_J\,]\Big] \;+\; \varepsilon\, T\,\sqrt{18\,\mathbb{E}_J\Big[\mathbb{E}_{V\sim Q_0}[\, N_J\,]\Big]}.$$

Since $\sum_{j=1}^{K} N_j = T$ and J is uniform over $[K]$, it follows that $\sum_{j=1}^{K} \mathbb{E}_{V\sim Q_0}[N_j] = T$, while on average $\mathbb{E}_J[\,\mathbb{E}_{Q_0}[N_J]]$ cannot exceed some fraction of T. Putting these together in (5)

$$\mathbb{E}_{J,V}\Big[\sum_{t=1}^{T} r_t(v_k) - \sum_{t=1}^{T} r_t(X_t)\Big] \;\geq\; \varepsilon\,\Big(T - \mathbb{E}_J\Big[\mathbb{E}_{V\sim Q_0}[\, N_J\,]\Big] \;-\; \varepsilon\, T\,\sqrt{18\,\mathbb{E}_J\Big[\mathbb{E}_{V\sim Q_0}[\, N_J\,]\Big]}\Big). \tag{6}$$

Substituting into the above expression and choosing ε to be $\frac{1}{6\sqrt{18}}\sqrt{K_\star/T}$ as long as $T \geq K_\star^3$ yields a final lower bound on the regret of the form

$$\mathbb{E}[R_{\mathcal{A}}(T)] \geq \varepsilon T\left(\frac{1}{3} - \varepsilon\sqrt{18\frac{T}{K_\star}}\right) \;\geq\; \Omega\Big(\sqrt{K_\star T}\Big),$$

as desired. □

Lemma C.10 (Stochastic Contexts)**.** *Fix any number of types $K_\star \geq 2$ and time horizon $T \geq d \cdot K_\star^3$. For any d-dimensional contextual pricing algorithm $\mathcal{A}$, there exists a problem instance with $K_\star$ types and stochastic contexts such that $\mathbb{E}[R_{\mathcal{A}}(T)] = \Omega(\sqrt{K_\star d T})$.*

Proof of Lemma C.10 We sample a selection $j = (j_1, \ldots, j_d) \in [K]^d$ uniformly at random and set $D_\star$ to the perturbed instance D_j. For each time step $t \in [T]$, we sample a coordinate $\ell \in [d]$ and set u_t to be the basis vector in the direction of ℓ. After T rounds, with probability at least $1 - o(1)$, each coordinate ℓ is sampled at least $T/d - \sqrt{T/d \log T}$ times by a Chernoff bound. The regret of the one dimension instance corresponding to ℓ during these rounds is at least $\Omega(\sqrt{K_\star T/d})$. Summing across dimensions gives the desired result. □

D Proofs for Section 4

Since our analysis exclusively reasons about the true type distribution $D_\star$, we abbreviate $\mathsf{dem} = \mathsf{dem}_\star$, $\mathsf{rev} = \mathsf{rev}_\star$, $\mathsf{gap} = \mathsf{gap}_\star$, and $\mathsf{br} = \mathsf{br}_\star$. For each price $p \in [0,1]$ and round $t \in [T]$, we define

ALGORITHM 4: `ZoomV`: Variance-Aware Zooming for Non-Contextual Pricing

Initialize: active price set $S \leftarrow \{\frac{2^i}{T} : i = 0, 1, \ldots, \lfloor \log_2 T \rfloor\} \cup \{1\}$;
for *each round* $t \in [T]$ **do**
    play $p_t \in \arg\max_{q \in S} \mathsf{index}_t(q)$;
    observe $y_t = \mathbb{1}\{\theta_t \le p_t\}$ and update $n_{t+1}(p_t)$;
    if *a price* $p > 1/T$ *becomes uncovered* **then**
        $q \leftarrow \min\{q' \in S : q' > p_t\}$;
        $S \leftarrow S \cup \{(p_t + q)/2\}$

$\mathcal{T}_t(p) := \{\tau \in [t-1] : p_\tau = p\}$ as the set of previous rounds where p was played, and let $n_t(p) := |\mathcal{T}_t(p)|$ denote the count of these rounds. We define $\mu_t(p) := \frac{1}{n_t(p)} \sum_{\tau \in \mathcal{T}_t(p)} p y_t$ as the average revenue during these rounds, $V_t(p) := \frac{1}{n_t(p)-1} \sum_{\tau \in \mathcal{T}_t(p)} (p y_t - \mu_t(p))^2$ as the sample variance, and $\sigma^2(p) := p^2 \mathsf{dem}(p)(1 - \mathsf{dem}(p))$ as the population variance, which remains unknown to the seller. When $n_t(p) = 0$, we set $\mu_t(p) = 0 = V_t(p) = 0$, and when $n_t(p) = 1$, we take $V_t(p) = \infty$. The confidence radius is given by $r_t(p) := \sqrt{\frac{10 V_t(p) \log T}{n_t(p)}} + \frac{12 \log(T)}{n_t(p)-1}$, which is taken as $+\infty$ if $n_t(p) \le 1$. A variant of Bernstein's inequality ensures that $|\mu_t(p) - \mathsf{rev}(p)| \le r_t(p)$ with high probability (see Lemma D.1). Defining the upper confidence bound as $\mathsf{UCB}_t(p) := \mu_t(p) + r_t(p)$, it follows that $\mathsf{rev}(p) \le \mathsf{UCB}_t(p)$ with high probability. A price p is said to be *covered* by some $q \in S$ if $p \in [q, q + r_t(q)]$ and q is the largest active price no greater than p, i.e., $q = \max\{q' \in S : q' \le p\}$. The one-sided Lipschitz property of the revenue function (Lemma 2.1) implies that $\mathsf{rev}(p) - \mathsf{rev}(q) \le r_t(q)$ with high probability. Finally, we define the index of a price $q \in S$ as $\mathsf{index}_t(q) := \mathsf{UCB}_t(q) + r_t(q)$, ensuring that any price p covered by some $q \in S$ satisfies $\mathsf{rev}(p) \le \mathsf{index}_t(q)$.

D.1 Proof of Lemma 4.2

We begin with a few helper lemmas. Throughout, we assume that $T \ge 3$ (otherwise the regret bound holds trivially). Our proofs mirror those of similar lemmas in [27], with small adjustments to handle the variance-adjusted confidence radii and the dyadic price selection rule.

Lemma D.1 (Concentration)**.** *Write $\mathcal{E}_{\text{clean}}$ for the event that*

$$|\mu_t(p) - \mathsf{rev}(p)| \le r_t(p) \le \sqrt{\frac{10\,\sigma^2(p) \log T}{n_t(p)}} + \frac{126 \log(T)}{n_t(p) - 1}$$

for all $t \in [T]$ and for all $p \in [0,1]$. Then $\Pr(\mathcal{E}_{\text{clean}}) \ge 1 - 8T^{-2}$.

Proof. For fixed $p \in [0,1]$ and $t \in [T]$, Theorems 10 and 11 of Maurer and Pontil [20] imply that

$$|\mu_t(p) - \mathsf{rev}(p)| \le r_t(p)$$
$$\left|\sigma_t(p) - \sqrt{V_t(p)}\right| \le \sqrt{\frac{11 \log T}{n_t(p) - 1}}$$

with probability $1 - 8T^{-5}$. We note that similar bounds appear in [2], which inspired our adjustments to the confidence intervals. Under this event, we further bound

$$\begin{aligned} r_t(p) &= \sqrt{\frac{10 V_n(p) \log T}{n_t(p)}} + \frac{12 \log T}{n_t(p) - 1} \\ &\le \sqrt{\frac{10\sigma^2(p) \log T}{n_t(p)}} + \sqrt{\frac{10 \log T}{n_t(p)}} \sqrt{\frac{11 \log T}{n_t(p) - 1}} + \frac{12 \log T}{n_t(p) - 1} \\ &\le \sqrt{\frac{10\sigma^2(p) \log T}{n_t(p)}} + \frac{126 \log T}{n_t(p) - 1}. \end{aligned}$$

Taking a union bound over t, the above must hold for all $t \in [T]$ with probability at least $1 - 8T^{-4}$. Now, the same Chernoff bound argument used in Claim 4.13 of Slivkins et al. [27] implies that $|\mu_t(p) - \mathsf{rev}(p)| \leq r_t(p)$ for all $p \in [0,1]$ and $t \in [T]$ with probability at least $1 - 8T^{-2}$. One technical observation is that Claim 4.13 requires that the set of arms every played by the algorithm is finite. This holds for `ZoomV` due to our dyadic arm activation rule. $\square$

Lemma D.2 (Covering invariant)**.** *At the beginning of each round, every price $p \geq 1/T$ is covered by some active arm.*

Proof. At round 1, $r_t(q) = \infty$ for all active arms $q \in S$, and so all arms larger than $1/T$ are covered by our choice of S. Now, suppose that the lemma holds up to round t, and that playing p_t causes a price $p \in R$ to become uncovered. Then, by the definition of covering, we must have

$$p_t \leq p_t + r_{t+1}(p_t) < p < q \leq p_t + r_t(p_t),$$

where q is the nearest active price to the right of p_t, selected at Step 6. First, verify that the added price, $p' := (p_t + q)/2$, is less than p. Since $r_{t+1}(p_t)$ must be less than one, we must have $n_{t+1}(p) - 1 = n_t(p) \geq 12$. Thus, we can bound

$$\begin{aligned} V_{t+1}(p) &= \frac{p^2}{n_{t+1}(p)(n_{t+1}(p)-1)} \sum_{\tau_1 \in \mathcal{T}_{t+1}(p)} \sum_{\tau_2 \in \mathcal{T}_{t+1}(p)} (y_{\tau_1} - y_{\tau_2})^2 \\ &\geq \frac{p^2}{(n_t(p)+1)n_t(p)} \sum_{\tau_1 \in \mathcal{T}_t(p)} \sum_{\tau_2 \in \mathcal{T}_t(p)} (y_{\tau_1} - y_{\tau_2})^2 \\ &= \frac{n_t(p)-1}{n_t(p)+1} V_t(p) \\ &\geq \frac{11}{13} V_t(p). \end{aligned}$$

Consequently, one can show that

$$r_{t+1}(p_t) = \sqrt{\frac{10 V_{t+1}(p)\log T}{n_t(p)+1}} + \frac{12\log(T)}{n_t(p)} \geq \frac{3}{4}\sqrt{\frac{10 V_t(p)\log T}{n_t(p)}} + \frac{12\log(T)}{n_t(p)-1} = \frac{3}{4} r_t(p_t).$$

Combining, we find that $p' = (p_t + q)/2 \leq p_t + \frac{1}{2} r_t(p_t) < p_t + r_{t+1}(p_t) < p$, as desired. Moreover, p' could not have already been active at round t; otherwise, p_t would not have been covering p. Finally, once p' is added, it covers p since $r_{t+1}(p') = \infty$. $\square$

Lemma D.3 (Gap bound)**.** *Condition on $\mathcal{E}_{\text{clean}}$. Then $\mathsf{gap}(p) \leq 5 r_t(p)$ and $n_t(p) \leq \frac{252\sigma^2(p)\log T}{\mathsf{gap}(p)^2} + \frac{504\log T}{\mathsf{gap}(p)}$ for all $p \in [0,1]$ and $t \in [T]$.*

Proof. Write $\hat{p} = \max\{\mathsf{br}, \frac{1}{T}\}$, and fix any price $p \in [0,1]$. Consider some round t at which p is played. By Lemma D.2, we know that $\hat{p}$ was covered at the beginning of round t by some price $q \in S$, and that p had a higher index than q. Hence,

$$\begin{aligned} \mathsf{index}_t(p) &\geq \mathsf{index}_t(q) && \text{(since } p \text{ was played)} \\ &= \mathsf{UCB}_t(q) + r_t(q) && \text{(by definition of index)} \\ &\geq \mu(q) + r_t(q) && \text{(concentration guarantee)} \\ &\geq \mathsf{rev}(\hat{p}). && (\hat{p} \text{ covered by } q) \end{aligned}$$

Moreover, $\mathsf{index}_t(p) \leq \mu_t(p) + 2r_t(p) \leq \mathsf{rev}(p) + 3r_t(p)$. Thus,

$$\mathsf{gap}(p) = \mathsf{rev}(p_\star) - \mathsf{rev}(p) \leq \frac{1}{T} + \mathsf{rev}(\hat{p}) - \mathsf{rev}(p) \leq \frac{1}{T} + 3r_t(p).$$

Now, if $n_t(p) \leq 12$, $\mathsf{gap}(p) \leq 1 < r_{t+1}(p)$. Otherwise, the bound above implies that $\mathsf{gap}(p) \leq 5r_{t+1}(p)$. Since $r_t(p)$ only changes when p is played and $\mathsf{gap}(p) \leq 5r_t(p)$ when $t = 1$, this guarantee holds for all t. For the other bound, we apply concentration to obtain

$$\mathsf{gap}(p) \leq 5\left(\sqrt{\frac{10\,\sigma^2(p)\log T}{n_t(p)}} + \frac{126\log(T)}{n_t(p)-1}\right).$$

Rearranging and solving the quadratic inequality in $n_t(p)$ gives the stated result. $\square$

Compared to the standard gap bound lemma for zooming (see, e.g., Lemma 4.14 of 27), Lemma D.3 is adapted to the variance of each price p. We can now prove the regret bound.

Lemma D.4 (Active arm separation)**.** *Conditioned on $\mathcal{E}_{\text{clean}}$, consider any three consecutive active arms $x < y < z$ which did not belong to S at initialization. Then $z - x > \frac{1}{10}\min\{\mathsf{gap}(x), \mathsf{gap}(y)\}$.*

Proof. If y was activated before z, then z must have been added as the midpoint of active arms y and $y + 2(z-y) = 2z - y$ at round τ_z, when y must have not covered $2z - y$. Thus, by Lemma D.3, we would have $2(z-x) > 2(z-y) = (2z-y) - y > \frac{1}{5}\mathsf{gap}(y)$. On the other hand, if y was activated after z, then it must have been added as the midpoint of x and z at round τ_y, when x must have not covered z. Again, by Lemma D.3, this would imply $z - x > \frac{1}{5}\Delta(y)$. □

Proof of Lemma 4.2 We freely condition on $\mathcal{E}_{\text{clean}}$, since the complement has negligible probability $O(T^{-2})$. For each $\delta > 0$, let $Y_\delta \subseteq X_{2\delta}$ denote the set of activated prices p with $\mathsf{gap}(p) \in [\delta, 2\delta)$. In what follows, we say that two prices are adjacent if they are neighboring within Y_δ. Note that at most $O(\log T)$ of the prices in Y_δ were activated at initialization. Consider the set Y_δ^0 which, for each such price, contains this price and up to two neighboring prices, such that the remaining prices $Y_\delta \setminus Y_\delta^0$ can be split into triples of neighboring prices. We then decompose

$$Y_\delta \setminus Y_\delta^0 = \{p_{1,1} < p_{1,2} < p_{1,3} < p_{2,1} < p_{2,2} < p_{2,3} < \cdots < p_{n,1} < p_{n,2} < p_{n,3}\},$$

where $p_{i,1}, p_{i,2}, p_{i,3}$ are neighboring for each $i \in [n]$. By Lemma D.4, we have $p_{i,3} - p_{i,1} > \delta/10$ for all $i \in [n]$, and so $p_{i,k} - p_{j,k} > \delta/10$ for all $k \in [3]$ whenever $i < j - 1$. Thus, we can partition $Y_\delta \setminus Y_\delta^0$ into at most 6 packings, each of which has separation at least $\delta/10$. Of course any $(\delta/10)$-packing of $Y_\delta \subseteq X_{2\delta}$ is contained within a $(\delta/5)$-cover of $X_{2\delta}$. Consequently, we have

$$\sum_{p \in Y_\delta} \sigma^2(p) \le 6N_{\text{var}}(X_{2\delta}, \delta/5) + O(\log T) \le 6c\delta^{-z} + O(\log T).$$

Noting that a $(\delta/10)$-packing within $[0,1]$ can have cardinality at most $10\delta^{-1}$, we further bound $|Y_\delta| = O(\log T + \delta^{-1})$. Thus, by Lemma D.3, the regret incurred due to posting prices in Y_δ is at most

$$\begin{aligned} 2\delta \cdot \sum_{p \in Y_\delta} O\big(\sigma^2(p)\log(T)\delta^{-2} + \log(T)\delta^{-1}\big) &= O\left(\log(T)\delta^{-1}\sum_{p \in Y_\delta}\sigma^2(p) + \log(T)|Y_\delta|\right) \\ &= O\Big(\log(T)\big(c\delta^{-1-z} + \log(T)\delta^{-1}\big) + \log(T)\big(\log T + \delta^{-1}\big)\Big) \\ &= O\big(c\log(T)\delta^{-1-z} + \log^2(T)\delta^{-1}\big) \end{aligned}$$

Now we sum over $\delta = 1/2, 1/4, \dots, \alpha$, where α will be tuned later, giving a total regret bound of

$$\sum_{j=1}^{\log(1/\alpha)} \left[c\log(T)2^{j(1+z)} + \log^2(T)2^j\right] + \alpha T = O\Big(c\log(T)\alpha^{-(1+z)} + \log^2(T)\alpha^{-1} + \alpha T\Big).$$

Taking $\alpha = (c\log(T)/T)^{1/(2+z)}$, we obtain the desired regret bound of $\widetilde{O}\big(c^{1/(2+z)}T^{1-1/(2+z)}\big)$. □

E Proofs for Section 5

We first recall some standard results from Vapnik–Chervonenkis (VC) theory. Given two distributions D, D' over a shared space $\mathcal{X}$, we define the total variation distance $\|D - D'\|_{\text{TV}} \coloneqq \sup_{A \subseteq \mathcal{X}} |D(A) - D'(A)|$.

Lemma E.1 (Section 28.1 of 25)**.** *Fix a finite set $\mathcal{X}$, a function family $\mathcal{F} \subseteq \{0,1\}^{\mathcal{X}}$, a distribution $D \in \Delta(\mathcal{X})$, and $\delta \in (0,1)$. Then, for $X_1, \dots, X_n$ sampled i.i.d. from D, we have*

$$\sup_{f \in \mathcal{F}} \left| \mathop{\mathbb{E}}_{x \sim D}[f(x)] - \frac{1}{n}\sum_{i=1}^n f(X_i) \right| = O\left(\sqrt{\frac{V\log(n/V) + \log(1/\delta)}{n}}\right)$$

with probability at least $1 - \delta$, where V is the VC dimension of $\mathcal{F}$.

Lemma E.2 (Theorem 9.3 of 25)**.** *The family $\mathcal{F} = \{0,1\}^{\mathcal{X}}$ has VC dimension $|\mathcal{X}|$, and the family of linear thresholds over $\mathbb{R}^d$ has VC dimension $d+1$. The former result implies that, under the setting above, we have*

$$\|D - \hat{D}_n\|_{\mathrm{TV}} = O\left(\sqrt{\frac{|\mathcal{X}|\log(n/|\mathcal{X}|) + \log(1/\delta)}{n}}\right)$$

with probability at least $1-\delta$.

We also recall Bernstein's inequality for the case of i.i.d. Bernoulli random variables.

Lemma E.3 (Theorem 2.10 of 3)**.** *For i.i.d. $X_1, \dots, X_n \sim \mathrm{Ber}(p)$ and $\delta > 0$,*

$$\sum_{i=1}^{n} X_i \le pn + \sqrt{2np\log(1/\delta)} + \log(1/\delta)/3$$

with probability at least $1-\delta$.

We now turn to the main proofs.

E.1 Observed Type Identifiers (Proof of Theorem 5.1)

To state our result for the first setting, we introduce an ε-ball performance metric for contextual search. This is a slight strengthening of the standard ε-ball metric $\sum_{t=1}^{T} \mathbb{1}\{|p_t - v_t| > \varepsilon\}$.

Definition E.4 (Strong ε-ball regret)**.** Let $\mathcal{A}$ be a contextual pricing algorithm which, at round $t \in [T]$ with context u_t, outputs a price $p_t \in [0,1]$ along with a confidence width $w_t \in [0,1]$. For $\varepsilon \in (0,1]$, we say that $\mathcal{A}$ achieves strong ε-ball regret $R(T)$ for contextual search if, when $K_\star = 1$ and $D_\star = \delta_{\theta_\star}$, we have $|p_t - v_t| \le w_t$ for each round t and $\sum_{t=1}^{T} \mathbb{1}\{w_t > \varepsilon\} \le R(T)$, where $v_t = \langle u_t, \theta_\star \rangle$.

That is, $\mathcal{A}$ produces ε-accurate estimates for the true values, outside of up to $R(T)$ rounds for exploration, and it can identify when these estimates are accurate. In practice, this tends to require that $\mathcal{A}$ maintain a confidence set around $\theta_\star$ whose width, when projected onto the current context, is greater than ε for at most $R(T)$ rounds. Fortunately, there are existing efficient algorithms which achieve low ε-ball regret.

Lemma E.5 (18)**.** *For $\varepsilon \in (0,1]$, there exists an ellipsoid-based contextual search algorithm* `ProjectedVolume`(ε) *with strong ε-ball regret $O(d\log(d/\varepsilon))$ and running time $\mathrm{poly}(d, 1/\varepsilon)$ per round.*[2]

We now present our algorithm (Algorithm 5), which uses `ProjectedVolume` as a subroutine.

Overview of Algorithm 5 We maintain a set $\mathcal{I}$ of observed types, initially empty, and an accuracy ε (tuned to minimize regret). We will initialize an independent copy $\mathcal{A}_i$ of `ProjectedVolume` for each i added to $\mathcal{I}$. Since these copies are simulated, we are free to query the price $\mathsf{price}(\mathcal{A}_i, u_t)$ and confidence width $\mathsf{width}(\mathcal{A}_i, u_t)$ for a context u_t without updating $\mathcal{A}_i$. Moreover, for each $i \in \mathcal{I}$, the algorithm maintains a frequency count $n_t(i)$, recording the number of rounds which we have followed the recommended price of $\mathcal{A}_i$, along with an exploration count $m_t(i)$, recording the number of rounds which we have played the price of $\mathcal{A}_i$ due to its lack of confidence along the current context. At each round t, we perform the following:

- If there is an observed type $i \in \mathcal{I}$ such that $\mathsf{width}(\mathcal{A}_i, u_t) > \varepsilon$ and that its number of exploration plays $m_t(i)$ is below a threshold of $\varepsilon T/K_\star$, the algorithm plays $\mathsf{price}(\mathcal{A}_i, u_t)$, observes the outcome and the realized type z_t, and updates $\mathcal{A}_i$ if $z_t = i$. In addition, we increment $n_t(i)$ and $m_t(i)$.
- Otherwise, it defines active set $\mathcal{S} = \{i \in \mathcal{I} : \mathsf{width}(\mathcal{A}_i, u_t) \le \varepsilon\}$, computes for each $i \in S$ the score
$$F(i) = \sum_{j \in \mathcal{S}} \frac{n_t(j)}{t-1} \mathbb{1}\Big\{\mathsf{price}(\mathcal{A}_j, u_t) \ge \mathsf{price}(\mathcal{A}_i, u_t)\Big\},$$

[2] Although [18] state a slightly weaker guarantee, instead bounding $\sum_{t=1}^{T} \mathbb{1}\{|p_t - v_t| > \varepsilon\} = O(d\log(d/\varepsilon))$, this strengthened result is immediate from their proof.

and plays $p_t = \max\{\mathsf{price}(\mathcal{A}_{i^*}, u_t) - \varepsilon, 0\}$ where $i^* \in \arg\max_{i\in\mathcal{S}}\{F(i)\cdot \mathsf{price}(\mathcal{A}_i, u_t)\}$. It then observes y_t and z_t and updates the frequency count $n_{t+1}(z_t)$. Here, F is an estimate for the demand at the price suggested by $\mathcal{A}_i$, and so $i^\star$ is an estimate for the revenue maximizing type. We pull back the price recommended by $\mathcal{A}_{i^\star}$ by ε to avoid issues due to estimation error.

ALGORITHM 5: Contextual Pricing with Ex-Post Type Identification

initialize: observed types $\mathcal{I} = \varnothing$, $\varepsilon = \sqrt{d\log(T)/T}$;
for *each round* $t \in [T]$ **do**
  observe context u_t;
  if *exists* $i \in \mathcal{I}$ *such that* $\mathsf{width}(\mathcal{A}_i, u_t) > \varepsilon$ *and* $m_t(i) < \varepsilon T$ **then**
    play $p_t = \mathsf{price}(\mathcal{A}_i, u_t)$ and observe y_t;
    observe type $z_t \in [K_\star]$;
    update algorithm $\mathcal{A}_i$ with y_t if $z_t = i$;
    increment $m_t(i)$ by 1;
  else
    let $\mathcal{S} = \{i \in \mathcal{I} : \mathsf{width}(\mathcal{A}_i, u_t) \le \varepsilon\}$;
    define $F(i) = \sum_{j\in\mathcal{S}} \frac{n_t(j)}{t-1} \cdot \mathbb{1}\{\mathsf{price}(\mathcal{A}_j, u_t) \ge \mathsf{price}(\mathcal{A}_i, u_t)\}$ for each $i \in \mathcal{S}$;
    set $i^* = \arg\max_{i\in\mathcal{S}} F(i)\cdot \mathsf{price}(\mathcal{A}_i, u_t)$;
    play $p_t = \max\{\mathsf{price}(\mathcal{A}_{i^*}, u_t) - \varepsilon, 0\}$ and observe y_t;
    observe type $z_t \in [K_\star]$;
  increment $n_t(z_t)$ by 1;
  if $z_t \notin \mathcal{I}$ **then**
    initialize copy $\mathcal{A}_{z_t}$ of `ProjectedVolume`(ε) and set $\mathcal{I} \leftarrow \mathcal{I} \cup \{z_t\}$;

Bounding exploration regret. Write $\mathcal{T}_1$ for the set of exploration rounds. By design, an exploration round is one in which some type i is used with $\mathsf{width}(\mathcal{A}_i, u_t) > \varepsilon$ and exploration counter satisfying $m_t(i) < \varepsilon T$. Trivially, $|\mathcal{T}_1| \le \varepsilon T$, so we can incur regret at most $\varepsilon T = \widetilde{O}(\sqrt{dT})$ during exploration.

Bounding mass of types which saturate exploration threshold. Next, consider any type i that has been explored sufficiently so that $m_T(i) = \varepsilon T$ after time T; denote by $\mathcal{S}'$ the set of all such types. We will show that $D_\star$ plays small mass on $\mathcal{S}'$. Fix $i \in \mathcal{S}'$ and write $\mathcal{T}_{1,i}$ for the exploration rounds where we follow $\mathcal{A}_i$. Conditioned on $\mathcal{T}_{1,i}$, we note that $X_t = \mathbb{1}\{z_t = i\}$, $t \in \mathcal{T}_{1,i}$, are i.i.d. Bernoulli random variables with $\Pr(X_t = 1) = D_\star(i)$. Defining

$$S_i = \sum_{\tau\in\mathcal{T}_{1,i}} X_\tau,$$

our guarantee for `ProjectedVolume` (Lemma E.5) and the width condition for exploration imply that $S_i = O(d\log(d/\varepsilon))$. On the other hand, by Bernstein's inequality (Lemma E.3), we have

$$S_i \ge m_T(i)D_\star(i) - \sqrt{2\,m_T(i)D_\star(i)\,\log(T)} - \frac{1}{3}\log(T).$$

with probability at least $1 - 1/T$. Since $m_T(i) = \varepsilon T = \sqrt{dT\log(T)}$, the dominant term is $m_T(i)D_\star(i)$ for T greater than a sufficiently large constant. Thus, we deduce that

$$\varepsilon T\, D_\star(i) \le O\Big(d\log\frac{d}{\varepsilon}\Big),$$

or equivalently,

$$D_\star(i) \le O\Big(\frac{d\log(d/\varepsilon)}{\varepsilon T}\Big).$$

Summing over all types in $\mathcal{S}'$ (of which there are at most $K_\star$) and taking a union bound, the total mass in $\mathcal{S}'$ is at most

$$D_\star(\mathcal{S}') = \sum_{i\in\mathcal{S}'} D_\star(i) \le O\Big(\frac{K_\star\, d\log(d/\varepsilon)}{\varepsilon T}\Big) = \widetilde{O}\Big(K_\star\sqrt{d/T}\Big) \tag{7}$$

with probability at least $1 - K_\star/T$. We condition on this bound holding for the remainder of the proof, since doing so contributes a negligible $K_\star$ to the regret. We also condition on the event that, for each round $t \in [T]$ the empirical frequencies of (all) types deviate from their true masses by at most $O\big(\sqrt{K_\star \log(T)/t}\big)$ in total variation. This is permissible by Lemma E.2 and a union bound over rounds.

Bounding exploitation regret. Fix an exploitation round t, and recall the set of accurately estimated types

$$\mathcal{S} = \{i \in \mathcal{I} : \mathsf{width}(\mathcal{A}_i, u_t) \le \varepsilon\}.$$

Write $v_1, \ldots, v_{K_\star} \in [0,1]$ for the true values at round t. By our construction and the ε-ball guarantee for `ProjectedVolume`, $\mathcal{A}_i$ returns a price that is an ε-accurate estimate of v_i, for each $i \in \mathcal{S}$. Moreover, by our analysis above, the mass on types outside of $\mathcal{S}$ is quite small. We thus bound

$$\begin{aligned}
\mathsf{dem}_\star(p_t, u_t) &= \sum_{i=1}^{K_\star} D_\star(i)\mathbb{1}\{v_i \ge p_t\} \\
&\ge \sum_{i \in \mathcal{S}} D_\star(i)\mathbb{1}\{v_i \ge p_t\} - \widetilde{O}\Big(K_\star\sqrt{d/T}\Big) && \text{(Eq. (7))} \\
&\ge \sum_{i \in \mathcal{S}} \frac{n_t(i)}{t-1}\mathbb{1}\{v_i \ge p_t\} - \widetilde{O}\Big(K_\star\sqrt{d/T}\Big) - O\big(\sqrt{K_\star \log(T)/t}\big) && \text{(TV bound)} \\
&\ge \sum_{i \in \mathcal{S}} \frac{n_t(i)}{t-1}\mathbb{1}\{v_i \ge \mathsf{price}(\mathcal{A}_{i^\star}, u_t) - \varepsilon\} - \widetilde{O}\Big(K_\star\sqrt{d\log(T)/t}\Big) && \text{(choice of } p_t) \\
&\ge \sum_{i \in \mathcal{S}} \frac{n_t(i)}{t-1}\mathbb{1}\{\mathsf{price}(\mathcal{A}_i, u_t) \ge \mathsf{price}(\mathcal{A}_{i^\star}, u_t)\} - \widetilde{O}\Big(K_\star\sqrt{d\log(T)/t}\Big) && (i \in \mathcal{S}) \\
&= F(i_\star) - \widetilde{O}\Big(K_\star\sqrt{d\log(T)/t}\Big). && \text{(choice of } F)
\end{aligned}$$

Consequently, we bound $\mathsf{rev}_\star(p_t, u_t) + \widetilde{O}\Big(K_\star\sqrt{d\log(T)/t}\Big)$ from below by

$$\begin{aligned}
p_t F(i^*) &\ge \mathsf{price}(\mathcal{A}_{i*}, u_t)F(i^*) - \varepsilon \\
&= \max_{j \in S} \mathsf{price}(\mathcal{A}_j, u_t)F(j) - \varepsilon && \text{(choice of } i^*) \\
&= \max_{p \in [0,1]} p \sum_{i \in \mathcal{S}} \frac{n_t(i)}{t-1}\mathbb{1}\{\mathsf{price}(\mathcal{A}_i, u_t) \ge p\} - \varepsilon && \text{(rev. maximized at jump)} \\
&\ge \max_{p \in [0,1]} p \sum_{i \in \mathcal{S}} D_\star(i)\mathbb{1}\{\mathsf{price}(\mathcal{A}_i, u_t) \ge p\} - \varepsilon - \widetilde{O}\Big(K_\star\sqrt{d\log(T)/t}\Big) && \text{(TV bound)} \\
&\ge \max_{p \in [0,1]} p \sum_{i \in \mathcal{S}} D_\star(i)\mathbb{1}\{v_i \ge p + \varepsilon\} - \varepsilon - \widetilde{O}\Big(K_\star\sqrt{d\log(T)/t}\Big) && (i \in S) \\
&\ge \max_{p \in [0,1]} p \sum_{i=1}^{K_\star} D_\star(i)\mathbb{1}\{v_i \ge p + \varepsilon\} - \varepsilon - \widetilde{O}\Big(K_\star\sqrt{d\log(T)/t}\Big) && \text{(Eq. (7))} \\
&= \max_{p \in [0,1]} p\, \mathsf{dem}_\star(p + \varepsilon, u_t) - \varepsilon - \widetilde{O}\Big(K_\star\sqrt{d\log(T)/t}\Big) \\
&= \max_{p \in [0,1]} \mathsf{rev}_\star(p + \varepsilon, u_t) - 2\varepsilon - \widetilde{O}\Big(K_\star\sqrt{d\log(T)/t}\Big) \\
&= \max_{p \in [0,1]} \mathsf{rev}_\star(p, u_t) - 3\varepsilon - \widetilde{O}\Big(K_\star\sqrt{d\log(T)/t}\Big).
\end{aligned}$$

All together, we see that playing p_t incurs regret at most $\widetilde{O}\Big(K_\star\sqrt{d\log(T)/t}\Big)$. Summing over exploitation rounds and adding the exploration regret gives a total bound of

$$R(T) = \sum_{t=1}^{T} \widetilde{O}\Big(K_\star\sqrt{d\log(T)/t}\Big) + \widetilde{O}(\sqrt{dT}) = \widetilde{O}(K_\star\sqrt{dT}),$$

as desired. $\square$

E.2 Observed Type Vectors (Proof of Theorem 5.2)

We first show that $\mathsf{rev}_{\hat{D}_\tau}$ concentrates tightly around $\mathsf{rev}_\star$, using a simple VC bound.

Lemma E.6. *Fix $D \in \Delta_K(\Theta)$ and let $\hat{D}_t$ be the empirical measure of t i.i.d. samples from D. We then have*

$$\sup_{p\in[0,1],u\in\mathcal{U}} |\mathsf{rev}_D(p,u) - \mathsf{rev}_{\hat{D}_t}(p,u)| = O\left(\sqrt{\frac{\min\{K,d\}\log(t) + \log(1/\delta)}{t}}\right)$$

with probability at least $1-\delta$.

Proof. We compute

$$\begin{aligned}\sup_{p\in[0,1],u\in\mathcal{U}} |\mathsf{rev}_D(p,u) - \mathsf{rev}_{\hat{D}_t}(p,u)| &\le \sup_{p\in[0,1],u\in\mathcal{U}} |\mathsf{dem}_D(p,u) - \mathsf{dem}_{\hat{D}_t}(p,u)| \\ &= \sup_{f\in\mathcal{F}}\left|\mathop{\mathbb{E}}_{\theta\sim D}[f(\theta)] - \mathop{\mathbb{E}}_{\theta\sim\hat{D}_t}[f(\theta)]\right|,\end{aligned}$$

where $\mathcal{F}$ is the space of linear threshold functions $f_{p,u} : \mathrm{supp}(D) \to \{0,1\}$ given by $f_{p,u}(\theta) = \mathbb{1}\{\langle u,\theta\rangle \ge p\}$. The result then follows by Lemma E.2. □

Now, our best response policy ensures that, at each round $t > 1$, we have

$$\begin{aligned}\mathsf{rev}_\star(p_t,u_t) &= \mathsf{rev}_{\hat{D}_{t-1}}(p_t,u_t) + \mathsf{rev}_\star(p_t,u_t) - \mathsf{rev}_{\hat{D}_{t-1}}(p_t,u_t) \\ &\ge \mathsf{rev}_{\hat{D}_{t-1}}(\mathsf{br}_\star(u_t),u_t) - \sup_{p\in[0,1],u\in\mathcal{U}} |\mathsf{rev}_\star(p,u) - \mathsf{rev}_{\hat{D}_{t-1}}(p,u)| \\ &\ge \mathsf{rev}_\star(\mathsf{br}_\star(u_t),u_t) - 2\sup_{p\in[0,1],u\in\mathcal{U}} |\mathsf{rev}_\star(p,u) - \mathsf{rev}_{\hat{D}_{t-1}}(p,u)|.\end{aligned}$$

Consequently, regret is at most

$$\begin{aligned}R(T) &\le 1 + 2\sum_{t=1}^{T-1} \sup_{p\in[0,1],u\in\mathcal{U}} |\mathsf{rev}_\star(p,u) - \mathsf{rev}_{\hat{D}_t}(p,u)| \\ &= O\left(\sqrt{\min\{K_\star,d\}}\sum_{t=1}^{T-1}\sqrt{\frac{\log(t)}{t}}\right) && \text{(Lemma E.6 with } \delta = t^{-2}\text{)} \\ &= \widetilde{O}(\sqrt{\min\{K_\star,d\}T}),\end{aligned}$$

as desired. □