OPTIMAL CONTEXTUAL PRICING IN NON-STATIONARY ENVIRONMENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study contextual dynamic pricing in a non-stationary environment. In each round, a seller observes a buyer's context, posts a price, and receives only binary purchase feedback. The buyer's value is linear in the context and an unobserved feature vector that may change over time, subject to a known total variation budget V. We measure dynamic regret against the first-best benchmark that extracts each buyer's full value.

For identical contexts, we develop a deterministic algorithm that combines binary-search exploration, lower-envelope pricing, and two complementary restart mechanisms. It achieves regret $O((\sqrt{TV}+1)\log T)$. With heterogeneous contexts, a new geometric difficulty arises: small feature changes can leave a relaxed feasible set narrow but far from the current feature vector. To address this difficulty, we introduce a multiscale shallow-pricing algorithm that maintains relaxed feasible sets and ellipsoidal approximations at several variation scales. Larger scales provide reliable localization certificates, while the smallest active scale determines the exploitation price. The algorithm achieves dynamic regret $O((d^2 + d^{3/2}\sqrt{TV})\log^2 T)$. Both guarantees match the identical-context lower bound $\Omega(\sqrt{TV} + \log\log T)$ up to logarithmic factors.

1 INTRODUCTION

Online sellers repeatedly make pricing decisions for buyers with different observable characteristics and changing willingness to pay. The economic scale of these decisions is substantial: U.S. consumers spent $257.8 billion online during November and December 2025, including $44.2 billion during the five-day Cyber Week (Adobe, 2026). Observable context can help a seller tailor prices to individual transactions, but the relationship between context and willingness to pay may change with seasonality, preferences, and market conditions. Learning this relationship is particularly challenging because pricing feedback is censored: a purchase or rejection reveals only whether the buyer's value exceeds the posted price. An effective pricing algorithm must therefore learn from limited feedback while adapting to a changing environment.

We study contextual dynamic pricing over T rounds. In each round, a new buyer arrives with an observable d-dimensional context and an unobserved d-dimensional feature vector. The buyer's value is the inner product of these two vectors. Contexts are nonnegative unit vectors, and hidden features are nonnegative vectors with Euclidean norm at most one, so values lie between zero and one. After observing the context, the seller posts a price and observes only whether the buyer purchases. Contexts may vary arbitrarily across rounds. The hidden features may also change, subject to a known upper bound V on the sum of their Euclidean changes between consecutive rounds. Our objective is to minimize *dynamic regret* relative to the *first-best benchmark*: the total revenue of an oracle that knows each buyer's value and posts that value as the price. This benchmark extracts the full surplus in every round and is stronger than comparison with the best single fixed price in hindsight (Kleinberg & Leighton, 2003).

Our problem builds on the study of online posted-price learning initiated by Kleinberg & Leighton (2003) and connects two lines of dynamic pricing research. Stationary contextual pricing studies heterogeneous contexts with a fixed hidden parameter (Paes Leme & Schneider, 2018; Cohen et al., 2020; Liu et al., 2021), whereas Paes Leme et al. (2021) study non-contextual pricing with changing

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

values under bounds on their per-round changes. We consider heterogeneous contexts together with changing hidden features under a known aggregate variation budget. Non-stationary contextual pricing has also been studied under stochastic valuation noise and demand models (Javanmard, 2017; Baby et al., 2023); our focus is the noiseless linear-valuation model and regret against the first-best benchmark. (A comprehensive discussion on the further related work can be found in Section A.) This leads to the following question:

> *What is the optimal dynamic regret in this setting, and what additional difficulties arise from the interaction between contextual heterogeneity and non-stationarity?*

Our results and techniques. We develop a deterministic contextual pricing algorithm with dynamic regret $O\Big((d^2 + d^{3/2}\sqrt{TV})\log^2 T\Big)$. Together with a lower bound of $\Omega(\sqrt{TV} + \log\log T)$, this establishes optimality up to logarithmic factors in T for fixed d among deterministic algorithms. The guarantee holds for every variation budget $V \in [0, T]$, including the stationary case. Our algorithm combines periodic and inconsistency restarts with a multiscale extension of shallow pricing.

We first develop the restart mechanism in the identical-context environment (Section 3). In this setting, binary search followed by lower-envelope pricing provides a simple exploration-then-exploitation procedure. Non-stationarity creates two distinct difficulties: an increase in values can leave an accepted price persistently too low, whereas even a small decrease can turn a previously accepted price into a rejection and cause a large revenue loss. We address these difficulties through periodic restarts and immediate restarts after rejected exploitation prices. A price margin controls the frequency of the latter by ensuring that each such rejection requires sufficient variation. This yields dynamic regret $O((\sqrt{TV} + 1)\log T)$.

Extending this argument to heterogeneous contexts requires additional geometric reasoning. A narrow interval directly localizes a scalar value, but localization in one context direction need not provide useful information in another. We therefore build on the shallow-pricing procedure of Cohen et al. (2020), which uses ellipsoidal approximations to guide exploration. To accommodate non-stationarity, we relax the halfspace constraints induced by purchase feedback. However, using a single relaxation scale introduces a new difficulty: a relaxed feasible set can become narrow while excluding the current hidden parameter. Nearly parallel context directions can amplify small parameter changes into a large localization error. Our Example 4.1 shows why the identical-context exploitation argument does not extend directly.

Our main algorithm resolves this difficulty by maintaining relaxed feasible sets and enclosing ellipsoids at multiple dyadic scales (Section 4). It explores whenever an active scale indicates substantial uncertainty and otherwise posts the lower-envelope price of the smallest active feasible set. The key observation is that a scale at least as large as the within-batch parameter diameter retains every hidden parameter in that batch. Its ellipsoid therefore provides a reliable certificate of localization. Nesting of the feasible sets then controls the regret of accepted exploitation prices, even when the smallest active set excludes the current parameter. Rejected exploitation prices trigger immediate restarts. Combining this multiscale argument with geometric bounds on exploration and variation-based bounds on the number of restarts yields our dynamic regret guarantee.

2 PRELIMINARIES

We introduce the contextual dynamic pricing problem in a non-stationary environment.

Environment. There is a single seller who interacts with buyers over T discrete rounds. In each round $t \in [T] \triangleq \{1, 2, \ldots, T\}$, a new buyer, indexed by t, arrives. Each buyer t has a public d-dimensional context vector $u_t \in [0, 1]^d$ with[1] $\|u_t\| = 1$ and a private d-dimensional feature vector $\theta_t \geq 0, \|\theta_t\| \leq 1$. After observing the public context u_t, the seller posts a take-it-or-leave-it price $p_t \in [0, 1]$ for a single product. Buyer t's value (i.e., willingness to pay) takes the linear form $v_t \triangleq \langle \theta_t, u_t \rangle \in [0, 1]$. Thus, buyer t purchases the product at price p_t if and only if her value is at least the price, i.e., $v_t \geq p_t$. The seller observes only whether the buyer purchases the product

[1] Throughout this paper, we use $\|\cdot\|$ to denote the Euclidean norm.

and does not observe her value. (We denote the purchase feedback by $f_t \triangleq \mathbb{I}\{v_t \geq p_t\} \in \{0,1\}$.) The seller's goal is to maximize her revenue, i.e., the total payments from all T buyers. We assume $T \geq 2$. No stochastic assumptions are imposed on the context or feature sequences.

Dynamic regret minimization. We evaluate an online learning algorithm by its regret relative to the *first-best benchmark* (i.e., full surplus), $\sum_{t\in[T]} v_t = \sum_{t\in[T]} \langle \theta_t, u_t \rangle$. This benchmark is achieved by posting the full-surplus-extraction price $p_t^* = v_t$ for each buyer t. Specifically, given a sequence of T buyers described by $\{u_t, \theta_t\}_{t\in[T]}$, the dynamic regret of an algorithm ALG is

$$\textsc{Regret}[\textsc{Alg}] \triangleq \sum\nolimits_{t\in[T]} \langle \theta_t, u_t \rangle - \sum\nolimits_{t\in[T]} p_t \cdot \mathbb{I}\{\langle \theta_t, u_t \rangle \geq p_t\}\,,$$

where $\{p_t\}_{t\in[T]}$ are the prices posted by ALG.

Non-stationarity upper bound. We measure the *non-stationarity* of the unknown feature sequence $\{\theta_t\}_{t\in[T]} = (\theta_t)_{t\in[T]}$ by its total variation, $\mathcal{V}(\{\theta_t\}_{t\in[T]}) \triangleq \sum_{t\in[T-1]} \|\theta_{t+1} - \theta_t\|$. Although the feature sequence $\{\theta_t\}_{t\in[T]}$ remains unobserved by the seller throughout the T rounds, we assume that the seller knows an upper bound V on its total variation in advance, i.e., $\mathcal{V}(\{\theta_t\}_{t\in[T]}) \leq V$. We focus on $V \in [0, T]$; larger variation budgets admit the trivial regret guarantee T. In the special case of zero non-stationarity, i.e., $V = 0$, our problem reduces to the classical stationary contextual dynamic pricing problem studied in Kleinberg & Leighton (2003); Paes Leme & Schneider (2018); Cohen et al. (2020); Liu et al. (2021).

3 Warm-Up: Optimal Dynamic Pricing for Identical Contexts

We begin with the identical-context environment, where all buyers have the same public context, denoted by u. This section serves as a warm-up for the more general environment with heterogeneous contexts: its algorithmic ingredients and regret analysis closely parallel, and lay the groundwork for, the design and analysis of our main algorithm in Section 4.

Although the public contexts are identical, the private features $\{\theta_t\}_{t\in[T]}$ may still vary across buyers. We can therefore view this setting as a pricing problem without public contexts, in which buyers have potentially different private values $\{v_t\}_{t\in[T]}$. For ease of presentation, we omit $\{u_t, \theta_t\}_{t\in[T]}$ in this section and work directly with the value sequence $\{v_t\}_{t\in[T]}$. Under the normalization in our model, $v_t \in [0, 1]$ for every buyer t. Moreover, since identical context u is unit-length (i.e., $\|u\| = 1$), we have $\sum_{t\in[T-1]} |v_{t+1} - v_t| \leq \sum_{t\in[T-1]} \|\theta_{t+1} - \theta_t\| \leq V$. We consider non-stationarity upper bounds $V \in [0, T]$.

Below, we present the algorithm for this environment and its dynamic regret guarantee. We then conclude the section with a sketch of the regret analysis.

Algorithm overview. For stationary online learning problems, a standard approach is *explore-then-exploit* (ETE), also commonly called *explore-then-commit* in the bandit literature (Lattimore & Szepesvári, 2020, Chapter 6). When all buyers have the same value, this approach can be implemented by first conducting binary search for $\lceil \log_2 T \rceil$ rounds to locate the value within an interval of width at most $1/T$. This exploration phase incurs at most $\lceil \log_2 T \rceil$ regret. During the subsequent exploitation phase, posting the lower endpoint of this interval in every remaining round incurs at most $O(1)$ additional regret.[2]

For non-stationary environments, a standard approach is to restart a learning algorithm *periodically*. For example, in non-stationary multi-armed bandit problems, Besbes et al. (2014) show that periodically restarting Exp3 achieves near-optimal dynamic regret under a variation-budget constraint. However, the revenue structure of dynamic pricing creates an additional difficulty. An arbitrarily small decrease in a buyer's value can reduce the revenue from a previously optimal price to zero: if the seller continues posting $p_{t+1} = v_t$ but $v_{t+1} < v_t$, the buyer rejects the offer. Thus, even an

[2] In the setting with an unknown identical value, the optimal regret bound of $O(\log \log T)$ can be achieved by a modified ETE approach that replaces ordinary binary search with an asymmetric search procedure (Kleinberg & Leighton, 2003).

Algorithm 1 Pricing for the identical-context environment

Input: $T \in \mathbb{N}, V \in [0, T]$
Parameters: $\delta \leftarrow \max\{1/T, \sqrt{V/T}\}$, $a \leftarrow \min\{T, \lceil \sqrt{T/V} \rceil\}$ ▷ set $a \leftarrow T$ when $V = 0$
$t \leftarrow 1, i \leftarrow 0$
while $t \leq T$ **do**
  $i \leftarrow i + 1, \tau \leftarrow \min\{T, t + a - 1\}, [p_i^{\min}, p_i^{\max}] \leftarrow [0, 1]$
  while $t \leq \tau$ and $p_i^{\max} - p_i^{\min} > \delta$ **do**
    Post $p_t \leftarrow (p_i^{\min} + p_i^{\max})/2$; observe f_t
    if $f_t = 1$, set $p_i^{\min} \leftarrow p_t$; **else** set $p_i^{\max} \leftarrow p_t$
    $t \leftarrow t + 1$
  end while
  $\underline{p}_i \leftarrow \max\{0, p_i^{\min} - \delta\}$
  while $t \leq \tau$ **do**
    Post $p_t \leftarrow \underline{p}_i$; observe f_t; set $t \leftarrow t + 1$
    if $f_{t-1} = 0$, **break** ▷ inconsistency restart
  end while
end while

arbitrarily small amount of variation can cause constant regret in every subsequent round until the next periodic restart, provided the buyer's value remains bounded away from zero.[3]

Motivated by this difficulty, our algorithm supplements periodic restarting with an *inconsistency restart*. Informally, if a buyer rejects a price that the algorithm posts during exploitation, the algorithm immediately restarts ETE. However, inconsistency restarting alone cannot guarantee sublinear regret either. To see this, consider a sequence in which all buyers during exploration have the same value. Immediately after exploration ends, the value increases by V and remains at this higher level for the rest of the horizon, where $V > 0$ is chosen so that all values remain in $[0, 1]$. No inconsistency restart occurs because buyers continue accepting the exploitation price. The regret in each exploitation round is at least V, leading to total regret of $\Omega(VT)$ when exploration takes $O(\log T)$ rounds. This regret is linear in T for any fixed admissible $V > 0$.

Our algorithm therefore combines periodic and inconsistency restarts. We call the consecutive rounds from initialization or a restart until the next restart a *batch*. Periodic restarting is controlled by a *window-size parameter* $a \in [T]$, which specifies the maximum number of rounds in a batch. This parameter balances the regret caused by non-stationarity during exploitation against the exploration cost of frequent periodic restarts.

Inconsistency restarting is controlled by a *price-margin parameter* $\delta \in (0, 1]$. This parameter determines both the target interval width during exploration and the price posted during exploitation. Specifically, once binary search produces an interval of width at most δ, the algorithm posts its lower endpoint minus δ, clipped at zero, during the subsequent exploitation phase. We call this price the *lower-envelope price* (LEP). Intuitively, δ balances the regret from conservative pricing during exploitation against the exploration cost of frequent inconsistency restarts. The formal algorithm is given in Algorithm 1.

Regret guarantee and its analysis. We now state the regret guarantee in Theorem 3.1.

Theorem 3.1. *In the identical-context environment, given any non-stationarity upper bound $V \in [0, T]$, set $a = \min\{T, \lceil \sqrt{T/V} \rceil\}$ and $\delta = \max\{1/T, \sqrt{V/T}\}$[4]. Then the dynamic regret of Algorithm 1 is at most $O(\sqrt{TV} \log T + \log T)$. Moreover, for every $V \in [0, T]$ and every deterministic online algorithm, there exists an admissible instance with total variation at most V on which the algorithm incurs dynamic regret of at least $\Omega(\sqrt{TV} + \log \log T)$.*

Combining the upper and lower bounds in Theorem 3.1, the regret guarantee of Algorithm 1 is optimal among deterministic online algorithms, up to logarithmic factors in T, for every $V \in [0, T]$. The lower bound combines the variation-dependent term $\Omega(\sqrt{TV})$, obtained by adapting the construc-

[3] For comparison, in a bandit problem with a fixed set of arms, an arm that was optimal at an earlier round has a current suboptimality gap at most twice the largest change in any arm's mean reward since that round.

[4] with the convention $a = T$ when $V = 0$.

tion in Paes Leme et al. (2021, Theorem 4.2), with the stationary lower bound $\Omega(\log \log T)$ from Kleinberg & Leighton (2003). The former considers the stronger restriction $|v_{t+1} - v_t| \leq V/T$ on changes between consecutive rounds,[5] while the latter applies when all buyers have the same value. We conclude this section with a sketch of the upper-bound proof. The full proof appears in Section C.

To bound the dynamic regret of Algorithm 1, we bound the regret from exploration and exploitation separately. During exploration, binary search takes at most $\lceil \log_2(1/\delta) \rceil$ rounds per batch. Since each round incurs at most one unit of regret, the total exploration regret is at most $m\lceil \log_2(1/\delta) \rceil$, where m is the number of batches. Periodic restarts contribute at most T/a completed batches, and each inconsistency restart requires at least δ variation within its batch. Hence, $m \leq 1 + T/a + V/\delta = O(\sqrt{TV} + 1)$. Since $\delta \geq 1/T$, the total exploration regret is therefore at most $O((\sqrt{TV} + 1) \log T) = O(\sqrt{TV} \log T + \log T)$.

To analyze exploitation, let $B_i \subseteq [T]$ denote the rounds in batch i, and define its value range as $\Delta_i \triangleq \max_{s,t \in B_i} |v_s - v_t|$. These within-batch value ranges satisfy $\sum_{i \in [m]} \Delta_i \leq V$. Each accepted exploitation round in batch i incurs regret of at most $2\delta + \Delta_i$. Each rejected exploitation round incurs at most one unit of regret and immediately ends the batch, so there is at most one such round per batch. Since each batch contains at most a rounds, the total exploitation regret is at most $2T\delta + \sum_{i \in [m]} |B_i| \Delta_i + m \leq 2T\delta + aV + m = O(\sqrt{TV} + 1)$. Combining the exploration and exploitation bounds proves the stated upper bound.

4 Optimal Dynamic Pricing using Multi-Scale Shallow Pricing

In this section, we turn to the more general environment in which buyers have heterogeneous d-dimensional context vectors. We present our main result: an online learning algorithm whose dynamic regret has optimal dependence on T and the non-stationarity upper bound V, up to logarithmic and dimension-dependent factors. Our algorithm (Algorithm 2) and its regret analysis extend the approach developed for the identical-context environment in Section 3.

Compared with the identical-context setting, our algorithm and its analysis introduce two main technical ingredients.

The first ingredient is to replace binary-search exploration with *shallow pricing*, introduced by Cohen et al. (2020) for stationary contextual dynamic pricing, while retaining lower-envelope pricing (LEP) for exploitation. With identical contexts, uncertainty concerns a single scalar value and can be represented by an interval. With heterogeneous contexts, uncertainty concerns a vector, and learning its projection in one context direction need not provide accurate information in another. Shallow pricing maintains an ellipsoidal approximation of the feasible region and uses its width in the current context direction to determine whether further exploration is needed. Thus, exploration and exploitation may be interleaved as different contexts arrive.

The second ingredient is to maintain *multiple copies of the geometric learning procedure, each with a price-margin parameter λ drawn from a dyadic set $\Lambda = \{\delta, 2\delta, \ldots, 2^L\delta\}$*, whose largest scale covers the diameter of the feature domain. This multiscale construction addresses a difficulty arising from the combination of heterogeneous contexts and non-stationary features when $d \geq 2$. As illustrated in Example 4.1, a relaxed feasible set can become narrow in the current context direction while remaining far from the current hidden feature. Maintaining multiple scales ensures that a larger-scale set can still provide a valid certificate of localization. This particular difficulty does not arise in either the stationary contextual setting of Cohen et al. (2020) or the non-stationary identical-context setting of Section 3, where a single price-margin scale suffices for the respective regret guarantees.

[5] Under the assumption that consecutive value changes are bounded by unknown, nonincreasing rates ϵ_t satisfying $\sum_{t \in [T-1]} \epsilon_t \leq V$, Paes Leme et al. (2021) propose a randomized pricing algorithm whose analysis yields expected regret of $O(\sqrt{VT} \log T + \sqrt{T} \log T)$. Their algorithm organizes learning into phases and uses randomly selected checking rounds to detect inconsistencies. Our inconsistency test is deterministic and also serves as a building block of our algorithm for heterogeneous contexts.

Shallow pricing with LEP. Our contextual algorithms build on the geometric approach of Cohen et al. (2020): each purchase or rejection provides a halfspace constraint on the hidden feature vector. For a nonempty compact convex set $K \subseteq \mathbb{R}^d$, define its width in context direction u as

$$\mathrm{width}(K, u) \triangleq \max_{x \in K} \langle x, u \rangle - \min_{x \in K} \langle x, u \rangle .$$

At scale λ, we maintain a relaxed feasible set $S_t^{(\lambda)}$ and an enclosing ellipsoid $\mathcal{E}_t^{(\lambda)}$, where

$$\mathcal{E}_t^{(\lambda)} \triangleq \left\{ x \in \mathbb{R}^d : (x - c_t^{(\lambda)})^\top (A_t^{(\lambda)})^{-1} (x - c_t^{(\lambda)}) \le 1 \right\}, \quad \text{and} \quad \mathrm{width}(\mathcal{E}_t^{(\lambda)}, u_t) = 2\sqrt{u_t^\top A_t^{(\lambda)} u_t}.$$

Here, $c_t^{(\lambda)}$ is the center and $A_t^{(\lambda)}$ is symmetric positive definite. The feasible set stores the feedback constraints and determines the exploitation price; the ellipsoid determines whether further exploration is needed. We initialize the feasible set and its enclosing ellipsoid as $S_0 = \{x \in \mathbb{R}^d : x \ge 0,\ \|x\| \le 1\}$ and $\mathcal{E}_0 = \{x \in \mathbb{R}^d : \|x\| \le 1\}$, respectively.

The SHALLOWPRICING procedure explores at scale λ only when $\mathrm{width}(\mathcal{E}_t^{(\lambda)}, u_t) > 4d\lambda$. It proposes the midpoint price $\widetilde{p}_t = \left\langle c_t^{(\lambda)}, u_t \right\rangle$; our algorithm posts $p_t = \min\{1, \max\{0, \widetilde{p}_t\}\}$. After observing feedback $f_t = \mathbb{I}\{\text{Buyer } t \text{ accept price } p_t\}$, we form the relaxed halfspace

$$H_t^{(\lambda)} \triangleq \begin{cases} \{x \in \mathbb{R}^d : \langle x, u_t \rangle \le p_t + \lambda\}, & f_t = 0, \\ \{x \in \mathbb{R}^d : \langle x, u_t \rangle \ge p_t - \lambda\}, & f_t = 1. \end{cases}$$

Every observation updates the relaxed feasible sets using the actual price and feedback: $S_{t+1}^{(\lambda)} = S_t^{(\lambda)} \cap H_t^{(\lambda)}$. On an exploration round, only the selected ellipsoid is updated, using the inferred midpoint feedback $\widetilde{f}_t = f_t \mathbb{I}\{\widetilde{p}_t \le 1\}$. The operator $\mathrm{ShallowUpdate}(\mathcal{E}, u; \lambda, \widetilde{f})$ denotes the standard shallow update at the unclipped midpoint, with acceptance if $\widetilde{f} = 1$ and rejection if $\widetilde{f} = 0$. For $\widetilde{p}_t < 0$, the posted price is necessarily accepted, and its relaxed constraint implies the midpoint-based constraint. For $\widetilde{p}_t > 1$, the midpoint would necessarily be rejected, and every $x \in S_0$ satisfies $\langle x, u_t \rangle \le 1 < \widetilde{p}_t + \lambda$. Thus, clipping preserves the ellipsoidal outer approximation and the exploration guarantee. The update formulas are given in Section B.

For a nonempty compact feasible set S, its *lower-envelope price* in context direction u is

$$\mathrm{LEP}(S, u) \triangleq \max\{0, \min_{x \in S} \langle x, u \rangle\} .$$

This generalizes lower-endpoint pricing in the identical-context setting. Since $S_t^{(\lambda)} \subseteq \mathcal{E}_t^{(\lambda)}$, the ellipsoid also certifies that $\mathrm{width}(S_t^{(\lambda)}, u_t) \le \mathrm{width}(\mathcal{E}_t^{(\lambda)}, u_t)$.

Identical-context approach does not extend directly. A natural extension of the identical-context algorithm fixes a single price-margin parameter δ and maintains a relaxed feasible set $S_t^{(\delta)}$ together with an enclosing ellipsoid $\mathcal{E}_t^{(\delta)}$. In each round, the algorithm explores using SHALLOWPRICING if $\mathrm{width}(\mathcal{E}_t^{(\delta)}, u_t) > 4d\delta$; otherwise, it posts $p_t = \mathrm{LEP}(S_t^{(\delta)}, u_t)$. Every observation updates the relaxed feasible set, while only exploration rounds update the ellipsoid. The algorithm restarts when the window of a rounds expires, the relaxed feasible set becomes empty, or an exploitation price is rejected.

A crucial difficulty arises from using a single price-margin parameter. In the identical-context setting, each accepted exploitation round in batch B_i incurs regret of at most $2\delta + \Delta_i$, where Δ_i is the within-batch value range. With heterogeneous contexts, the analogous quantity is the within-batch parameter diameter, $\Delta_i \triangleq \max_{s,t \in B_i} \|\theta_s - \theta_t\|$. If $\Delta_i \le \delta$, every parameter in the batch satisfies all δ-relaxed feedback constraints. Thus, whenever the ellipsoid is sufficiently narrow for exploitation, the regret is at most $4d\delta$. However, when $\Delta_i > \delta$, even by only a constant factor, small parameter changes can cause a large localization error, as the following example illustrates. Section D provides a formal analysis of this single-scale extension. Optimizing the parameters in the resulting upper bound yields a dynamic regret guarantee of $O((d^2 + d^{3/2} V^{1/3} T^{2/3}) \log T)$. For fixed dimension, a matching lower bound up to logarithmic factors shows that this suboptimal rate is unavoidable under the selected tuning.

Example 4.1 (A narrow but misplaced relaxed feasible set)**.** *Let $d = 2$, fix $\delta \in (0, 1/16]$, and write $q = \sqrt{1+\delta^2}$. Initialize the relaxed feasible set to $S_0 = \{x \in \mathbb{R}^d : x \geq 0,\ \|x\| \leq 1\}$. Consider two price queries within a batch.*[6]

In round 1, let $u_1 = (0, 1)$, $\theta_1 = (1/2, 1/4 + \delta)$, and $p_1 = 1/4 + \delta$. The buyer accepts because $v_1 = p_1$, giving the relaxed constraint $x_2 \geq 1/4$. In round 2, let $u_2 = (\delta, 1)/q$, $\theta_2 = (1/2, 1/4-2\delta)$, and $p_2 = (1/4+\delta^2)/q-\delta$. The value is $v_2 = (1/4-3\delta/2)/q$, so $p_2-v_2 = \delta(\delta+3/2-q)/q > 0$. The buyer therefore rejects, giving the relaxed constraint $\delta x_1+x_2 \leq 1/4+\delta^2$. After these observations, the relaxed feasible set is $S^{(\delta)} = S_0 \cap \{x \in \mathbb{R}^d : x_2 \geq 1/4,\ \delta x_1 + x_2 \leq 1/4 + \delta^2\}$.

From round 3 onward, let $u_t = (1, 0)$ and $\theta_t = (1/2, 1/4 + \delta)$. All contexts, parameters, and prescribed prices satisfy the model assumptions. The parameter diameter is 3δ, and the total parameter variation is 6δ. Yet the two relaxed constraints imply $0 \leq x_1 \leq \delta$ for every $x \in S^{(\delta)}$. Since both $(0, 1/4)$ and $(\delta, 1/4)$ belong to this set, $\mathrm{width}(S^{(\delta)}, u_t) = \delta$.

For every round $t \geq 3$, the true value is $1/2$, whereas $\mathrm{LEP}(S^{(\delta)}, u_t) = 0$. This price is accepted and incurs regret $1/2$. Its relaxed feedback constraint, $x_1 \geq -\delta$, leaves the feasible set unchanged. Thus, if LEP continues to be used without a restart, every subsequent round incurs regret $1/2$, despite the parameter diameter being only 3δ.

The example shows that a small parameter diameter and a small directional width of the relaxed feasible set do not, by themselves, imply a regret bound of $O(\Delta_i + \delta)$ on accepted exploitation rounds. Here, $\Delta_i + \delta = 4\delta$, while the regret remains $1/2$ as δ tends to zero. The two initial contexts are nearly parallel. Combining their feedback constraints requires dividing by δ, which can amplify an $O(\delta)$ change in one coordinate into a constant localization error in another. With identical contexts, all constraints concern the same scalar value, so this amplification does not occur.

This example illustrates a limitation of reasoning from the relaxed feasible set alone. It does not establish that the single-scale algorithm generates this history: the prescribed queries need not be its midpoint prices, and a narrow feasible set need not have a narrow enclosing ellipsoid. Accordingly, the example explains the obstacle to extending the one-dimensional exploitation argument, without proving a regret lower bound for every single-scale tuning.

This example motivates maintaining several relaxation scales, as we explain after introducing the multiscale algorithm.

Algorithm for general environment: multi-scale shallow pricing. Our algorithm retains the two restart mechanisms from the identical-context setting: periodic restarts controlled by a window-size parameter a and inconsistency restarts triggered by rejected exploitation prices. We again call the consecutive rounds between restarts a *batch*. Within each batch, the algorithm combines shallow-pricing exploration with lower-envelope pricing (LEP), using multiple relaxation scales to account for the unknown amount of parameter variation.

The algorithm is parameterized by a *base price-margin parameter* δ, which determines the smallest relaxation scale. We maintain the dyadic scale set $\Lambda = \{\delta, 2\delta, \ldots, 2^L\delta\}$, where L is the smallest nonnegative integer such that $2^L\delta \geq \sqrt{2}$, the diameter of the feature domain for $d \geq 2$. Each scale $\lambda \in \Lambda$ maintains a relaxed feasible set $S_t^{(\lambda)}$ and an enclosing ellipsoid $\mathcal{E}_t^{(\lambda)}$. At the beginning of each batch, all feasible sets and ellipsoids are reset to their initial states. Each scale represents a different allowance for parameter variation within the batch. A scale is *active* if its relaxed feasible set is nonempty; the largest scale always remains active. We initialize $S_0 = \{x \in \mathbb{R}^d : x \geq 0,\ \|x\| \leq 1\}$ and $\mathcal{E}_0 = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$.

For exploration, the algorithm checks the ellipsoid width in the current context direction at every active scale. If any active scale λ satisfies $\mathrm{width}(\mathcal{E}_t^{(\lambda)}, u_t) > 4d\lambda$, the algorithm selects the smallest such scale and posts the corresponding shallow-pricing price. The resulting feedback updates the relaxed feasible sets at all scales, using their respective margins, while only the ellipsoid at the selected scale receives a shallow-cut update. Since a new context may reveal uncertainty in a different direction, exploration and exploitation can be interleaved within a batch.

[6] The first two prices are prescribed to illustrate the geometry of the relaxed feedback constraints. We do not claim that they are generated by the shallow-pricing midpoint rule.

Algorithm 2 Multiscale shallow pricing with LEP

Input: $T \in \mathbb{N}$, $d \in \mathbb{N}$ with $d \geq 2$, $V \in [0, T]$
Parameters: $\delta \leftarrow \max\{1/T, \sqrt{dV/T}\}$, $a \leftarrow \min\{T, \lceil \sqrt{dT/V} \rceil\}$ ▷ set $a \leftarrow T$ when $V = 0$
$\Lambda \leftarrow \{\delta, 2\delta, \ldots, 2^L\delta\}$, where L is the smallest nonnegative integer with $2^L\delta \geq \sqrt{2}$
$t \leftarrow 1$, $i \leftarrow 0$
while $t \leq T$ **do**
  $i \leftarrow i + 1$, $\tau \leftarrow \min\{T, t + a - 1\}$
  For all $\lambda \in \Lambda$, set $S_t^{(\lambda)} \leftarrow S_0$ and $\mathcal{E}_t^{(\lambda)} \leftarrow \mathcal{E}_0$
  while $t \leq \tau$ **do**
    Observe u_t; set $\Lambda_t \leftarrow \{\lambda \in \Lambda : S_t^{(\lambda)} \neq \varnothing\}$
    $\mathcal{X}_t \leftarrow \{\lambda \in \Lambda_t : \mathrm{width}(\mathcal{E}_t^{(\lambda)}, u_t) > 4d\lambda\}$
    if $\mathcal{X}_t \neq \varnothing$ **then**
      $\lambda_t \leftarrow \min \mathcal{X}_t$, $\widetilde{p}_t \leftarrow \left\langle c_t^{(\lambda_t)}, u_t \right\rangle$, $p_t \leftarrow \min\{1, \max\{0, \widetilde{p}_t\}\}$
    else
      $\lambda_t^{\min} \leftarrow \min \Lambda_t$, $p_t \leftarrow \mathrm{LEP}(S_t^{(\lambda_t^{\min})}, u_t)$
    end if
    Post p_t and observe f_t
    For all $\lambda \in \Lambda$, set $S_{t+1}^{(\lambda)} \leftarrow S_t^{(\lambda)} \cap H_t^{(\lambda)}$ and $\mathcal{E}_{t+1}^{(\lambda)} \leftarrow \mathcal{E}_t^{(\lambda)}$
    if $\mathcal{X}_t \neq \varnothing$ **then**
      $\widetilde{f}_t \leftarrow f_t \mathbb{I}\{\widetilde{p}_t \leq 1\}$
      $\mathcal{E}_{t+1}^{(\lambda_t)} \leftarrow \mathrm{ShallowUpdate}(\mathcal{E}_t^{(\lambda_t)}, u_t; \lambda_t, \widetilde{f}_t)$
    end if
    $t \leftarrow t + 1$
    if $\mathcal{X}_{t-1} = \varnothing$ and $f_{t-1} = 0$, **break** ▷ inconsistency restart
  end while
end while

For exploitation, the algorithm requires every active scale to satisfy its width condition. It then posts the lower-envelope price computed from the smallest active scale. This choice uses the nesting of the relaxed feasible sets: a smaller scale gives a smaller feasible set and therefore a weakly higher lower-envelope price. Meanwhile, a sufficiently large scale still contains the current hidden feature vector and certifies that the remaining uncertainty is small. Together, these properties control the regret on accepted exploitation rounds even when the smallest active set does not contain the current parameter.

As in the identical-context setting, the algorithm restarts after at most a rounds or immediately after an exploitation price is rejected. Periodic restarts limit the regret from changes that remain undetected because buyers continue accepting the posted prices. An inconsistency restart responds to overpricing and certifies more than δ parameter variation within the batch, allowing its frequency to be controlled by V. The window-size parameter a balances the cost of retaining outdated information against the exploration cost of frequent restarts, while the base margin δ balances pricing accuracy against exploration and inconsistency-restart costs. The formal algorithm is given in Algorithm 2.

Before stating the dynamic regret guarantee, we revisit Example 4.1 to explain how multiple scales address the illustrated difficulty.

Insert the first two observations from Example 4.1 into the relaxed feasible sets at every scale. At scale δ, the resulting set is narrow in direction $u_t = (1, 0)$ but excludes the current parameter $\theta_t = (1/2, 1/4 + \delta)$. At scale 4δ, the current parameter remains feasible because the within-batch parameter diameter is only 3δ. This larger-scale set also contains $(0, 1/4)$, so its width in direction u_t is at least $1/2$. For $\delta < 1/64$, this exceeds the exploration threshold $4d(4\delta) = 32\delta$. Any enclosing ellipsoid therefore also violates the width condition, and the multiscale rule would explore after this feedback history. The larger scale provides a reliable certificate of localization, while the smallest active scale determines the exploitation price. To see why this combination suffices, consider a batch B_i with parameter diameter $\Delta_i \triangleq \max_{s,t \in B_i} \|\theta_s - \theta_t\|$. Let λ_i be the smallest maintained scale satisfying $\lambda_i \geq \Delta_i$. Then $\lambda_i \leq 2\max\{\Delta_i, \delta\}$, and $\theta_t \in S_t^{(\lambda_i)}$ throughout the batch. Thus, scale λ_i remains active. Whenever the algorithm exploits, every

active scale, including λ_i, satisfies its ellipsoid width condition. Let $\lambda_t^{\min}$ denote the smallest active scale. Since $\lambda_t^{\min} \le \lambda_i$, nesting gives $S_t^{(\lambda_t^{\min})} \subseteq S_t^{(\lambda_i)}$. Consequently, the posted price satisfies $p_t = \mathrm{LEP}(S_t^{(\lambda_t^{\min})}, u_t) \ge \mathrm{LEP}(S_t^{(\lambda_i)}, u_t)$. On each accepted exploitation round, $v_t - p_t \le \mathrm{width}(S_t^{(\lambda_i)}, u_t) \le \mathrm{width}(\mathcal{E}_t^{(\lambda_i)}, u_t) \le 4d\lambda_i \le 8d\max\{\Delta_i, \delta\}$. Multiple scales therefore recover a regret bound controlled by the parameter diameter and the base margin, even when the smallest active set excludes the current parameter. A misplaced smaller-scale set may also cause overpricing. The resulting rejection immediately triggers an inconsistency restart, and these rejected rounds are accounted for separately in the regret analysis. The bound above applies to accepted exploitation rounds without requiring the smallest active set to contain the current parameter.

Regret guarantee and its analysis. We now state the dynamic regret guarantee of Algorithm 2.

Theorem 4.1. *In the contextual environment, given any non-stationarity upper bound $V \in [0, T]$, set $a = \min\{T, \lceil \sqrt{dT/V} \rceil\}$ and $\delta = \max\{1/T, \sqrt{dV/T}\}$ with the convention $a = T$ when $V = 0$. Then the dynamic regret of Algorithm 2 is at most $O((d^2 + d^{3/2}\sqrt{TV})\log^2 T)$.*

Moreover, for every $V \in [0, T]$ and every deterministic online algorithm, there exists an admissible instance with total variation at most V on which the algorithm incurs dynamic regret of at least $\Omega(\sqrt{TV} + \log\log T)$.

Combining the upper and lower bounds in Theorem 4.1, the regret guarantee of Algorithm 2 is optimal among deterministic online algorithms, up to logarithmic factors in T, for every $V \in [0, T]$ when d is fixed. The term $O(d^2 \log^2 T)$ accounts for the initial exploration cost and ensures that the guarantee also covers the stationary case $V = 0$. The remaining term scales as $\sqrt{TV}$, matching the variation-dependent lower bound up to logarithmic and dimension-dependent factors. We do not claim that the dependence on d is optimal.

The lower bound follows directly from Theorem 3.1, since the identical-context environment is a special case of the contextual environment. Specifically, any scalar value sequence $\{v_t\}_{t\in[T]}$ can be embedded by setting $u_t = (1, 0, \ldots, 0)$ and $\theta_t = (v_t, 0, \ldots, 0)$. For the upper bound, it suffices to analyze $V \le T/d$. When $V > T/d$, the stated bound follows from the trivial regret bound T, since $d^{3/2}\sqrt{TV} \ge T$. We next sketch the analysis for the nontrivial regime; the full proof is deferred to the Section E.

Consider the regime $V \le T/d$. Let $B_1, \ldots, B_m$ denote the batches, and define their parameter diameters by $\Delta_i \triangleq \max_{s,t\in B_i} \|\theta_s - \theta_t\|$. Since the batches are disjoint intervals, $\sum_{i=1}^m \Delta_i \le V$. We bound the regret from exploration and exploitation separately.

We first bound the number of batches. Periodic restarts contribute at most T/a completed batches. Every rejected exploitation round certifies more than δ parameter variation within its batch: otherwise, the current parameter would remain feasible at scale δ, and the smallest-scale LEP could not exceed the current value. Thus, there are at most V/δ inconsistency restarts, giving $m \le 1 + T/a + V/\delta = O(1 + \sqrt{TV/d})$.

For exploration, the shallow-pricing exploration bound gives at most $O(d^2 \log T)$ exploration rounds per scale in each batch. Since $\delta \ge 1/T$, the number of maintained scales is $O(\log T)$. Each exploration round incurs at most one unit of regret, so the total exploration regret is at most $O(md^2 \log^2 T) = O((d^2 + d^{3/2}\sqrt{TV})\log^2 T)$.

For exploitation, let λ_i be the smallest maintained scale satisfying $\lambda_i \ge \Delta_i$. As established above, this scale remains active throughout batch B_i, and each accepted exploitation round incurs regret of at most $4d\lambda_i \le 8d\max\{\Delta_i, \delta\}$. Using $|B_i| \le a$ and $\sum_{i=1}^m \Delta_i \le V$, the total regret from accepted exploitation rounds is at most $8dT\delta + 8d\sum_{i=1}^m |B_i|\Delta_i \le 8dT\delta + 8daV = O(d + d^{3/2}\sqrt{TV})$. Each rejected exploitation round incurs at most one unit of regret, and there are at most V/δ such rounds. Their total regret is therefore at most $V/\delta \le \sqrt{TV/d}$.

Combining these bounds yields $\textsc{Regret}[\textsc{Alg}] = O((d^2 + d^{3/2}\sqrt{TV})\log^2 T)$, as claimed.

AI USE STATEMENT

Generative AI tools were used as an assistive tool during the development of this work, including for improving exposition and organization, checking notation and internal consistency, discussing proof presentation and examples, and assisting with literature exploration. All research ideas, mathematical claims, proofs, and cited references included in the final manuscript were reviewed and verified by the authors. The authors take full responsibility for the final content of the paper.

REFERENCES

Adobe. Adobe: Holiday shopping season drove a record $257.8 billion online with consumers embracing generative AI tools, January 2026. URL https://news.adobe.com/news/2026/01/adobe-holiday-shopping-season.

Dheeraj Baby and Yu-Xiang Wang. Optimal dynamic regret in exp-concave online learning. In *Proceedings of the 34th Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pp. 359–409, 2021. URL https://proceedings.mlr.press/v134/baby21a.html.

Dheeraj Baby, Jianyu Xu, and Yu-Xiang Wang. Non-stationary contextual pricing with safety constraints. *Transactions on Machine Learning Research*, 2023. URL https://openreview.net/forum?id=fWIQ9Oaao0.

Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420, 2009. doi: 10.1287/opre.1080.0640.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Stochastic multi-armed-bandit problem with non-stationary rewards. In *Advances in Neural Information Processing Systems*, volume 27, 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/hash/91ba7292e5388b90b58d0b839a7f19ec-Abstract.html.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, 2015. doi: 10.1287/opre.2015.1408.

Yifang Chen, Chung-Wei Lee, Haipeng Luo, and Chen-Yu Wei. A new algorithm for non-stationary contextual bandits: Efficient, optimal and parameter-free. In *Proceedings of the 32nd Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 696–726, 2019. URL https://proceedings.mlr.press/v99/chen19b.html.

Wang Chi Cheung, David Simchi-Levi, and Ruihao Zhu. Learning to optimize under non-stationarity. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1079–1087, 2019. URL https://proceedings.mlr.press/v89/cheung19b.html.

Maxime C. Cohen, Ilan Lobel, and Renato Paes Leme. Feature-based dynamic pricing. *Management Science*, 66(11):4921–4943, 2020. doi: 10.1287/mnsc.2019.3485.

Adel Javanmard. Perishability of data: Dynamic pricing under varying-coefficient models. *Journal of Machine Learning Research*, 18(53):1–31, 2017. URL https://jmlr.org/papers/v18/17-061.html.

Adel Javanmard and Hamid Nazerzadeh. Dynamic pricing in high-dimensions. *Journal of Machine Learning Research*, 20(9):1–49, 2019. URL https://jmlr.org/papers/v20/17-357.html.

Feiyu Jiang and Zifeng Zhao. On non-stationary dynamic pricing: Adaptivity and optimality, 2026. URL https://arxiv.org/abs/2607.24115. arXiv preprint arXiv:2607.24115.

N. Bora Keskin and Assaf Zeevi. Chasing demand: Learning and earning in a changing environment. *Mathematics of Operations Research*, 42(2):277–307, 2017. doi: 10.1287/moor.2016.0807.

Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pp. 594–605, 2003. doi: 10.1109/SFCS.2003.1238232.

Akshay Krishnamurthy, Thodoris Lykouris, Chara Podimata, and Robert E. Schapire. Contextual search in the presence of adversarial corruptions. *Operations Research*, 71(4):1120–1135, 2023. doi: 10.1287/opre.2022.2365.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

Allen Liu, Renato Paes Leme, and Jon Schneider. Optimal contextual pricing and extensions. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms*, pp. 1059–1078, 2021. doi: 10.1137/1.9781611976465.66.

Ilan Lobel, Renato Paes Leme, and Adrian Vladu. Multidimensional binary search for contextual decision-making. *Operations Research*, 66(5):1346–1361, 2018. doi: 10.1287/opre.2018.1722.

Yiyun Luo, Will Wei Sun, and Yufeng Liu. Contextual dynamic pricing with unknown noise: Explore-then-UCB strategy and improved regrets. In *Advances in Neural Information Processing Systems*, volume 35, pp. 37445–37457, 2022. doi: 10.52202/068431-2714. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/f38d1fd25c15a0ad9ba758de4e7b1819-Abstract-Conference.html.

Renato Paes Leme and Jon Schneider. Contextual search via intrinsic volumes. In *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science*, pp. 268–282, 2018. doi: 10.1109/FOCS.2018.00034.

Renato Paes Leme, Balasubramanian Sivan, Yifeng Teng, and Pratik Worah. Learning to price against a moving target. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 6223–6232, 2021. URL https://proceedings.mlr.press/v139/leme21a.html.

Renato Paes Leme, Chara Podimata, and Jon Schneider. Corruption-robust contextual search through density updates. In *Proceedings of the 35th Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pp. 3504–3505, 2022. URL https://proceedings.mlr.press/v178/leme22a.html.

Jianyu Xu and Yu-Xiang Wang. Logarithmic regret in feature-based dynamic pricing. In *Advances in Neural Information Processing Systems*, volume 34, pp. 13898–13910, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/742141ceda6b8f6786609d31c8ef129f-Abstract.html.

Jianyu Xu and Yu-Xiang Wang. Towards agnostic feature-based dynamic pricing: Linear policies vs linear valuation with unknown noise. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 9643–9662, 2022. URL https://proceedings.mlr.press/v151/xu22d.html.

Jianyu Xu and Yu-Xiang Wang. Optimal contextual pricing under agnostic non-Lipschitz demand, 2026. URL https://arxiv.org/abs/2605.05609. arXiv preprint arXiv:2605.05609.

Zifeng Zhao, Feiyu Jiang, Yi Yu, and Xi Chen. High-dimensional dynamic pricing under nonstationarity: Learning and earning with change-point detection. *Management Science*, 2026. doi: 10.1287/mnsc.2023.00889. Articles in Advance.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning*. AAAI Press, 2003. URL https://cdn.aaai.org/ICML/2003/ICML03-120.pdf.

A Further Related Work

We discuss additional connections to online pricing, contextual search, and learning in non-stationary environments. The main distinctions concern the valuation model, the measure of non-stationarity, the information available to the seller, and the revenue benchmark.

Online posted pricing and demand learning. Kleinberg & Leighton (2003) study online posted-price auctions under identical, random, and worst-case valuations. For an unknown identical value, they establish the optimal $\Theta(\log\log T)$ regret. In this special case, the best fixed price and the first-best benchmark coincide. For arbitrary valuation sequences, their comparison is instead with the best fixed price in hindsight. A complementary operations research literature studies the exploration–exploitation tradeoff in learning an unknown demand function. For example, Besbes & Zeevi (2009) develop pricing policies and regret bounds relative to a seller who knows the demand function, in a setting with finite inventory. Our model has no intertemporal inventory constraint and compares revenue with the sum of buyers' realized values. Thus, differences in regret rates across these settings reflect differences in both information and benchmarks.

Stationary contextual pricing and geometric search. Cohen et al. (2020) study feature-based pricing with an unknown linear valuation model and develop ellipsoid-based algorithms with regret polynomial in d and logarithmic in T. They also introduce the robust ShallowPricing procedure, which relaxes feedback constraints to accommodate small valuation noise. Our algorithms use this geometric primitive. Related geometric methods appear in multidimensional binary search (Lobel et al., 2018) and contextual search via intrinsic volumes (Paes Leme & Schneider, 2018). Liu et al. (2021) obtain an $O(d\log\log T + d\log d)$ regret bound for stationary contextual pricing under their normalization and develop extensions to more general contextual search problems. These works exploit a fixed underlying parameter or hypothesis. In our setting, feedback constraints collected at different times may correspond to different hidden parameters. Consequently, shrinking an uncertainty set need not localize the current parameter. Our contribution is to control this loss of consistency while retaining the exploration guarantees of geometric search. Multiscale methods themselves also appear in prior contextual search algorithms, including Liu et al. (2021); here, the scales represent different tolerances for parameter drift.

Contextual pricing with stochastic valuation noise. A substantial literature assumes that valuations consist of a linear contextual component and an additional random shock. Javanmard & Nazerzadeh (2019) study high-dimensional pricing with sparse parameters and develop regularized likelihood methods. Xu & Wang (2021) establish logarithmic regret guarantees with a known noise distribution, including for adversarial contexts. When the noise distribution is unknown, Xu & Wang (2022) distinguish learning a linear pricing policy from learning a linear valuation model with agnostic noise, while Luo et al. (2022) combine parameter exploration with a UCB procedure for learning the noise distribution and maximizing revenue. More recently, Xu & Wang (2026) study unknown noise distributions that induce discontinuous demand, allowing jumps and atoms under stochastic, well-conditioned contexts. These models introduce statistical uncertainty beyond the unknown linear parameter. Their distribution-aware oracle generally knows the demand model but does not observe each buyer's realized noise before pricing. Our noiseless model instead permits an oracle that knows the current parameter and context to extract the buyer's full value. Accordingly, the associated regret guarantees address different benchmarks and cannot be compared solely by their dependence on T.

Non-contextual pricing against a moving value. The closest predecessor to our identical-context setting is Paes Leme et al. (2021), who study changing scalar values with binary purchase feedback and the same first-best benchmark. They analyze adversarial evolution under pointwise bounds on consecutive value changes, as well as stochastic evolution. Their adversarial revenue guarantees cover a fixed changing rate and unknown, non-increasing sequences of changing-rate bounds. The monotonicity condition concerns these bounds, not the values themselves. Their algorithms combine localization with randomized checking rounds to detect discrepancies between the current value and the maintained interval. Our formulation assumes a known aggregate variation budget V and permits this variation to be distributed arbitrarily over time. Our identical-context algorithm uses periodic restarts and deterministic inconsistency restarts, together with a price margin that controls their

frequency. It achieves $O((\sqrt{TV}+1)\log T)$ regret and supplies the restart mechanism used in our contextual algorithm. The two formulations also differ in prior information: knowing the aggregate budget is an assumption of our algorithm, whereas their unknown-rate algorithms adapt without this information.

Non-stationary pricing with stochastic demand. Non-stationary demand learning has also been studied beyond the moving-value model. Keskin & Zeevi (2017) analyze pricing with an unknown changing demand model and characterize regret under variation constraints. In the contextual setting, Javanmard (2017) study a linear valuation model with random shocks and time-varying coefficients, using projected stochastic gradient methods. Their guarantees cover both adversarial and stochastic context sequences and explicitly account for changes in the coefficients. Baby et al. (2023) study non-stationary contextual pricing with safety constraints. With known strictly log-concave market noise, they derive dynamic regret guarantees through proper online learning with exp-concave losses and adapt to an unknown parameter variation budget. Their benchmark is the sequence of optimal expected-revenue prices under the current demand model. These papers already address contextual heterogeneity together with non-stationarity. Our focus is the noiseless linear-valuation model, an aggregate Euclidean variation budget, and regret against full surplus, without assuming the smooth demand probabilities used in their statistical analyses.

Change detection provides another approach to non-stationary pricing. Zhao et al. (2026) develop a policy based on online change-point detection for high-dimensional generalized linear demand models, with guarantees that account for sparsity and the number of changes. Jiang & Zhao (2026) develop multiscale change-point detection that adapts to both piecewise stationarity and a design-adjusted parameter variation budget, without knowing these measures in advance. Their model uses stochastic contexts and generalized linear demand, and their oracle maximizes expected revenue. Their scales organize statistical detection over time. In our algorithm, scales index simultaneous relaxations of the same geometric feedback constraints and determine whether localization is reliable enough for exploitation. This distinction also clarifies that our contribution is not the general use of restarts or multiple scales in non-stationary pricing.

Non-stationary bandits and restart methods. Besbes et al. (2014) study stochastic multi-armed bandits under a variation budget on mean rewards and establish near-optimal dynamic regret through periodic restarts of Exp3. Cheung et al. (2019) develop sliding-window methods and a bandit-over-bandit approach for non-stationary linear bandits. Chen et al. (2019) give an adaptive contextual bandit algorithm that uses replay phases to detect changes without prior knowledge of the amount of non-stationarity. These works motivate forgetting outdated observations and combining exploration with change detection. However, their variation measures and reward structures differ from ours. Even with a fixed context, an arbitrarily small decrease in the hidden parameter can change a purchase into a rejection at a fixed positive price, causing a constant revenue drop. Thus, a small parameter variation budget does not imply a comparably small uniform variation budget for revenue functions. Generic reward-variation guarantees therefore do not directly yield our regret bound. Our analysis instead uses the geometry of purchase feedback and the asymmetric consequences of underpricing and overpricing.

Dynamic regret and online convex optimization. Dynamic regret against a changing comparator is a classical objective in online optimization. Zinkevich (2003) analyze online convex optimization with moving comparators, and Besbes et al. (2015) study stochastic optimization under variation budgets on the objective functions. Baby & Wang (2021) establish optimal dynamic regret guarantees for exp-concave online learning when improper predictions are allowed. The pricing results of Baby et al. (2023) connect this literature to constrained estimation and stochastic demand. Our revenue loss is discontinuous and is not a convex function of the posted price, so these guarantees do not apply directly. Nevertheless, the distinction between comparator movement, objective variation, and parameter variation is relevant here: our budget constrains hidden features, while contexts may vary arbitrarily. In particular, even when $V = 0$, both buyer values and first-best prices may change across rounds.

Corruption-robust contextual search. Krishnamurthy et al. (2023) study contextual search when some responses are adversarially corrupted and design algorithms whose performance degrades with the number of corruptions. Paes Leme et al. (2022) develop density-based methods for adversari-

ally noisy contextual search, obtaining guarantees for symmetric and tolerance-based losses. These works share the concern that intersecting all observed constraints can eliminate the correct hypothesis. The source of inconsistency differs in our model: feedback is accurate for the current buyer, but the hidden parameter itself moves. A single persistent parameter shift may make many subsequent observations inconsistent with an earlier parameter while contributing only once to our variation budget. Moreover, a small symmetric prediction error does not ensure small pricing regret, because even slight overpricing can eliminate the entire revenue from a round. Thus, corruption budgets and search-loss guarantees require additional arguments before they can be used for our objective.

Position of our contribution. Our work combines geometric contextual pricing with variation-controlled adaptation under a first-best benchmark. The main technical issue is that a narrow relaxed feasible set may exclude the current hidden parameter without producing an immediate rejection. Maintaining several relaxation scales provides a scale that remains consistent with the parameters within a batch, while nesting relates its localization guarantee to the price computed from the smallest active scale. Periodic and inconsistency restarts then control the effects of accumulated drift. This yields the optimal dependence on T and V, up to logarithmic factors for fixed d among deterministic algorithms, under the assumption that the seller knows V.

B Technical Background: Geometric Search and Shallow Pricing

B.1 Background: geometric search in feature-based pricing

Our contextual algorithms build on the geometric viewpoint introduced in feature-based dynamic pricing by Cohen et al. (2020). In that model, each arriving product is described by an observed feature vector, the buyer's valuation is linear in an unknown parameter, and the seller observes only binary accept/reject feedback after posting a price. Thus, each observation induces a halfspace constraint on the unknown parameter. Pricing can therefore be viewed as a geometric search problem: the seller maintains a region of parameters that remain consistent with past feedback, and each new context direction asks for the range of possible valuations induced by this region.

For a convex set $K \subseteq \mathbb{R}^d$ and a context direction u, define the directional width of K along u as

$$\mathrm{width}(K, u) := \max_{x \in K} \langle x, u \rangle - \min_{x \in K} \langle x, u \rangle .$$

This is the length of the interval of valuations that are still possible in context u when the hidden parameter is only known to lie in K. Equivalently, if

$$b^-(K, u) := \min_{x \in K} \langle x, u \rangle , \qquad b^+(K, u) := \max_{x \in K} \langle x, u \rangle ,$$

then

$$\mathrm{width}(K, u) = b^+(K, u) - b^-(K, u).$$

The algorithmic primitive in Cohen et al. (2020) is EllipsoidPricing. Instead of working directly with a possibly complicated polyhedral uncertainty set, the algorithm maintains an ellipsoidal outer approximation

$$\mathcal{E}_t = \mathcal{E}(A_t, c_t) := \left\{ x \in \mathbb{R}^d : (x - c_t)^\top A_t^{-1} (x - c_t) \le 1 \right\} ,$$

where c_t is the center and A_t is a positive definite matrix. When this primitive is run at scale λ in our algorithms, we write these quantities as $E_t^{(\lambda)}$, $c_t^{(\lambda)}$, and $A_t^{(\lambda)}$. For a context u_t, the minimum and maximum possible valuations over this ellipsoid have the closed forms

$$b_t^- = \langle c_t, u_t \rangle - \sqrt{u_t^\top A_t u_t}, \qquad b_t^+ = \langle c_t, u_t \rangle + \sqrt{u_t^\top A_t u_t}.$$

Thus

$$\mathrm{width}(\mathcal{E}_t, u_t) = b_t^+ - b_t^- = 2\sqrt{u_t^\top A_t u_t}.$$

Given a target accuracy $\eta > 0$, EllipsoidPricing behaves as follows. If

$$b_t^+ - b_t^- \le \eta,$$

the algorithm exploits by posting the lower endpoint

$$p_t = b_t^-.$$

If instead

$$b_t^+ - b_t^- > \eta,$$

the algorithm explores by posting the midpoint price

$$p_t = \frac{b_t^- + b_t^+}{2} = \langle c_t, u_t \rangle .$$

An acceptance, $f_t = 1$, implies the central halfspace constraint

$$\langle x, u_t \rangle \geq p_t,$$

while a rejection, $f_t = 0$, implies

$$\langle x, u_t \rangle \leq p_t.$$

The next ellipsoid is chosen as the minimum-volume ellipsoid containing the corresponding half-ellipsoid.

More explicitly, define

$$q_t := \frac{A_t u_t}{\sqrt{u_t^\top A_t u_t}}.$$

For an acceptance, the central-cut update is

$$c_{t+1} = c_t + \frac{1}{d+1} q_t,$$

and for a rejection it is

$$c_{t+1} = c_t - \frac{1}{d+1} q_t.$$

In both cases,

$$A_{t+1} = \frac{d^2}{d^2-1} \left(A_t - \frac{2}{d+1} q_t q_t^\top \right).$$

The key fact behind the exploration bound is that each central cut decreases the ellipsoid volume by a constant factor, while exploration occurs only when the current direction has large ellipsoidal width.

The robust primitive we use is SHALLOWPRICING, also from Cohen et al. (2020). It modifies the central-cut update by adding a safety buffer to the inferred halfspace. At scale $\lambda > 0$, let

$$\widetilde{p}_t := \frac{b_t^- + b_t^+}{2} = \langle c_t, u_t \rangle$$

denote the standard, unclipped midpoint proposal. Under the standard SHALLOWPRICING update, an acceptance produces the shifted constraint

$$\langle x, u_t \rangle \geq \widetilde{p}_t - \lambda,$$

and a rejection produces the shifted constraint

$$\langle x, u_t \rangle \leq \widetilde{p}_t + \lambda.$$

Thus the update no longer cuts the ellipsoid through its center; it is a shallow cut.

The shallow-cut update can also be written in closed form. Let

$$\alpha_t := -\frac{\lambda}{\sqrt{u_t^\top A_t u_t}}.$$

The quantity α_t is the signed depth of the cut. When the algorithm explores only if

$$\mathrm{width}(\mathcal{E}_t, u_t) > 4d\lambda,$$

we have

$$\sqrt{u_t^\top A_t u_t} > 2d\lambda,$$

and hence

$$\alpha_t > -\frac{1}{2d}.$$

This is the regime in which the shallow-cut ellipsoid update is valid.

For an acceptance, the shallow-cut center update is

$$c_{t+1} = c_t + \frac{1 + d\alpha_t}{d+1} q_t,$$

whereas for a rejection,

$$c_{t+1} = c_t - \frac{1 + d\alpha_t}{d+1} q_t.$$

In both cases, the shape matrix is updated as

$$A_{t+1} = \frac{d^2}{d^2-1}(1-\alpha_t^2)\left(A_t - \frac{2(1+d\alpha_t)}{(d+1)(1+\alpha_t)} q_t q_t^\top\right).$$

Clipping the exploration price to $[0,1]$. Our algorithms post

$$p_t := \min\{1, \max\{0, \widetilde{p}_t\}\}.$$

Let

$$\widetilde{f}_t := f_t \, \mathbf{1}\{\widetilde{p}_t \le 1\}$$

denote the inferred feedback at the unclipped midpoint. We denote the corresponding ellipsoid update by

$$E_{t+1} = \mathrm{ShallowUpdate}(E_t, u_t; \lambda, \widetilde{f}_t).$$

The relaxed feasible set is updated using the actual posted price p_t and observed feedback f_t, while the ellipsoid is updated using the standard shallow cut associated with $\widetilde{p}_t$ and $\widetilde{f}_t$.

If $0 \le \widetilde{p}_t \le 1$, then $p_t = \widetilde{p}_t$ and $\widetilde{f}_t = f_t$, so the standard shallow-pricing update applies unchanged.

Suppose instead that $\widetilde{p}_t < 0$. Since realized valuations are nonnegative, $p_t = 0$ is necessarily accepted, and the actual λ-relaxed feedback constraint is

$$\langle x, u_t \rangle \ge -\lambda.$$

Since $\widetilde{p}_t < 0$,

$$\{x : \langle x, u_t \rangle \ge -\lambda\} \subseteq \{x : \langle x, u_t \rangle \ge \widetilde{p}_t - \lambda\}.$$

Thus the actual relaxed constraint is stronger than the standard acceptance constraint associated with $\widetilde{p}_t$.

Finally, suppose that $\widetilde{p}_t > 1$. Then $\widetilde{f}_t = 0$. Since every $x \in S_0$ satisfies $\langle x, u_t \rangle \le 1$, we have

$$\langle x, u_t \rangle \le 1 < \widetilde{p}_t + \lambda.$$

Hence the updated relaxed feasible set is contained in the standard rejection halfspace

$$\{x : \langle x, u_t \rangle \le \widetilde{p}_t + \lambda\},$$

regardless of the feedback observed at the clipped price $p_t = 1$.

Therefore, in all cases, the standard shallow cut associated with $(\widetilde{p}_t, \widetilde{f}_t)$ contains the updated relaxed feasible set. Thus clipping the posted price to $[0,1]$ preserves the ellipsoidal outer-approximation invariant and the usual shallow-cut exploration guarantee.

In our non-stationary setting, the relaxation parameter λ has a different interpretation from statistical noise: it is a robustness margin that allows past feedback constraints to remain compatible with a parameter that may have moved within the current batch. The single-scale baseline uses one fixed radius $\lambda = \delta$, while our final algorithm maintains a dyadic hierarchy of such radii. The SHALLOW-PRICING primitive itself is not our contribution; our use of it together with scale-dependent relaxed feasible sets and LEP pricing is described in Section 4.

B.2 Relaxed feasible sets versus ellipsoidal outer approximations

In our non-stationary algorithms, the relaxed feasible set and its ellipsoidal outer approximation have distinct roles. At each relaxation scale λ we maintain

$$S_t^{(\lambda)} \subseteq \mathcal{E}_t^{(\lambda)}.$$

The relaxed feasible set is the intersection of the shifted feedback halfspaces and is therefore the object used for LEP pricing. The ellipsoid is used only to decide whether the current context direction is sufficiently uncertain to require exploration and, on such a round, to perform the shallow-cut update. In particular,

$$\mathrm{width}(S_t^{(\lambda)}, u_t) \le \mathrm{width}(\mathcal{E}_t^{(\lambda)}, u_t).$$

On a multiscale exploration round at scale λ_t, only $\mathcal{E}_t^{(\lambda_t)}$ is shallow-updated. This is because the ellipsoid update uses $\mathrm{ShallowUpdate}$ with the unclipped midpoint $\widetilde{p}_t$ and its inferred feedback $\widetilde{f}_t$. Since $\widetilde{p}_t$ is centered with respect to $E_t^{(\lambda_t)}$, the shallow-cut depth guarantee above applies directly at that scale. For every other scale λ, leaving the ellipsoid unchanged preserves the outer-approximation invariant because

$$S_{t+1}^{(\lambda)} \subseteq S_t^{(\lambda)} \subseteq \mathcal{E}_t^{(\lambda)} = \mathcal{E}_{t+1}^{(\lambda)}.$$

Throughout the remainder of the analysis, let

$$R := \sqrt{2},$$

which upper-bounds the diameter of the initial feasible set S_0.

Lemma B.1 (Shallow-cut exploration bound)**.** *Fix a relaxation scale $\lambda > 0$ and an exploration threshold η satisfying*

$$4d\lambda \le \eta < 20R(d+1).$$

If an ellipsoid is updated only on rounds for which

$$\mathrm{width}(E_t, u_t) > \eta,$$

using the standard λ-shifted shallow cut associated with the unclipped midpoint proposal, then within a fixed batch the number of such updates is at most

$$20d^2 \ln\left(\frac{20R(d+1)}{\eta}\right).$$

Proof. Inspecting the proof of Theorem 3 of Cohen et al. (2020), the exploration-count argument uses $\eta \ge 4d\lambda$ to ensure that the shallow-cut depth satisfies $\alpha_t \ge -1/(2d)$, and then gives

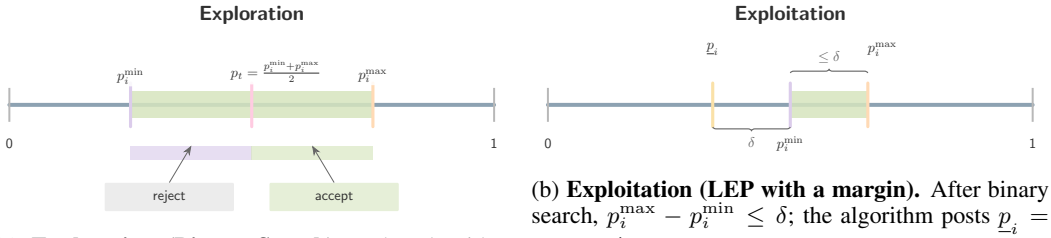
$$N \le 20d^2 \ln\left(\frac{20R(d+1)}{\eta}\right).$$

The clipping argument above ensures that the ellipsoid update used here is the same valid shallow update at the unclipped midpoint. □

C Identical Context Upper Bound

C.1 Complete proof of Theorem 3.1

Illustration. Figure 1 summarizes the two phases of Algorithm 1. During exploration, binary search progressively narrows $[p_i^{\min}, p_i^{\max}]$ according to the observed accept/reject feedback. The final endpoints retain accepted/rejected witness rounds that are used in the regret analysis below. Once its width is at most δ, the algorithm switches to LEP with a margin and repeatedly posts $\underline{p}_i = \max\{p_i^{\min} - \delta, 0\}$. The margin makes a rejection informative: if this LEP price is rejected, the valuation must have decreased sufficiently relative to an earlier exploration round, and the resulting reset can be charged to the total variation budget. The hard window-size reset, in contrast, limits the duration of undetectable upward movements. These are precisely the two ingredients used in the regret analysis below.



(a) **Exploration (Binary Search).** The algorithm queries $p_t \in [p_i^{\min}, p_i^{\max}]$ and shrinks $[p_i^{\min}, p_i^{\max}]$ based on accept/reject feedback.

(b) **Exploitation (LEP with a margin).** After binary search, $p_i^{\max} - p_i^{\min} \leq \delta$; the algorithm posts $\underline{p}_i = \max\{p_i^{\min} - \delta, 0\}$ until either an inconsistency triggers a reset or the window-size cap is reached.

Figure 1: Illustration of Algorithm 1 in the identical-context setting. Exploration tightens $[p_i^{\min}, p_i^{\max}]$, after which exploitation repeatedly posts the LEP price with margin $\underline{p}_i$. A rejection triggers an inconsistency reset, whereas the window-size cap limits how long the same price can persist without such a signal.

Proof. Let

$$B_1 = \{1, \ldots, t_1\}, \quad B_2 = \{t_1 + 1, \ldots, t_2\}, \quad \ldots, \quad B_r = \{t_{r-1} + 1, \ldots, T\}$$

denote the batches generated by the algorithm, i.e., consecutive rounds between resets. Let R_i denote the regret incurred in batch B_i, and let $E_i \subseteq B_i$ denote its exploration rounds. By construction,

$$|B_i| \leq a.$$

Also define the variation diameter of batch B_i by

$$\Delta_i := \max_{j,k \in B_i} |v_j - v_k|.$$

For every exploitation round $j \in B_i \setminus E_i$, the algorithm posts the same LEP price with margin,

$$p_j = \underline{p}_i = \max\{p_i^{\min} - \delta, 0\}.$$

For any batch that reaches exploitation, let

$$L_i := p_i^{\min}, \qquad U_i := p_i^{\max}$$

denote the final binary-search interval, so that

$$U_i - L_i \leq \delta.$$

We first bound the regret on an accepted LEP round $j \in B_i \setminus E_i$. Recall that

$$p_j = \underline{p}_i = \max\{L_i - \delta, 0\}.$$

If $U_i < 1$, then the final upper endpoint must have been created by a rejected binary-search query. Hence there exists an exploration round $t_1 \in E_i$ such that

$$v_{t_1} < U_i.$$

Therefore, by the definition of Δ_i,

$$v_j \leq v_{t_1} + \Delta_i < U_i + \Delta_i \leq L_i + \delta + \Delta_i.$$

Consequently,

$$v_j - p_i \leq \Delta_i + 2\delta.$$

Indeed, if $L_i \geq \delta$, then $p_i = L_i - \delta$ and the claim follows directly; while if $L_i < \delta$, then $p_i = 0$ and

$$v_j < L_i + \delta + \Delta_i < \Delta_i + 2\delta.$$

If instead $U_i = 1$, then $U_i - L_i \leq \delta$ implies $L_i \geq 1 - \delta$. Hence

$$p_i = \max\{L_i - \delta, 0\} \geq 1 - 2\delta,$$

and since $v_j \le 1$,

$$v_j - p_i \le 2\delta \le \Delta_i + 2\delta.$$

Thus every accepted LEP round incurs regret at most $\Delta_i + 2\delta$.

The regret of batch B_i can be decomposed as

$$\begin{aligned} R_i &= \sum_{j \in B_i} \Big(v_j - \mathcal{I}\{p_j \le v_j\} p_j \Big) \\ &= \sum_{j \in E_i} \Big(v_j - \mathcal{I}\{p_j \le v_j\} p_j \Big) + \sum_{j \in B_i \setminus E_i} \Big(v_j - \mathcal{I}\{p_j \le v_j\} p_j \Big) \\ &\le |E_i| + |B_i|(\Delta_i + 2\delta) + 1 \\ &\le O(\log(T) + |B_i|\delta + a\Delta_i). \end{aligned}$$

Here we used that binary search requires at most $\log(T)$ exploration rounds, and that on every exploitation round before a reset the regret is at most $\Delta_i + 2\delta$.

Summing over all batches gives

$$\begin{aligned} \textsc{Regret}[\textsc{Alg}] = \sum_{i=1}^{r} R_i &\le \sum_{i=1}^{r} O\left(|B_i|\delta + a\Delta_i + \log(T) \right) \\ &= O(T\delta + a \sum_{i=1}^{r} \Delta_i + r \log(T)). \end{aligned}$$

Since the batches are disjoint intervals,

$$\sum_{i=1}^{r} \Delta_i \le \mathcal{V}(\mathbf{v}) \le V.$$

It therefore remains to bound the number of batches.

Lemma C.1. *The two reset mechanisms imply*

$$r \le \frac{V}{\delta} + \frac{T}{a} + 1.$$

Proof. Every batch B_i ends for one of the following reasons:

- the window-size cap is reached, i.e., $|B_i| = a$;
- $i = r$, so this is the final incomplete batch;
- an inconsistency occurs on an LEP round and triggers an immediate reset.

There are at most T/a batches of the first type and at most one batch of the second type. We now show that there are at most V/δ inconsistency resets.

Consider a batch B_i that ends because an LEP price is rejected at round t_i. Since the algorithm resets only after the binary-search phase, this means that LEP has become overpricing:

$$v_{t_i} < p_{t_i} = \underline{p}_i.$$

Moreover, an inconsistency can only occur when $\underline{p}_i > 0$, so

$$\underline{p}_i = p_i^{\min} - \delta.$$

Hence

$$v_{t_i} < p_i^{\min} - \delta.$$

During binary search, the lower endpoint $p_i^{\min}$ can increase only after an accepted query. Therefore, there exists an exploration round $j \in E_i$ at which a price equal to the resulting lower endpoint was accepted, so

$$p_i^{\min} \le v_j.$$

Consequently,

$$v_{t_i} + \delta < p_i^{\min} \le v_j,$$

and therefore

$$\delta < |v_j - v_{t_i}| \le \Delta_i.$$

In particular,

$$\delta < \sum_{k=t_{i-1}+1}^{t_i-1} |v_{k+1} - v_k|.$$

Thus every inconsistency reset can be charged to more than δ total variation inside its batch. Since these batches are disjoint and

$$\mathcal{V}(\mathbf{v}) \le V,$$

there can be at most V/δ such resets. Combining the three cases gives

$$r \le \frac{T}{a} + \frac{V}{\delta} + 1.$$

□

Using Lemma C.1, we obtain

$$\textsc{Regret}[\textsc{Alg}] \le O(T\delta + aV + \left(\frac{T}{a} + \frac{V}{\delta} + 1\right)\log(T)).$$

For notational simplicity, in the tuning calculation below we ignore the ceiling in the definition of a. This affects the bounds only by an absolute constant factor: for $x \ge 1$,

$$x \le \lceil x \rceil \le 2x,$$

and the additional truncation by T preserves this constant-factor relation. Thus replacing $\min\{T, \lceil\sqrt{T/V}\rceil\}$ by $\min\{T, \sqrt{T/V}\}$ does not affect the stated big-O regret bound. Setting

$$a = \min\{T, \sqrt{\frac{T}{V}}\}, \qquad \delta = \max\{\frac{1}{T}, \sqrt{\frac{V}{T}}\},$$

gives

$$\begin{aligned}\textsc{Regret}[\textsc{Alg}] &\le O(T\delta + aV + \left(\frac{T}{a} + \frac{V}{\delta}\right)\log(T) + \log(T)) \\ &= O(\sqrt{TV}\log(T) + \log(T)).\end{aligned}$$

for the case $V > \frac{1}{T}$ and when $V \le \frac{1}{T}$ we get

$$\begin{aligned}\textsc{Regret}[\textsc{Alg}] &\le O(T\delta + aV + \left(\frac{T}{a} + \frac{V}{\delta}\right)\log(T) + \log(T)) \\ &= O(1 + \log(T)) = O(\log(T)).\end{aligned}$$

This proves the claimed dynamic-regret bound. □

Remark C.1. *Alternatively for $V > 0$ choosing*

$$a = \min\left\{T, \sqrt{\frac{T\log T}{V}}\right\}, \qquad \delta = \max\left\{\frac{1}{T}, \sqrt{\frac{V\log T}{T}}\right\},$$

in the bound above yields the slightly sharper guarantee

$$\textsc{Regret}[\textsc{Alg}] = O\Big(\sqrt{TV\log T} + \log T\Big).$$

For $V = 0$, we take $a = T$ and $\delta = 1/T$.

Algorithm 3 Single-scale shallow pricing with LEP

Input: $T \in \mathbb{N}$, $d \in \mathbb{N}$ with $d \geq 2$, $V \in [0, T]$
Parameters: $\delta \leftarrow \max\{1/T, d^{1/2}(V/T)^{1/3}\}$, $a \leftarrow \min\{T, \lceil d^{1/2}(T/V)^{1/3}\rceil\}$, $\eta \leftarrow 4d\delta$ ▷ set $a \leftarrow T$ when $V = 0$
$t \leftarrow 1$, $i \leftarrow 0$
while $t \leq T$ **do**
  $i \leftarrow i + 1$, $\tau \leftarrow \min\{T, t + a - 1\}$
  Set $S_t \leftarrow S_0$, $S_t^+ \leftarrow S_0$, and $\mathcal{E}_t \leftarrow \mathcal{E}_0$
  while $t \leq \tau$ **do**
    if $S_t^+ = \varnothing$, **break** ▷ inconsistency restart
    Observe u_t
    if $\mathrm{width}(\mathcal{E}_t, u_t) > \eta$ **then**
      $\widetilde{p}_t \leftarrow \langle c_t, u_t \rangle$, $p_t \leftarrow \min\{1, \max\{0, \widetilde{p}_t\}\}$
    else
      $p_t \leftarrow \mathrm{LEP}(S_t^+, u_t)$
    end if
    Post p_t and observe f_t
    Construct H_t and H_t^+ according to p_t and f_t
    $S_{t+1} \leftarrow S_t \cap H_t$, $S_{t+1}^+ \leftarrow S_t^+ \cap H_t^+$, and $\mathcal{E}_{t+1} \leftarrow \mathcal{E}_t$
    if $\mathrm{width}(\mathcal{E}_t, u_t) > \eta$ **then**
      $\widetilde{f}_t \leftarrow f_t \mathbb{I}\{\widetilde{p}_t \leq 1\}$
      $\mathcal{E}_{t+1} \leftarrow \mathrm{ShallowUpdate}(\mathcal{E}_t, u_t; \delta, \widetilde{f}_t)$
    end if
    $t \leftarrow t + 1$
    if $\mathrm{width}(\mathcal{E}_{t-1}, u_{t-1}) \leq \eta$ and $f_{t-1} = 0$, **break** ▷ inconsistency restart
  end while
end while

D Single-Scale Contextual Baseline

This appendix gives the complete single-scale contextual algorithm and its analysis. As discussed in Section 4, this is the most direct geometric extension of the identical-context warm-up: it maintains a single relaxed feasible set at radius δ, uses shallow pricing with the midpoint proposal clipped to $[0, 1]$, and exploits using LEP.

Within a batch, let S_t denote the set of parameters consistent with the exact feedback observed so far. In parallel, let $S_t^+ = S_t^{(\delta)}$ denote the δ-relaxed feasible set. After posting p_t in context u_t, define

$$H_t = \begin{cases} \{x : \langle x, u_t \rangle \leq p_t\}, & f_t = 0, \\ \{x : \langle x, u_t \rangle \geq p_t\}, & f_t = 1, \end{cases} \qquad H_t^+ = \begin{cases} \{x : \langle x, u_t \rangle \leq p_t + \delta\}, & f_t = 0, \\ \{x : \langle x, u_t \rangle \geq p_t - \delta\}, & f_t = 1. \end{cases}$$

The feasible sets are updated according to

$$S_{t+1} = S_t \cap H_t, \qquad S_{t+1}^+ = S_t^+ \cap H_t^+.$$

The algorithm also maintains an ellipsoidal outer approximation

$$S_t^+ \subseteq \mathcal{E}_t.$$

As in the multiscale algorithm, the relaxed feasible set determines the LEP price, while the ellipsoid determines whether exploration is necessary.

The empty-set restart in Algorithm 3 is a technical form of inconsistency reset: if the relaxed feedback constraints become mutually infeasible, the current batch is discarded. As shown below, such a restart, like a rejected LEP round, can be charged to at least δ variation within the batch.

Theorem D.1 (Single-scale contextual regret)**.** *For $4d\delta \leq \eta < 20R(d+1)$, Algorithm 3 has dynamic regret*

$$\textsc{Regret}[\textsc{Alg}] \leq O(d^2 \log \frac{20R(d+1)}{\eta}) \left(\frac{T}{a} + \frac{V}{\delta} + 1 \right) + T\eta + a\frac{V}{\delta}.$$

with the convention $a = T$ when $V = 0$. And by choosing the parameters as

$$a = \min\{T, \lceil d^{1/2}\left(\frac{T}{V}\right)^{1/3}\rceil\}, \qquad \delta = \max\{\frac{1}{T}, d^{1/2}\left(\frac{V}{T}\right)^{1/3}\}, \qquad \eta = 4d\delta,$$

One can prove a regret bound of

$$\textsc{Regret}[\textsc{Alg}] \le O\Big((d^2 + d^{3/2}V^{1/3}T^{2/3})\log(T)\Big).$$

D.1 Shallow-pricing exploration bound

Lemma D.2 (Shallow-pricing exploration bound)**.** *Suppose $4d\delta \le \eta < 20R(d+1)$. Within any fixed batch, the number of exploration rounds in Algorithm 3 is at most*

$$O\left(d^2 \log \frac{20R(d+1)}{\eta}\right).$$

Proof. Consider an exploration round t. By Algorithm 3,

$$\mathrm{width}(\mathcal{E}_t, u_t) > \eta.$$

Writing

$$\mathcal{E}_t = E(A_t, c_t),$$

we have

$$\mathrm{width}(\mathcal{E}_t, u_t) = 2\sqrt{u_t^\top A_t u_t},$$

and therefore

$$\sqrt{u_t^\top A_t u_t} > \frac{\eta}{2}.$$

Let

$$\widetilde{p}_t := \langle c_t, u_t \rangle.$$

The posted exploration price is

$$p_t = \min\{1, \max\{0, \widetilde{p}_t\}\},$$

and the ellipsoid is updated using

$$\widetilde{f}_t = f_t \mathbb{I}\{\widetilde{p}_t \le 1\}.$$

By Section B, this clipping preserves the ellipsoidal outer-approximation invariant and the validity of the standard shallow update.

The normalized depth of this shallow cut satisfies

$$\frac{\delta}{\sqrt{u_t^\top A_t u_t}} < \frac{2\delta}{\eta} \le \frac{1}{2d}.$$

Thus every exploration round uses a valid shallow ellipsoid update, and the standard shallow-cut potential argument recalled in Section B gives the displayed bound. □

D.2 Reset count

Lemma D.3 (Reset count)**.** *The number of non-window resets in Algorithm 3 is at most V/δ. Consequently, the total number of batches satisfies*

$$r \le 1 + \frac{T}{a} + \frac{V}{\delta}.$$

Proof. First consider a rejected LEP round t. Rejection means

$$\langle \theta_t, u_t \rangle < p_t.$$

Since the realized valuation is nonnegative, this implies $p_t > 0$. Hence the clipping at zero is inactive and

$$p_t = \mathrm{LEP}(S_t^+, u_t) = \min_{x \in S_t^+} \langle x, u_t \rangle.$$

It follows that

$$\theta_t \notin S_t^+.$$

Since S_t^+ is the intersection of the δ-relaxed feedback halfspaces generated earlier in the same batch, there exists some previous round $j < t$ such that

$$\theta_t \notin H_j^+.$$

At round j, however, θ_j satisfied the corresponding exact halfspace H_j.

If $f_j = 0$, then

$$\langle \theta_j, u_j \rangle \le p_j, \qquad \langle \theta_t, u_j \rangle > p_j + \delta,$$

and hence, using $\|u_j\| = 1$,

$$\|\theta_t - \theta_j\| \ge \langle \theta_t - \theta_j, u_j \rangle > \delta.$$

If $f_j = 1$, then

$$\langle \theta_j, u_j \rangle \ge p_j, \qquad \langle \theta_t, u_j \rangle < p_j - \delta,$$

and similarly

$$\|\theta_t - \theta_j\| > \delta.$$

Thus every rejected-LEP reset certifies more than δ variation within the batch.

The same conclusion holds if a batch is restarted because $S_t^+ = \varnothing$. Suppose that $S_t^+ = \emptyset$ at the beginning of round t. If the variation diameter of the completed batch were at most δ, then for every earlier round j in that batch, θ_{t-1} would satisfy the δ-relaxed feedback halfspace generated at round j. It also satisfies the relaxed halfspace generated at round $t-1$. Hence $\theta_{t-1} \in S_t^+$, contradicting $S_t^+ = \emptyset$.

Since different reset batches are disjoint intervals and the total variation budget satisfies

$$\mathcal{V}(\Theta) \le V,$$

there can be at most V/δ non-window resets. The hard window cap contributes at most T/a additional batches, plus one final incomplete batch. □

D.3 Proof of Theorem D.1

Proof. Let B_i denote the i-th batch and define its variation diameter by

$$\Delta_i := \max_{j,k \in B_i} \|\theta_j - \theta_k\|.$$

We split the batches into *good* batches, for which $\Delta_i \le \delta$, and *bad* batches, for which $\Delta_i > \delta$.

First consider a good batch. For every round t in this batch and every previous round $j < t$ in the same batch, the exact feedback halfspace generated at j contains θ_j. Since

$$\|\theta_t - \theta_j\| \le \Delta_i \le \delta,$$

the current parameter θ_t satisfies the corresponding δ-relaxed halfspace. Therefore

$$\theta_t \in S_t^+$$

throughout the batch meaning that every LEP round in a good batch is accepted

$$p_t \le \langle \theta_t, u_t \rangle$$

Now consider an LEP exploitation round t in this good batch. Since the round is not exploratory,

$$\mathrm{width}(\mathcal{E}_t, u_t) \le \eta.$$

Because

$$S_t^+ \subseteq \mathcal{E}_t,$$

we have

$$\mathrm{width}(S_t^+, u_t) \le \mathrm{width}(\mathcal{E}_t, u_t) \le \eta.$$

The LEP price is

$$p_t \geq \min_{x \in S_t^+} \langle x, u_t \rangle .$$

Since $\theta_t \in S_t^+$ and the realized valuation is nonnegative, the clipped LEP price cannot exceed the current valuation, and hence

$$\begin{aligned} \langle \theta_t, u_t \rangle - p_t &\leq \max_{x \in S_t^+} \langle x, u_t \rangle - \min_{x \in S_t^+} \langle x, u_t \rangle \\ &= \mathrm{width}(S_t^+, u_t) \leq \eta. \end{aligned}$$

Thus every accepted LEP round in a good batch incurs regret at most η.

Since exploration prices are clipped to $[0, 1]$, exploration rounds incur regret at most 1. By Lemma D.2, each batch contains only $\widetilde{O}(d^2)$ exploration rounds. Hence every good batch B_i satisfies

$$R_i \leq O(d^2 \log \frac{20R(d+1)}{\eta}) + |B_i|\eta.$$

For a bad batch, we use the crude bound

$$R_i \leq |B_i| \leq a.$$

Since every bad batch satisfies $\Delta_i > \delta$ and the batches are disjoint, the number of bad batches is at most V/δ.

By Lemma D.3, the total number of batches is at most

$$r \leq 1 + \frac{T}{a} + \frac{V}{\delta}.$$

Therefore

$$\textsc{Regret}[\textsc{Alg}] \leq O(d^2 \log \frac{20R(d+1)}{\eta}) \left(\frac{T}{a} + \frac{V}{\delta} + 1 \right) + T\eta + a\frac{V}{\delta}.$$

For notational simplicity, we ignore the ceiling in a in the tuning calculation below. Since the unrounded quantity is at least one, this changes the bounds only by an absolute constant factor. With

$$a = \min\{T, d^{1/2} \left(\frac{T}{V} \right)^{1/3}\}, \qquad \delta = \max\{\frac{1}{T}, d^{1/2} \left(\frac{V}{T} \right)^{1/3}\}, \qquad \eta = 4d\delta,$$

If $\eta \geq 20R(d+1)$, the claimed tuned bound follows from the trivial regret bound $\textsc{Regret}[\textsc{Alg}] \leq T$, since

$$d^{3/2} V^{1/3} T^{2/3} = \frac{\eta T}{4} \geq 5R(d+1)T.$$

Hence it suffices to consider $\eta < 20R(d+1)$, where Lemma D.2 applies.

One may introduce

$$Q := V^{1/3} T^{2/3},$$

and see that

$$\frac{T}{a} \leq 1 + \frac{Q}{\sqrt{d}}, \qquad \frac{V}{\delta} \leq \frac{Q}{\sqrt{d}},$$

Then

$$T\eta \leq 4d + 4d^{3/2} Q, \qquad a\frac{V}{\delta} \leq Q,$$

Overall proving

$$\textsc{Regret}[\textsc{Alg}] \leq O\Big((d^2 + d^{3/2} V^{1/3} T^{2/3}) \log(T) \Big), \tag{1}$$

as claimed. □

We next show that the selected single-scale tuning can incur $\Omega(V^{1/3} T^{2/3})$ regret even on a stationary instance.

Proposition D.4 (Lower bound for the selected single-scale tuning)**.** *Consider the single-scale shallow pricing algorithm in dimension $d = 2$, with parameters*

$$\delta = \max\left\{\frac{1}{T}, \sqrt{2}\left(\frac{V}{T}\right)^{1/3}\right\}, \qquad a = \min\left\{T, \left\lceil\sqrt{2}\left(\frac{T}{V}\right)^{1/3}\right\rceil\right\}, \qquad \eta = 8\delta,$$

where $V \in (0, T]$. At each restart, initialize the relaxed feasible set to

$$S_0 = \{x \in \mathbb{R}^2 : x \geq 0,\ \|x\| \leq 1\}$$

and the ellipsoid to

$$\mathcal{E}_0 = \{x \in \mathbb{R}^2 : \|x\| \leq 1\}.$$

Suppose each batch contains at most a rounds.

There exists a stationary instance with identical contexts on which the algorithm incurs dynamic regret of at least

$$\frac{3}{8\sqrt{2}} V^{1/3} T^{2/3}.$$

The same lower bound holds when exploration prices are clipped to $[0, 1]$.

Proof. For every round $t \in [T]$, set

$$u_t = u = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \qquad \theta_t = \theta = \left(\frac{3}{4\sqrt{2}}, \frac{3}{4\sqrt{2}}\right).$$

These contexts and parameters satisfy the model assumptions, and every buyer has value

$$v_t = \langle \theta, u \rangle = \frac{3}{4}.$$

Moreover, the parameter sequence is stationary, so its total variation is zero and hence does not exceed V.

Consider the first round of any batch. If the algorithm explores, it posts the midpoint of the projection of $\mathcal{E}_0$ onto u, namely

$$\langle 0, u \rangle = 0.$$

This price is accepted and incurs regret $3/4$. If the algorithm instead exploits, it posts the lower-envelope price of S_0 in direction u, which is

$$\max\left\{0, \min_{x \in S_0} \langle x, u \rangle\right\} = 0.$$

This also incurs regret $3/4$. Thus, regardless of whether the algorithm initially explores or exploits, the first round of each batch incurs $3/4$ regret. Clipping exploration prices to $[0, 1]$ does not change this conclusion, since the initial exploration price is 0.

Let r denote the number of batches. Since each batch contains at most a rounds,

$$r \geq \frac{T}{a}.$$

All other rounds incur nonnegative regret, so

$$\textsc{Regret}[\textsc{Alg}] \geq \frac{3r}{4} \geq \frac{3T}{4a}.$$

Since $V \leq T$,

$$\sqrt{2}\left(\frac{T}{V}\right)^{1/3} \geq \sqrt{2} > 1.$$

Therefore,

$$a \leq \left\lceil\sqrt{2}\left(\frac{T}{V}\right)^{1/3}\right\rceil \leq 2\sqrt{2}\left(\frac{T}{V}\right)^{1/3}.$$

Consequently,

$$\textsc{Regret}[\textsc{Alg}] \geq \frac{3}{8\sqrt{2}} V^{1/3} T^{2/3},$$

as claimed.

This lower bound matches the $V^{1/3}T^{2/3}$ dependence of the stated upper bound for the exact tuning of Algorithm 3, up to logarithmic factors when the dimension is fixed.

In particular, when $V = 1$, this tuning incurs $\Omega(T^{2/3})$ regret even on a stationary instance. The result concerns the selected parameter choice; it does not rule out improved guarantees under a different tuning of the single-scale algorithm. □

E Multiscale Contextual Analysis

This appendix gives the complete analysis of Algorithm 2. We first establish the nesting and containment properties of the relaxed feasible sets, then bound the exploration complexity and LEP regret, and finally combine these ingredients to prove Theorem 4.1.

E.1 Structural properties of the relaxed sets

Recall the dyadic scale set

$$\Lambda = \{\delta, 2\delta, \dots, 2^L\delta\},$$

where $2^L\delta \geq R$.

Lemma E.1 (Nesting across scales)**.** *If $\lambda \leq \lambda'$, then for every round t within a batch,*

$$S_t^{(\lambda)} \subseteq S_t^{(\lambda')}.$$

Consequently, the active set

$$\Lambda_t = \{\lambda \in \Lambda : S_t^{(\lambda)} \neq \varnothing\}$$

is a suffix of Λ.

Proof. For each observed feedback sign, increasing the relaxation radius only enlarges the corresponding halfspace:

$$H_j^{(\lambda)} \subseteq H_j^{(\lambda')} \qquad \text{whenever } \lambda \leq \lambda'.$$

Intersecting these halfspaces over previous rounds preserves the inclusion. The suffix property follows immediately. □

Lemma E.2 (Existence of an active scale)**.** *For every round t within a batch,*

$$\Lambda_t \neq \varnothing.$$

In particular,

$$\lambda_t^{\min} = \min \Lambda_t$$

is well-defined.

Proof. Let

$$\lambda_{\max} = 2^L\delta \geq R.$$

Fix any current round t and any previous round $j < t$ in the same batch. The parameter θ_j satisfies the exact feedback halfspace generated at round j. Since both θ_j and θ_t lie in S_0,

$$\|\theta_t - \theta_j\| \leq \operatorname{diam}(S_0) \leq R \leq \lambda_{\max}.$$

Using $\|u_j\| = 1$, the current parameter θ_t therefore satisfies the $\lambda_{\max}$-relaxed version of the feedback halfspace from round j. This holds for every previous round, so

$$\theta_t \in S_t^{(\lambda_{\max})}.$$

Hence the largest scale is active. □

E.2 Containment at the retrospective right scale

For a batch B_i, define its variation diameter by

$$\Delta_i := \max_{j,k \in B_i} \|\theta_j - \theta_k\|,$$

and define the retrospective right scale

$$\lambda_i := \min\{\lambda \in \Lambda : \lambda \geq \Delta_i\}.$$

This scale exists because

$$\Delta_i \leq \operatorname{diam}(S_0) \leq R.$$

Lemma E.3 (Multiscale containment)**.** *For every $t \in B_i$ and every $s \in B_i$ with $s \leq t$,*

$$\theta_s \in S_t^{(\lambda_i)}.$$

Moreover,

$$\lambda_i \leq 2 \max\{\Delta_i, \delta\}.$$

Proof. Fix $s \leq t$ and a previous round $j < t$ in the same batch. If $f_j = 0$, then

$$\langle \theta_j, u_j \rangle \leq p_j.$$

Because

$$\|\theta_s - \theta_j\| \leq \Delta_i \leq \lambda_i,$$

we obtain

$$\langle \theta_s, u_j \rangle \leq \langle \theta_j, u_j \rangle + \|\theta_s - \theta_j\| \leq p_j + \lambda_i.$$

Thus θ_s satisfies $H_j^{(\lambda_i)}$.

The case $f_j = 1$ is symmetric:

$$\langle \theta_s, u_j \rangle \geq \langle \theta_j, u_j \rangle - \|\theta_s - \theta_j\| \geq p_j - \lambda_i.$$

Hence θ_s satisfies every relaxed halfspace defining $S_t^{(\lambda_i)}$.

The dyadic definition of λ_i gives

$$\lambda_i < 2\Delta_i \qquad \text{whenever } \Delta_i \geq \delta,$$

while

$$\lambda_i = \delta \qquad \text{when } \Delta_i < \delta.$$

Therefore

$$\lambda_i \leq 2 \max\{\Delta_i, \delta\}.$$

□

Corollary E.4 (The right scale is active)**.** *For every $t \in B_i$,*

$$\lambda_i \in \Lambda_t \qquad \textit{and hence} \qquad \lambda_t^{\min} \leq \lambda_i.$$

Proof. Lemma E.3 gives

$$\theta_t \in S_t^{(\lambda_i)},$$

so the set is nonempty and λ_i is active. The second claim follows from the definition of $\lambda_t^{\min}$. □

E.3 EXPLORATION COMPLEXITY

Lemma E.5 (Multiscale exploration bound)**.** *Suppose $V \le T/d$. Within any fixed batch, Algorithm 2 performs at most*

$$20d^2(L+1)\ln\left(\frac{5R(d+1)}{d\delta}\right)$$

exploration rounds.

Proof. Fix a scale $\lambda \in \Lambda$. The algorithm selects this scale only when it is active and

$$\mathrm{width}(\mathcal{E}_t^{(\lambda)}, u_t) > 4d\lambda.$$

Whenever it is selected, only $\mathcal{E}_t^{(\lambda)}$ is shallow-updated using the standard λ-shifted shallow cut associated with the unclipped midpoint proposal. By Section B, this update remains a valid outer approximation when the posted exploration price is clipped at $[0,1]$. Since $V \le T/d$, we have $\delta \le 1 < R$. By the minimality of L, every $\lambda \in \Lambda$ satisfies $\lambda < 2R$. Hence

$$4d\lambda < 8dR < 20R(d+1),$$

so Lemma B.1 applies with $\eta = 4d\lambda$, bounding the number of times this scale can be selected by

$$20d^2\ln\left(\frac{20R(d+1)}{4d\lambda}\right) \le 20d^2\ln\left(\frac{5R(d+1)}{d\delta}\right),$$

where we used $\lambda \ge \delta$. Summing over the $L+1$ dyadic scales proves the claim. □

E.4 ACCEPTED-LEP REGRET

Lemma E.6 (Accepted LEP regret at the right scale)**.** *Consider an accepted LEP round t in batch B_i. Then*

$$\langle \theta_t, u_t\rangle - p_t \le 8d\max\{\Delta_i, \delta\}.$$

Proof. By Corollary E.4, λ_i is active and

$$\lambda_t^{\min} \le \lambda_i.$$

Since round t is an LEP round, no active scale violates its width threshold. In particular,

$$\mathrm{width}(\mathcal{E}_t^{(\lambda_i)}, u_t) \le 4d\lambda_i.$$

Using

$$S_t^{(\lambda_i)} \subseteq \mathcal{E}_t^{(\lambda_i)},$$

we also have

$$\mathrm{width}(S_t^{(\lambda_i)}, u_t) \le 4d\lambda_i.$$

By Lemma E.1,

$$S_t^{(\lambda_t^{\min})} \subseteq S_t^{(\lambda_i)}.$$

Therefore, by monotonicity of $z \mapsto \max\{0, z\}$,

$$\begin{aligned} p_t &= \mathrm{LEP}\big(S_t^{(\lambda_t^{\min})}, u_t\big) \\ &\ge \mathrm{LEP}\big(S_t^{(\lambda_i)}, u_t\big) \\ &\ge \min_{x \in S_t^{(\lambda_i)}} \langle x, u_t\rangle . \end{aligned}$$

Lemma E.3 gives

$$\theta_t \in S_t^{(\lambda_i)}.$$

Since the LEP round is accepted,

$$\begin{aligned} \langle \theta_t, u_t\rangle - p_t &\le \langle \theta_t, u_t\rangle - \min_{x \in S_t^{(\lambda_i)}} \langle x, u_t\rangle \\ &\le \mathrm{width}(S_t^{(\lambda_i)}, u_t) \\ &\le 4d\lambda_i \\ &\le 8d\max\{\Delta_i, \delta\}. \end{aligned}$$

□

E.5 RESET COUNT

Lemma E.7 (Multiscale reset count)**.** *The number of LEP-rejection resets in Algorithm 2 is at most*

$$\frac{V}{\delta}.$$

Consequently, the total number of batches satisfies

$$r \le 1 + \frac{T}{a} + \frac{V}{\delta}.$$

Proof. Consider a rejected LEP round t. Rejection means

$$\langle \theta_t, u_t \rangle < p_t.$$

Since the realized valuation is nonnegative, we must have $p_t > 0$. Hence the clipping at zero is inactive, and therefore

$$p_t = \mathrm{LEP}\big(S_t^{(\lambda_t^{\min})}, u_t\big) = \min_{x \in S_t^{(\lambda_t^{\min})}} \langle x, u_t \rangle .$$

It follows that

$$\theta_t \notin S_t^{(\lambda_t^{\min})}.$$

Hence θ_t violates some earlier $\lambda_t^{\min}$-relaxed feedback halfspace generated at a round j in the same batch, while θ_j satisfied the corresponding exact halfspace.

If $f_j = 0$, then

$$\langle \theta_j, u_j \rangle \le p_j, \qquad \langle \theta_t, u_j \rangle > p_j + \lambda_t^{\min},$$

so

$$\|\theta_t - \theta_j\| > \lambda_t^{\min} \ge \delta.$$

The case $f_j = 1$ is symmetric. Thus every LEP-rejection reset certifies more than δ variation within its batch.

Since reset batches are disjoint and the total variation budget is at most V, there can be at most V/δ such resets. Window-size resets contribute at most T/a additional batches, plus one final batch. □

E.6 PROOF OF THEOREM 4.1

Proof. If $V > T/d$, the claimed bound follows from the trivial regret bound $\mathrm{REGRET}[\mathrm{ALG}] \le T$, since

$$d^{3/2}\sqrt{TV} \ge T.$$

Hence, for the remainder of the proof, it suffices to consider $V \le T/d$. Let

$$B_1, \dots, B_r$$

be the batches. Since the batches are disjoint intervals,

$$\sum_{i=1}^{r} \Delta_i \le \mathcal{V}(\Theta) \le V.$$

We decompose regret into exploration rounds, accepted LEP rounds, and rejected LEP rounds.

Since exploration prices are nonnegative and realized valuations lie in $[0, 1]$, each exploration round incurs regret at most 1. Then by Lemma E.5, exploration contributes at most

$$20d^2(L+1)\ln\left(\frac{5R(d+1)}{d\delta}\right) r.$$

For accepted LEP rounds in batch B_i, Lemma E.6 gives per-round regret at most

$$8d\max\{\Delta_i, \delta\} \le 8d(\Delta_i + \delta).$$

Since

$$|B_i| \le a,$$

summing over batches yields

$$\begin{aligned}\text{accepted-LEP regret} &\le 8da\sum_{i=1}^{r}\Delta_i + 8d\delta\sum_{i=1}^{r}|B_i| \\ &\le 8daV + 8d\delta T.\end{aligned}$$

By Lemma E.7, there are at most V/δ rejected LEP rounds, each with regret at most 1. Moreover,

$$r \le 1 + \frac{T}{a} + \frac{V}{\delta}.$$

Therefore

$$\begin{aligned}\textsc{Regret}[\textsc{Alg}] \le 20d^2(L+1)\ln\!\left(\frac{5R(d+1)}{d\delta}\right)\left(1+\frac{T}{a}+\frac{V}{\delta}\right) \\ + 8daV + 8d\delta T + \frac{V}{\delta}.\end{aligned}$$

For notational simplicity, in the tuning calculation below we ignore the ceiling in the definition of a. This affects the bounds only by an absolute constant factor: for $x \ge 1$,

$$x \le \lceil x \rceil \le 2x,$$

and the additional truncation by T preserves this constant-factor relation. Thus replacing $\min\{T, \lceil\sqrt{dT/V}\rceil\}$ by $\min\{T, \sqrt{dT/V}\}$ does not affect the stated big-O regret bound. Choose

$$a = \min\{T, \sqrt{\frac{dT}{V}}\}, \qquad \delta = \max\{\frac{1}{T}, \sqrt{\frac{dV}{T}}\}.$$

Then

$$\frac{T}{a} \le 1 + \sqrt{\frac{VT}{d}}, \qquad \frac{V}{\delta} \le \sqrt{\frac{VT}{d}},$$

and

$$daV \le d^{3/2}\sqrt{VT}, \qquad d\delta T \le d + d^{3/2}\sqrt{VT}.$$

Finally,

$$L + 1 = O\!\left(1 + \log\frac{R}{\delta}\right),$$

Hence

$$\textsc{Regret}[\textsc{Alg}] = O\!\left((d^2 + d^{3/2}\sqrt{VT})\log^2(T)\right).$$

□