Intrinsic-VAPE: Contextual Dynamic Pricing on Doubling Metrics with Cross-Learned Demand

Liz Lemma Future Detective

January 16, 2026

Abstract

Modern pricing systems condition on high-dimensional embeddings of users, queries, and products. Existing nonparametric regret guarantees for contextual dynamic pricing typically scale exponentially with the ambient dimension because they rely on Euclidean coverings of the context space. Building on the VAPE framework (Valuation Approximation–Price Elimination), which separates learning the valuation function from learning a transferable demand curve over price increments, we develop an intrinsic-complexity version that replaces fixed coverings with adaptive partitions driven by the metric structure of contexts. We model contexts as points in a metric space with doubling dimension m and assume the expected valuation g is Hölder-continuous in that metric, while valuations are distorted by bounded i.i.d. noise with Lipschitz CDF. Our algorithm combines (i) zooming-style adaptive sampling to estimate g only where needed and (ii) a global successive-elimination routine over price increments that shares demand information across contexts as in VAPE. We prove a high-probability regret bound of order $\tilde{O}\big(T^{(m+2\beta)/(m+3\beta)}\big)$, replacing ambient dimension by intrinsic dimension. The analysis highlights a distinctive challenge of pricing relative to standard bandits: accurate demand estimation is needed over a wide range of increments because optimal markups vary with $g(x)$. We formalize how adaptive valuation learning can be coupled with cross-context demand learning without losing transfer benefits, providing a theory aligned with embedding-era (2026) pricing pipelines.

Table of Contents

ations; noise and Lipschitz demand; regret benchmark and discussion of what is and is not identified.

3. 3. From Euclidean coverings to intrinsic complexity: doubling metrics, hierarchical nets/trees, zooming dimension vs doubling dimension; implications for sampling valuation approximation.

4. 4. Intrinsic-VAPE algorithm: (i) adaptive valuation approximation module on a context tree (when/where to sample uniform prices), (ii) global increment grid and demand estimation shared across contexts, (iii) price elimination with time-varying feasible increments; computational complexity and implementable variants.

5. 5. Main theorems: regret bound with intrinsic dimension; high-probability events and parameter choices; interpretation of the exponent $(m + 2\beta)/(m + 3\beta)$ and log factors.

6. 6. Proof sketch: decomposition into valuation-approximation regret + price-elimination regret; controlling valuation error via tree confidence radii; coupling lemma ensuring demand estimates remain nearly unbiased despite context-dependent estimation error; elimination counting argument.

7. 7. Extensions: (a) manifold assumption (smooth submanifold in $\mathbb{R}^d$), (b) learned representation with online metric refinement (requires numerical methods / representation learning), (c) partial pooling across segments as optional add-on; limitations.

8. 8. Simulations / empirical protocol: synthetic doubling-metric contexts and embedding-like contexts; comparison to fixed-cover VAPE; sensitivity to intrinsic dimension and smoothness.

9. 9. Conclusion: implications for 2026 pricing systems; open problems (lower bounds in intrinsic dimension; drift; fairness constraints).

1 Introduction

Digital platforms increasingly set prices in environments where the seller observes rich, high-dimensional signals before quoting a take-it-or-leave-it offer. A marketplace sees a vector of browsing events; a mobility platform sees a spatio-temporal state, weather, and supply indicators; a subscription service sees a learned embedding summarizing a user's history. From an econometric viewpoint, these signals are "contexts" that shift willingness to pay; from a systems viewpoint, they are features produced by upstream machine-learning models. In both cases, the practical promise is the same: better targeting should improve revenue relative to uniform pricing.

The theoretical bottleneck is that the standard nonparametric route to contextual pricing—treating the valuation function as Lipschitz (or Hölder) in an ambient Euclidean space and then covering that space uniformly—scales poorly with the raw feature dimension. If the context lives in $\mathbb{R}^d$ and we insist on guarantees that hold uniformly over the space, the number of regions required to localize the valuation at resolution ϵ is on the order of $\epsilon^{-d/\beta}$. Even in moderately sized embeddings, this curse of dimensionality makes the implied sample requirements astronomical. The mismatch with practice is stark: modern representations are often intentionally overparameterized, but the induced data geometry is typically low-dimensional in an intrinsic sense (for instance, because behavior concentrates near a manifold, or because only a few latent factors matter). A theory that takes the embedding dimension d at face value therefore risks being pessimistic precisely in the regimes where practitioners most rely on representation learning.

This paper develops a regret analysis for contextual dynamic pricing that depends on *intrinsic* complexity of the context space rather than its ambient description. Our organizing hypothesis is that the seller should be able to learn as quickly as the metric geometry permits: if the arrival sequence of contexts explores only a low-dimensional subset, the algorithm should refine only there; if the metric space itself has bounded growth (in the sense of a doubling condition), then adaptive partitioning should exploit that structure. This viewpoint is not merely aesthetic. It aligns the model with how platforms actually deploy personalization: they do not attempt to learn a separate price for every point in a huge feature space, but instead generalize across "nearby" contexts defined by a similarity metric (often the one implicitly used by the embedding).

Two additional features of the pricing problem sharpen the need for a geometry-aware approach. First, the seller typically observes only *binary feedback*: whether the buyer purchased at the posted price. In contrast to contextual regression, the valuation is never directly observed, and in contrast to standard stochastic bandits, the "reward" is censored and endogenous to the price. Second, there are (at least) two unknown objects interacting: a context-dependent valuation level and a context-invariant noise distribu-

tion (equivalently, a demand curve). A seller who conflates these sources of uncertainty may learn slowly even in low-dimensional settings, because the same data must simultaneously support cross-context generalization and price-response estimation.

Our starting point is a structural idea introduced in the VAPE line of work: separate the problem into (i) learning the *valuation level* $g(x)$ that shifts with context, and (ii) learning a one-dimensional *demand curve over price increments* that is shared across contexts. Intuitively, even if buyers differ systematically across contexts, the seller can view each transaction as: "value equals a baseline determined by x plus a noise shock," so the purchase decision depends on the increment $\delta = p - g(x)$. When this decomposition is valid, data collected at different contexts can be pooled to estimate demand at the same increment, which is substantially easier than learning a separate demand curve for each context.

Yet implementing this decomposition in a regret-minimizing algorithm raises a subtle difficulty that we refer to as the *wide increment range* challenge. In an idealized world where $g(x)$ were known, the seller could choose an increment δ and post $p = g(x)+\delta$, thus generating a clean Bernoulli observation with success probability $D(\delta)$. In reality, the seller replaces $g(x)$ by an estimate $\hat{g}(x)$ and posts $p = \hat{g}(x) + \delta$. The resulting purchase probability is then $D(\delta+\hat{g}(x)-g(x))$, so the demand observation is *biased* by the valuation error. This bias is not a mere nuisance term: if increments are chosen from a wide grid (which they must be, ex ante, because the optimal increment is unknown), then even a modest valuation error can shift the effective increment enough to blur the algorithm's comparisons among candidate prices. Put differently, the seller is trying to learn a one-dimensional curve $D(\cdot)$, but the covariate noise induced by mis-estimating $g(\cdot)$ can destroy the very pooling that makes the approach attractive.

VAPE addresses this tension by interleaving two modules. The first is a *valuation approximation* routine that occasionally posts randomized prices to obtain an unbiased signal for $g(x)$ from one-bit feedback. The second is a *price increment elimination* routine that treats the residual problem as a one-dimensional bandit over a discretized increment set, pooling demand estimates across contexts whenever the same increment is played. The conceptual appeal is clear: valuation learning is "local" in context space, while demand learning is "global" and therefore statistically efficient. The limitation of the original ambient-dimension analysis is also clear: making valuation errors uniformly small requires a covering of the context space at resolution $\epsilon^{1/\beta}$, and the number of cells scales with the ambient dimension.

Our main contribution is to show that this architecture can be upgraded from ambient-dimension coverings to *intrinsic* adaptive partitions in a general metric space. Concretely, we replace a fixed ϵ-net by a zooming-style hierarchical tree whose refinement is governed by the doubling structure of the metric. When contexts concentrate in a low-complexity region, the tree

expands primarily there; when the metric admits small covers at fine scales, the total number of active cells grows like $\epsilon^{-m/\beta}$, where m is the doubling dimension. This delivers regret bounds that scale with m rather than the ambient embedding dimension, thereby reconciling theory with the empirical reality that many high-dimensional representations have low intrinsic complexity.

A second contribution is analytical: we formalize how valuation error propagates into demand estimation under *adaptive* data collection. The wide increment range challenge is inherently sequential: the algorithm's increment choices depend on past outcomes and on the current valuation estimates, which themselves depend on past randomized probing. We develop a coupling argument showing that, on a high-probability event where valuation errors are controlled, the induced bias in demand estimation is Lipschitz in that error and can be absorbed into the confidence bounds used for elimination. This result clarifies why cross-context pooling remains valid even though each demand observation is generated at an *effective* increment that is slightly perturbed by estimation error. Economically, it says that if the demand curve is not too steep, then small mis-calibrations of the valuation level do not fundamentally undermine learning about price sensitivity.

The resulting algorithm, which we call Intrinsic-VAPE, inherits the desirable modularity of VAPE while adapting to the geometry of contexts. At a high level, it (i) maintains a context tree and triggers additional valuation-approximation probes only when the current cell is too large to guarantee sufficiently small valuation bias, and (ii) runs a global elimination procedure over a grid of increments using pooled purchase outcomes. The regret bound decomposes into interpretable pieces: a discretization term from approximating the continuous increment choice by a grid, an estimation term from learning demand over that grid, and a valuation-approximation term whose magnitude is controlled by the intrinsic covering growth. Optimizing the discretization level yields a rate governed by the intrinsic dimension m and the Hölder smoothness parameter β, capturing the natural tradeoff between how fast valuation varies across contexts and how complex the context space is.

We also view the analysis as offering guidance for practice. When a platform adopts a new embedding or similarity metric, the question is not only predictive accuracy but also *learnability under feedback constraints*. Our results suggest that what matters for dynamic pricing is the metric's doubling behavior (or related intrinsic complexity measures), because it controls how many distinct "pricing neighborhoods" must be explored to calibrate valuations. Moreover, the decomposition into valuation and increment learning highlights a design principle: invest exploration budget in a way that stabilizes the mapping from price to increment (i.e., keep $|\hat{g}(x) - g(x)|$ small enough), and then exploit pooling to learn the demand curve quickly. This is especially relevant when experimentation is costly or constrained by policy,

since one-bit feedback already limits information per interaction.

At the same time, the model deliberately abstracts from several real-world complications. We assume an additive valuation structure with a context-invariant noise distribution, which rules out heteroskedasticity or context-dependent price sensitivity beyond a location shift. We focus on static regret relative to the best context-dependent posted price, rather than dynamic benchmarks that would allow the policy to react to latent state variables. We also rely on regularity of the demand curve (Lipschitzness) to control the bias induced by valuation errors. These assumptions are motivated by tractability and by the goal of isolating how intrinsic geometry affects learning, but they should be revisited when applying the approach to domains with strong nonstationarities, strategic behavior, or richer feedback.

Roadmap. Section 2 formalizes the interaction protocol, the valuation model, and the regret benchmark, and clarifies which objects are and are not identified from binary outcomes. Section 3 develops the structural decomposition that motivates learning demand over increments and pooling across contexts. Section 4 presents Intrinsic-VAPE, emphasizing the interface between the adaptive context tree and the global increment elimination. Section 5 provides the regret analysis, including the intrinsic-dimension bound on valuation approximation complexity and the coupling argument for demand estimation under adaptive valuation error. We conclude with extensions and discussion of when intrinsic-geometry guarantees are likely to be informative in applied pricing systems.

2 Model

2.1 Contextual posted-price interaction

We study a seller who repeatedly posts take-it-or-leave-it prices to a sequence of myopic buyers over a finite horizon of T rounds. In round $t \in \{1, \dots, T\}$, the seller first observes a *context* x_t—a feature vector, state, or embedding produced by an upstream system—and then chooses a price p_t from a known feasible interval. After the price is posted, the buyer either purchases or not; the seller observes only this binary outcome and the realized revenue.

Formally, contexts take values in a metric space $(\mathcal{X}, \rho)$ with bounded diameter,

$$\operatorname{diam}(\mathcal{X}) \le B_x.$$

We allow the context sequence $(x_t)_{t=1}^T$ to be arbitrary, and in particular not drawn i.i.d.; the seller's learning guarantee is therefore meaningful in settings where the platform faces shifting traffic patterns, targeted experimentation, or other forms of nonstationary arrivals. The seller's policy maps the current context and the past interaction history into a price,

$$p_t \in [0, B_y], \qquad p_t = \phi_t\big(x_t; (x_s, p_s, o_s)_{s<t}\big),$$

where $o_s \in \{0, 1\}$ denotes whether a purchase occurred at round s. The seller then observes

$$o_t = \mathbf{1}\{p_t \le y_t\}, \qquad r_t = p_t o_t,$$

where y_t is the buyer's valuation and r_t is the realized revenue.

The restriction to binary feedback is central. It captures the common operational situation in which the platform observes whether a transaction occurred but not the buyer's willingness to pay, and it emphasizes the informational bottleneck: each interaction provides at most one bit about demand at the quoted price. This differs from contextual regression or auction settings, where richer price–quantity variation (or bids) may be available.

2.2 Additive valuations and intrinsic regularity

We assume valuations follow an additive, context-shifted structure:

$$y_t = g(x_t) + \xi_t. \tag{1}$$

The unknown function $g : \mathcal{X} \to \mathbb{R}$ represents the systematic component of willingness to pay that varies with context, while ξ_t captures idiosyncratic dispersion around that level. We impose the boundedness condition

$$|g(x)| \le B_g \quad \text{for all } x \in \mathcal{X},$$

and we assume the noise terms are i.i.d., centered, and bounded:

$$\mathbb{E}[\xi_t] = 0, \qquad |\xi_t| \le B_\xi \text{ a.s.},$$

independent of $(x_s, p_s, o_s)_{s<t}$ and of x_t. Consequently valuations are bounded as $|y_t| \le B_y$ with $B_y := B_g + B_\xi$, and it is without loss to restrict prices to $[0, B_y]$.

The additive structure in (1) can be viewed as a location-shift model for heterogeneity: context changes the "baseline" valuation through $g(x)$, while the distribution of deviations is stable across contexts. This assumption is common in empirical demand modeling and often a reasonable approximation when contexts are engineered to summarize buyer type. At the same time, it deliberately rules out context-dependent slopes (e.g., different price sensitivities in different segments) and heteroskedastic dispersion. We return to this limitation in the discussion below; for now, we emphasize that our goal is to isolate a setting where cross-context pooling is in principle possible and to study how geometry of contexts affects learnability under one-bit feedback.

To capture regularity of the systematic component across similar contexts, we assume g is Hölder continuous with respect to the metric ρ. That is, for parameters $L_g > 0$ and $\beta \in (0, 1]$,

$$|g(x) - g(x')| \le L_g \, \rho(x, x')^\beta \qquad \text{for all } x, x' \in \mathcal{X}. \tag{2}$$

When $\beta = 1$ this is the familiar Lipschitz condition. Economically, (2) formalizes that small changes in context imply small changes in willingness to pay, a premise that underlies most personalization systems. Importantly, we do not assume $\mathcal{X}$ is Euclidean; ρ can be an application-driven similarity metric (for example, a distance between embeddings). This flexibility is critical because, in practice, the relevant geometry is often induced by representation learning rather than by raw covariates.

Finally, we will measure the intrinsic complexity of $(\mathcal{X}, \rho)$ via a doubling condition. We say $\mathcal{X}$ has doubling dimension m if every ball of radius r can be covered by at most 2^m balls of radius $r/2$. We use this property only as a structural assumption on the space—it expresses that the number of "distinct neighborhoods" does not explode too quickly as we zoom in. The algorithmic consequences of doubling structure, and the contrast with Euclidean coverings, are developed in the next section.

2.3 Noise, demand, and the role of Lipschitzness

Let F denote the cumulative distribution function of ξ_t. The purchase decision under price p and context x is

$$o = \mathbf{1}\{p \leq g(x) + \xi\}.$$

Equivalently, purchases occur when the noise shock exceeds the *increment* $\delta := p - g(x)$. This motivates defining the *demand function over increments*

$$D(\delta) := \mathbb{P}(\xi \geq \delta) = 1 - F(\delta).$$

We assume F is Lipschitz with constant L_ξ, which is equivalent to D being Lipschitz with the same constant:

$$|D(\delta) - D(\delta')| \leq L_\xi |\delta - \delta'|. \tag{3}$$

This regularity plays a dual role. First, it is economically interpretable: demand should not change discontinuously with a small perturbation of price relative to the valuation level. Second, and more technically, it controls how errors in estimating $g(x)$ translate into errors in interpreting observed purchase outcomes as information about $D(\cdot)$. Because the seller will typically post p using an estimate $\hat{g}(x)$ rather than the truth, Lipschitzness is what allows us to bound the resulting "increment miscalibration" effect.

The centered-noise assumption $\mathbb{E}[\xi] = 0$ is a normalization that pins down $g(x)$ as the conditional mean valuation level in (1). Without some normalization, (g, F) is not uniquely parameterized: shifting g by a constant and shifting ξ by the negative of that constant leaves y unchanged. We emphasize, however, that even with centering, the seller does not *observe* y_t, so $g(x)$ is not directly identifiable from a single context without price

variation. Identification in this environment is inherently tied to experimentation: learning requires posting prices that induce informative accept/reject outcomes.

A deeper identification issue concerns the extent to which one can disentangle $g(\cdot)$ from F using only binary purchases. In our model, such disentanglement is possible in principle because contexts shift valuations while the noise distribution is stable, so variation across contexts can be used to learn the shape of $D(\cdot)$ once levels are controlled. But this is precisely where the feedback constraint bites: the seller must simultaneously (i) localize $g(x)$ well enough that posted prices correspond to known increments, and (ii) sample enough increments to recover the revenue-relevant features of D. Our regret analysis is designed to make this tradeoff explicit.

2.4 Expected revenue and the regret benchmark

Given context x and posted price p, the seller's expected revenue is

$$\begin{aligned}\pi(x,p) &:= \mathbb{E}\big[p\,\mathbf{1}\{p \le y\} \mid x,p\big] = p\,\mathbb{P}(p \le g(x)+\xi) \\ &= p\,D\big(p-g(x)\big). \end{aligned} \tag{4}$$

The dependence on x is entirely through the shift $g(x)$; the demand curve $D(\cdot)$ is shared across contexts. This representation clarifies what the seller is trying to do. For each x, the seller would like to choose a price that balances a higher margin against a lower purchase probability, but it can do so efficiently only if it can map the observed context into an accurate valuation level, thereby converting prices into increments at which demand information is comparable across rounds.

Let

$$p^*(x) \in \arg\max_{p\in[0,B_y]} \pi(x,p)$$

denote an optimal context-dependent posted price. We evaluate performance by *static regret* relative to this oracle that knows g and D:

$$R_T := \sum_{t=1}^{T} \Big(\pi(x_t,p^*(x_t)) - \pi(x_t,p_t)\Big). \tag{5}$$

This benchmark is natural for two reasons. First, it corresponds to the best myopic posted-price policy under the assumed stationary valuation model; it is therefore the right target if we view each round as an independent selling opportunity with no intertemporal constraints. Second, it isolates the learning problem created by one-bit feedback: even achieving the static optimum requires exploration because the seller does not know which prices are profitable at which contexts.

We stress what this benchmark does *not* capture. It does not reward dynamic price discrimination that conditions on latent state beyond x_t, nor

does it incorporate forward-looking buyer responses or strategic waiting. It also does not allow the oracle to choose randomized prices that might be optimal under alternative objectives (e.g., information acquisition). Our focus is the canonical online-learning goal: compete with the best feasible deterministic price as a function of observed context, under a stationary but unknown demand environment.

2.5 Discussion: what the model captures, and what it abstracts from

The combination of (1), (2), and (3) is intended to reflect a pragmatic view of personalization in platforms. Contexts are high-dimensional but organized by a similarity metric; valuation varies smoothly in that metric; and residual heterogeneity is stable enough across contexts that pooling is meaningful once we account for level shifts. Under these conditions, the fundamental obstacle is not the absence of structure but rather the scarcity of information per interaction. Binary feedback forces the seller to "probe" the valuation level to calibrate increments, and it must do so while also learning where the demand curve makes revenue highest.

At the same time, we view these assumptions as a disciplined starting point rather than as a complete description of applied pricing systems. Context-dependent dispersion (heteroskedasticity), context-dependent slopes (non-parallel demand shifts), and time variation in F would all break the clean separation implicit in (4). In such environments, cross-context pooling can become biased in a way that does not vanish with finer partitions. Our analysis therefore should be read as characterizing when pooling is justified and when geometry-aware exploration can yield meaningful gains—not as claiming that all personalization problems reduce to this structure.

With the model and benchmark in place, we next develop the geometric perspective that motivates our algorithmic design: rather than covering a high-dimensional ambient space uniformly, we exploit intrinsic properties of $(\mathcal{X}, \rho)$ that control how many distinct "valuation neighborhoods" must be explored to make increment-based demand learning reliable.

3 From Euclidean coverings to intrinsic complexity

The regret rates in contextual pricing problems are often stated in terms of an *ambient* dimension: one discretizes a d-dimensional covariate space at resolution h, and the number of regions to explore scales like h^{-d}. This viewpoint is natural when $\mathcal{X} \subset \mathbb{R}^d$ with a standard norm and when d is modest. But it is a poor guide for modern applications, where contexts are typically high-dimensional embeddings, sparse vectors, or composite objects (queries, baskets, user histories) for which the relevant similarity geometry is neither Euclidean nor low-dimensional in the coordinate sense. In such

settings, the key empirical fact is that *local neighborhoods* may still be small and well-structured even when the ambient representation is huge. Our goal in this section is to formalize this idea through intrinsic covering growth, and to explain why it directly governs the number of rounds needed to "calibrate" valuations well enough for increment-based demand learning.

Why coverings enter: valuation approximation as localization. Recall that the seller never observes y_t; it only observes whether a posted price clears. To interpret a purchase outcome as information about the increment $\delta = p - g(x)$, the seller must know $g(x)$ to reasonable accuracy. If we were willing to ignore geometry, we could treat each context as unrelated and attempt to learn a separate valuation level for each distinct x_t, but this is hopeless when $\mathcal{X}$ is large or continuous. The Hölder condition (2) is the bridge: if two contexts are close in ρ, then their valuations are close, and we can share information locally.

A convenient way to operationalize local sharing is to partition $\mathcal{X}$ into regions (cells) whose ρ-diameter is at most h, estimate $g(\cdot)$ on each cell, and then extend the estimate to all points in that cell. The approximation error decomposes into a *bias* term and a *statistical* term. If $\mathrm{rad}(u)$ denotes a scale parameter for a cell u (for instance, an upper bound on its diameter), then Hölder continuity implies that within that cell,

$$\sup_{x,x' \in \mathrm{cell}(u)} |g(x) - g(x')| \;\le\; L_g \,\mathrm{rad}(u)^{\beta}.$$

Thus, even with infinite data, using a single "representative" valuation for the whole cell induces a bias on the order of $L_g \mathrm{rad}(u)^{\beta}$. On the other hand, given only one-bit outcomes, estimating the representative valuation requires repeated probing (as in VAPE-style uniform pricing), and concentration yields a statistical error that scales like $n_u^{-1/2}$ after n_u probes in that cell (up to logarithmic factors). A canonical target is therefore to choose n_u large enough that

$$(\text{statistical error}) \;\lesssim\; (\text{geometric bias}) \;\approx\; L_g \,\mathrm{rad}(u)^{\beta},$$

since pushing statistical error below the intrinsic cell bias yields diminishing returns. The central question becomes: *how many cells must we maintain at each scale*, and hence how many total valuation-probing rounds are needed to achieve a uniform approximation accuracy?

The Euclidean benchmark and its limitations. If $\mathcal{X} \subset \mathbb{R}^d$ and ρ is induced by a norm, a standard construction is an h-net (or grid) of $\mathcal{X}$: a set of points whose h-balls cover the space. The size of a minimal h-cover is governed by the covering number $N(\mathcal{X}, \rho, h)$, which in Euclidean settings behaves like h^{-d} up to constants. If we wish to ensure a valuation bias at most ϵ, we choose $h \asymp (\epsilon / L_g)^{1/\beta}$, producing $N(\mathcal{X}, \rho, h) \asymp \epsilon^{-d/\beta}$ regions.

If each region requires $\tilde{O}(\epsilon^{-2})$ one-bit probes to estimate its representative valuation to accuracy ϵ, the valuation-approximation budget scales like

$$\tilde{O}\Big(\epsilon^{-d/\beta}\,\epsilon^{-2}\Big),$$

which is precisely the kind of dimension dependence that becomes prohibitive when d is large.

Two practical points make this benchmark even less appealing. First, in embedding-based systems, Euclidean distances in the ambient coordinates often exaggerate degrees of freedom: many directions correspond to noise or are rarely explored by arriving contexts. Second, fixed covers are *wasteful* when contexts arrive non-uniformly. A uniform grid pays for cells that are never visited, yet in online pricing the seller only receives feedback along the realized context trajectory. These observations motivate a geometry that captures intrinsic neighborhood growth and an algorithmic mechanism that refines space only where data arrive.

Doubling metrics: intrinsic neighborhood growth. The doubling condition provides a succinct, representation-invariant way to encode that "balls do not explode" as we zoom in. Formally, $(\mathcal{X}, \rho)$ has doubling dimension m if every ball of radius r can be covered by at most 2^m balls of radius $r/2$. Iterating this property implies polynomial (rather than exponential) growth of covering numbers: there exist constants (depending on the doubling constant) such that for all $0 < h \le B_x$,

$$N(\mathcal{X}, \rho, h) \;\le\; C\left(\frac{B_x}{h}\right)^m.$$

Thus, if the intrinsic doubling dimension is $m \ll d$, then the number of h-scale regions needed to control the Hölder bias is on the order of h^{-m}, not h^{-d}. Substituting $h \asymp (\epsilon/L_g)^{1/\beta}$ yields the intrinsic analogue of the Euclidean complexity:

$$N(\mathcal{X}, \rho, (\epsilon/L_g)^{1/\beta}) \;\lesssim\; \left(\frac{L_g^{1/\beta} B_x}{\epsilon^{1/\beta}}\right)^m \;\asymp\; \epsilon^{-m/\beta}.$$

This is the first place where intrinsic dimension enters our analysis: it controls how many distinct "valuation neighborhoods" must be separately calibrated to guarantee a uniform ϵ-accurate approximation to g.

Economically, one can interpret m as a bound on the effective richness of segmentation induced by the similarity metric. When m is small, a finite number of localized experiments can, in principle, calibrate valuation levels across the entire population of contexts. When m is large, personalization is intrinsically expensive: the platform faces many locally distinct valuation regimes and must spend more of its horizon learning them.

Hierarchical nets and trees: organizing multiscale exploration. Covering-number bounds are informative, but to convert them into an online learning algorithm we need a *data structure* that supports (i) multiscale localization, (ii) refinement where needed, and (iii) efficient updates along the realized context path. Hierarchical nets provide such a structure. At a high level, we build a sequence of nets at scales $r_0 > r_1 > r_2 > \cdots$ (typically geometrically decreasing, e.g. $r_\ell = 2^{-\ell} B_x$). At each scale r_ℓ, we select a set of centers $C_\ell \subset \mathcal{X}$ such that (a) C_ℓ is an r_ℓ-cover (every point is within r_ℓ of some center), and (b) centers are well-separated (a packing property). The doubling assumption ensures that each center at scale r_ℓ has only a bounded number (on the order of 2^m) of descendants at the next finer scale, yielding a tree-like refinement graph.

This construction induces a *context tree* $\mathcal{T}$ whose nodes correspond to cells $\mathrm{cell}(u)$ at various radii $\mathrm{rad}(u)$. Each round, the observed context x_t is assigned to a leaf at the finest active scale by descending the tree along the unique chain of cells containing x_t. Crucially, we do not need to materialize the full tree down to a target resolution everywhere. We can create and refine nodes only when contexts actually arrive in their vicinity, which is how we avoid paying for unvisited regions.

From an implementation viewpoint, such trees are not exotic: they are closely related to cover trees and navigating nets used in nearest-neighbor search, and they align naturally with the way platforms already bucket users or queries at multiple granularities. The mathematical role of the tree is simply to make the doubling geometry operational for learning.

Zooming versus doubling: data-dependent refinement and worst-case guarantees. It is useful to distinguish two notions of intrinsic complexity that appear in related bandit literatures. The doubling dimension m is a *uniform* property of the entire space, providing worst-case control on covering growth. A "zooming" notion, by contrast, is *data-dependent*: it measures the effective dimension of the subset of contexts (or near-optimal regions) actually explored by the algorithm. In stochastic settings where contexts are drawn from a fixed distribution, or in Lipschitz bandits where the learner focuses on near-optimal actions, zooming-type quantities can yield sharper rates than uniform covering bounds.

In our setting, however, contexts may be arbitrary, and the seller must guarantee regret control even under adversarial arrivals. This pushes us toward a worst-case geometry: we cannot assume the sequence concentrates on a low-dimensional subset, and we must be prepared to refine wherever traffic goes. The doubling dimension is therefore the appropriate baseline: it is weak enough to apply to non-Euclidean spaces, yet strong enough to yield finite-sample complexity bounds that do not blow up with the ambient representation. That said, our tree-based approach is inherently adaptive,

and in benign instances it effectively behaves like a zooming algorithm by refining only the regions that are visited frequently. The formal regret bound we state later is thus a conservative guarantee; in practice, one should expect better performance when arrivals are concentrated.

Implications for valuation-approximation sampling. We now connect these geometric ideas to the number of rounds needed to estimate $g(\cdot)$ sufficiently well for increment-based demand learning. Fix a target uniform valuation accuracy $\epsilon > 0$. A natural design is to ensure that whenever we treat two contexts as belonging to the same cell, the cell radius satisfies

$$L_g \operatorname{rad}(u)^\beta \;\lesssim\; \epsilon,$$

so that geometric bias is at most ϵ. Under doubling structure, the number of active cells at this radius scale is on the order of $\epsilon^{-m/\beta}$. Within each such cell, we need enough binary probes to estimate the cell's representative valuation up to statistical error ϵ. Because the feedback is one-bit, the variance per probe is bounded and concentration yields a sample requirement of order $\tilde{O}(\epsilon^{-2})$ per cell (the logarithmic factors account for uniform control over times and cells). Putting the pieces together gives the intrinsic valuation-approximation budget

$$\mathrm{TVA}(\epsilon) \;\approx\; \tilde{O}\Big(\epsilon^{-m/\beta}\,\epsilon^{-2}\Big),$$

which replaces the Euclidean $\epsilon^{-d/\beta}\epsilon^{-2}$ with its intrinsic counterpart. This term is not yet a regret bound, but it already captures the core economic tradeoff: finer personalization (smaller ϵ) reduces the bias with which we interpret purchases as demand information at particular increments, but it requires more localized experiments to calibrate g across the context space.

A practical interpretation is that the platform must spend a nontrivial fraction of its horizon performing "calibration" experiments that may not be immediately revenue-maximizing, especially when the intrinsic segmentation complexity m is large or when g is rough (small β). Conversely, when m is small and g is smooth, calibration can be amortized across many nearby contexts, enabling the seller to quickly enter a regime where most rounds can be devoted to optimizing increments rather than learning levels.

Limitations and what geometry does not solve. Doubling structure addresses *where* valuation experiments need to occur (how many distinct neighborhoods exist), but it does not remove the one-bit bottleneck: even in a single cell, estimating a valuation level to precision ϵ requires on the order of ϵ^{-2} probes. Moreover, doubling assumptions are only as meaningful as the metric ρ. If ρ fails to reflect true similarity in valuations (for example, if embeddings are misaligned with willingness-to-pay), then small

doubling dimension may not translate into small valuation variation, and any geometry-aware method will suffer. This is not merely a technical caveat: it points to a systems-level complementarity between representation learning (choosing ρ) and pricing experimentation (learning g and D).

With these geometric tools in hand, we are ready to describe our algorithmic construction. The key idea is to couple a tree-based valuation module, which uses the intrinsic metric structure to decide *when* and *where* to probe, with a global increment-learning module, which pools demand information across contexts once increments are properly calibrated.

4 Intrinsic-VAPE: adaptive calibration with pooled increment learning

Our algorithmic construction has two separable objectives that must be coordinated online. First, we need a *local* procedure that decides when a region of contexts has been "calibrated" well enough that purchase outcomes can be interpreted as information about the increment $\delta = p - g(x)$. Second, conditional on that calibration, we need a *global* procedure that learns the one-dimensional demand curve over increments, and then translates that knowledge into context-dependent prices via the identity $p = g(x) + \delta$. Intrinsic-VAPE implements this separation explicitly through (i) an adaptive valuation-approximation module organized on a context tree $\mathcal{T}$, and (ii) a pooled increment-learning module over a grid $\mathcal{K}$, equipped with successive elimination and time-varying feasibility constraints.

A two-mode view of each round. At round t we observe x_t and then place the round into one of two modes:

1. *Calibration mode* (valuation approximation): we post a price chosen solely to sharpen our estimate of $g(x_t)$ (more precisely, of the local valuation level within the cell containing x_t).

2. *Optimization mode* (increment learning/selection): we post a price of the form $p_t = \hat{g}_t(x_t) + \delta$ with δ chosen to learn or exploit the global demand curve $D(\cdot)$.

The conceptual discipline is that calibration rounds are charged to the "TVA" term, while optimization rounds are where pooled learning of D occurs. In benign instances, the algorithm quickly transitions to optimization mode for most contexts; in worst-case instances, the doubling geometry controls how many distinct neighborhoods must be calibrated before this can happen.

4.1 Module I: adaptive valuation approximation on a context tree

Tree representation and active cells. We maintain a hierarchical net/tree $\mathcal{T}$ over $(\mathcal{X}, \rho)$. Each node u corresponds to a cell $\mathrm{cell}(u) \subseteq \mathcal{X}$ and carries a radius parameter $\mathrm{rad}(u)$, typically on a geometric scale $\mathrm{rad}(u) \in \{2^{-\ell} B_x : \ell \geq 0\}$. A context x is associated with a unique *active leaf* by descending the tree along the chain of cells containing x until no finer active child exists. We emphasize that $\mathcal{T}$ is maintained *lazily*: a node is created (and later refined) only when contexts arrive in its vicinity, so the data structure grows with the realized trajectory rather than with an ex ante cover.

A one-bit valuation signal via uniform probing. During calibration, we use a VAPE-style randomization to extract an unbiased (or clipped-unbiased) signal about the latent valuation $y_t = g(x_t) + \xi_t$ from a single purchase bit. Concretely, we draw a probing price $P_t \sim \mathrm{Unif}[0, B_y]$ independent of the past, post $p_t = P_t$, and observe $o_t = \mathbf{1}\{P_t \leq y_t\}$. Define the pseudo-observation

$$Z_t \;=\; B_y \, o_t.$$

When valuations lie in $[0, B_y]$, we have $\mathbb{E}[Z_t \mid y_t] = y_t$; more generally, Z_t is an unbiased signal for y_t after truncation at 0, which is sufficient for our purposes because negative valuations optimally map to near-zero prices in the feasible set. Since ξ_t is centered and independent, $\mathbb{E}[Z_t \mid x_t]$ tracks $g(x_t)$ up to the same truncation, and the boundedness $|Z_t| \leq B_y$ yields sub-Gaussian concentration for cell averages.

Cell-wise estimators and resolution criteria. Each node u maintains a count $n(u)$ of calibration probes assigned to it and an empirical mean

$$\hat{g}(u) \;=\; \frac{1}{n(u)} \sum_{s \leq t : x_s \in \mathrm{cell}(u),\; s \text{ calibration}} Z_s,$$

with the convention that $\hat{g}(u) = 0$ when $n(u) = 0$, and with clipping to $[-B_g, B_g]$ if desired. To decide whether a cell is "resolved," we track a confidence radius that mirrors the bias–variance decomposition:

$$\mathrm{conf}_t(u) \;=\; c_1 \sqrt{\frac{\log(T)}{n(u) \vee 1}} \;+\; L_g \, \mathrm{rad}(u)^{\beta}.$$

The first term is statistical (one-bit concentration), and the second is geometric (Hölder bias within the cell). We declare u *calibrated* at time t if $\mathrm{conf}_t(u) \leq \epsilon_u$ for a target tolerance ϵ_u at that scale; in the simplest uniform-accuracy implementation, we take $\epsilon_u \equiv \epsilon$ for all active leaves.

When to probe and when to refine. Given x_t, let u_t be its active leaf. If u_t is not calibrated, we enter calibration mode and post a probing price P_t. The resulting Z_t is used to update statistics along the visited path (either only at u_t, or at all ancestors as a variance-reduction device). Refinement is triggered when a coarse cell has accumulated enough probes that its statistical uncertainty is already below its geometric bias, i.e.,

$$c_1\sqrt{\frac{\log(T)}{n(u)}} \;\le\; c_2\, L_g\, \mathrm{rad}(u)^{\beta},$$

at which point further probing without splitting would be inefficient. We then create children of u at radius $\mathrm{rad}(u)/2$ (or the next net scale), and subsequently route future contexts to these finer cells. The doubling property ensures that each node has only a bounded number of children (on the order of 2^m up to constants), so the total number of active nodes at a given finest radius is controlled by intrinsic rather than ambient dimension.

4.2 Module II: pooled increment grid and demand estimation

Increment discretization and feasibility. In optimization mode, we restrict attention to an increment grid

$$\mathcal{K} \;=\; \{\delta_k = k\epsilon : k \in \mathbb{Z},\ |\delta_k| \le B_y + 1\},$$

and translate an increment into a price through the current valuation estimate:

$$p_t \;=\; \mathrm{clip}_{[0,B_y]}\big(\hat{g}_t(x_t) + \delta_k\big).$$

The feasibility constraint is *time-varying* because $\hat{g}_t(x_t)$ evolves and because clipping can distort the intended increment. Operationally, we define the feasible increment set at (t, x_t) as

$$\mathcal{K}_t(x_t) \;=\; \{\delta_k \in \mathcal{K} : 0 \le \hat{g}_t(x_t) + \delta_k \le B_y\}\,,$$

and we only play increments in $\mathcal{K}_t(x_t)$ during optimization rounds. This detail matters both computationally (we avoid meaningless arms) and analytically (we preserve the structural identity $\pi(x, g(x)+\delta) = (g(x)+\delta)D(\delta)$ up to controlled estimation error).

Pooled demand statistics. For each grid point δ_k, we maintain global counters:

$$N_t(k) \;=\; \sum_{s\le t} \mathbf{1}\{\text{round } s \text{ played } \delta_k\}, \qquad S_t(k) \;=\; \sum_{s\le t} \mathbf{1}\{\text{round } s \text{ played } \delta_k\}\, o_s,$$

and the pooled estimator $\hat{D}_t(\delta_k) = S_t(k)/(N_t(k) \vee 1)$. The key economic reason pooling is valid is that $D(\cdot)$ depends only on the noise distribution and not on x. The key statistical caveat is that when we post $p_t = \hat{g}_t(x_t) + \delta_k$, the purchase probability is $D(\delta_k + \hat{g}_t(x_t) - g(x_t))$, not exactly $D(\delta_k)$. Lipschitzness of D converts this into a bias term of order $L_\xi|\hat{g}_t(x_t) - g(x_t)|$, which is precisely why the calibration module is a prerequisite for reliable demand learning.

Revenue confidence bounds for increment selection. On an optimization round in cell u_t, we use the calibrated estimate $\hat{v}_t = \hat{g}(u_t)$ and its associated tolerance ϵ_{u_t} to form an interval for the valuation level,

$$v \in [\hat{v}_t - \epsilon_{u_t},\ \hat{v}_t + \epsilon_{u_t}].$$

For each feasible increment $\delta_k \in \mathcal{K}_t(x_t)$ we also maintain a confidence band for demand,

$$D(\delta_k) \in \left[\hat{D}_t(\delta_k) - b_t(k),\ \hat{D}_t(\delta_k) + b_t(k)\right], \qquad b_t(k) = c_3\sqrt{\frac{\log(T)}{N_t(k) \vee 1}} + L_\xi \epsilon_{u_t},$$

where the second term is the calibration-induced bias cushion. Combining these, we obtain conservative upper and lower bounds on the expected revenue of increment δ_k at the current context:

$$\mathrm{UCB}_t(k) = (\hat{v}_t + \epsilon_{u_t} + \delta_k)\left(\hat{D}_t(\delta_k) + b_t(k)\right), \qquad \mathrm{LCB}_t(k) = (\hat{v}_t - \epsilon_{u_t} + \delta_k)\left(\hat{D}_t(\delta_k) - b_t(k)\right),$$

with the understanding that negative prices are excluded by feasibility. These bounds formalize the sense in which uncertainty about g and uncertainty about D enter multiplicatively in revenue.

4.3 Module III: successive elimination with a time-varying active set

Why elimination rather than plain exploration. A naive approach would pick k via an upper-confidence rule each round. We instead use successive elimination because it cleanly separates "learning to the current tolerance" from "shrinking the decision set," and because it interfaces naturally with our calibration tolerances: once all surviving increments have demand confidence width at most the cell-specific bias scale, further demand exploration is not economically valuable in that cell.

Elimination rule and phase structure. We maintain an active set $\mathcal{A}_t \subseteq \mathcal{K}$ of increments that have not yet been ruled out. Periodically (e.g. when all $k \in \mathcal{A}_t$ have been sampled enough that $c_3\sqrt{\log(T)/N_t(k)}$ falls below

a phase threshold), we compare increments using the current revenue confidence bounds. Given x_t and its calibrated valuation interval, we eliminate any $k \in \mathcal{A}_t \cap \mathcal{K}_t(x_t)$ such that

$$\mathrm{UCB}_t(k) \;<\; \max_{j \in \mathcal{A}_t \cap \mathcal{K}_t(x_t)} \mathrm{LCB}_t(j),$$

since such an increment cannot be optimal at this context within the current uncertainty. Importantly, elimination is *context-conditional*: an increment may be infeasible or dominated at one valuation level and yet remain potentially optimal at another, so we eliminate only when the confidence bounds imply dominance at the valuation levels that actually arise. This is where the intrinsic tree matters: by restricting attention to calibrated cells, the valuation range relevant for elimination is narrow, which speeds up pruning in practice.

Choosing which increment to play. On optimization rounds, we select an increment from $\mathcal{A}_t \cap \mathcal{K}_t(x_t)$ that balances exploitation and information acquisition. One implementable rule is to pick the maximizer of $\mathrm{UCB}_t(k)$; another, closer to textbook successive elimination, is to cycle through increments whose $N_t(k)$ is below a phase quota. Both variants ensure that increments with large uncertainty are sampled often enough to be eliminated (or confirmed) quickly, while increments that are already well-estimated are played primarily when they are plausibly revenue-maximizing at the current $\hat{v}_t$.

4.4 Computational complexity and implementable variants

Per-round cost and data structures. The tree module requires updating statistics only on the path from the root to the active leaf containing x_t, which has depth $O(\log(B_x/\mathrm{rad}_{\min}))$ and is $O(\log T)$ under geometric scaling. Each node update is constant time, so valuation maintenance is polylogarithmic per round. The increment module maintains $(N_t(k), S_t(k))$ for active grid points and can support UCB/LCB queries with either (i) a linear scan over the currently feasible active set, or (ii) a priority queue keyed by $\mathrm{UCB}_t(k)$ when $\mathcal{A}_t$ is large. Since $\mathcal{K}_t(x_t)$ is an interval in k (feasibility is a price-range constraint), we can also maintain active increments in an ordered structure and query maxima in logarithmic time.

Practical variants. Several implementation choices do not affect the conceptual separation of modules. First, one can replace an explicit hierarchical net with standard metric trees (cover trees, navigating nets) that support approximate nearest-center assignment in $O(\log|\text{active nodes}|)$ time. Second, one can batch elimination checks at dyadic times to avoid constant recomputation of UCB/LCB; this changes only logarithmic factors. Third,

when B_y is large and clipping becomes frequent, it is often preferable to shrink $\mathcal{K}_t(x_t)$ by a safety margin (e.g. enforce $p_t \in [\eta, B_y - \eta]$) so that the intended increment is realized without truncation; economically, this corresponds to avoiding boundary prices where revenue sensitivity is dominated by feasibility rather than demand.

What the algorithm does and does not buy us. Intrinsic-VAPE leverages geometry only to reduce the number of *distinct valuation neighborhoods* that require calibration; it does not circumvent the fundamental one-bit nature of the feedback. In particular, if contexts arrive so that many fine cells must be activated (large intrinsic complexity along the realized path), calibration will consume a substantial fraction of the horizon. Conversely, once calibration is achieved, the increment module enjoys strong cross-context sharing: every optimization round contributes to learning the same function $D(\cdot)$, so incremental improvements in demand estimation are immediately usable across the entire space. The next section formalizes this tradeoff in the regret bound by combining the tree-controlled calibration budget with high-probability confidence events for demand estimation and elimination.

5 Main theorems: intrinsic-dimension regret and how to read it

We now state the performance guarantees for Intrinsic-VAPE and clarify the roles of the intrinsic dimension m, the Hölder smoothness (L_g, β), and the one-bit nature of feedback. The main message is that the algorithm pays for *learning the valuation surface* only through the number of context neighborhoods that must be resolved along the realized trajectory, and this number is governed by doubling geometry rather than any ambient embedding dimension. Once valuations are locally calibrated, *demand learning over increments* becomes a one-dimensional problem with global data sharing.

Regret bound in tunable form. Fix a target uniform valuation accuracy parameter $\epsilon \in (0, 1]$ and an increment grid resolution of the same order.[1] Consider the event on which (i) every active cell u declared calibrated indeed satisfies $\sup_{x \in \mathrm{cell}(u)} |\hat{g}(x) - g(x)| \lesssim \epsilon$, and (ii) the pooled increment-demand estimates satisfy simultaneous concentration with a bias term proportional to the contemporaneous valuation error. On this event, the regret of Intrinsic-

[1] One can decouple these choices; we keep them comparable to simplify the statement since both ultimately trade off a linear-in-T approximation loss against inverse-polynomial estimation costs.

VAPE admits the decomposition

$$R_T \;\le\; C\Big(T\epsilon \;+\; \underbrace{\epsilon^{-2}\,\mathrm{polylog}(T)}_{\text{increment identification}} \;+\; \underbrace{\mathrm{TVA}(\epsilon)}_{\text{valuation calibration}}\Big), \tag{6}$$

where C depends polynomially on (B_y, L_ξ, L_g) and

$$\mathrm{TVA}(\epsilon) \;=\; \tilde{O}\Big(\epsilon^{-m/\beta}\,\epsilon^{-2}\Big). \tag{7}$$

The term $T\epsilon$ is the discretization/approximation loss: even with perfect knowledge of $D(\cdot)$ on the grid, restricting to increments spaced at ϵ (and allowing residual valuation uncertainty of order ϵ) can cost at most order ϵ revenue per round. The ϵ^{-2} term is the statistical cost of learning a one-dimensional demand curve from binary outcomes at grid points (up to logarithms and constants). The TVA term is the total number of calibration rounds needed to make the bias in demand observations uniformly smaller than ϵ in all regions that become active.

Main regret theorem (optimized rate). Optimizing the upper bound in (6) over ϵ yields the intrinsic-dimension rate.

Theorem (Intrinsic-VAPE static regret). *Under the standing assumptions on $(\mathcal{X}, \rho)$, g, and ξ_t, there exists an implementation of Intrinsic-VAPE and absolute constants $c, C > 0$ such that with probability at least $1 - c/T$,*

$$R_T \;\le\; \tilde{O}\Big(T^{\frac{m+2\beta}{m+3\beta}}\Big).$$

The hidden constants depend polynomially on $(B_x, B_g, B_\xi, L_g, L_\xi)$, and $\tilde{O}(\cdot)$ hides polylogarithmic factors in T and in the doubling-covering constants.

The exponent arises by balancing the linear approximation term $T\epsilon$ against the dominant estimation burden $\epsilon^{-m/\beta-2}$ coming from calibration plus demand learning. Setting

$$\epsilon \;\asymp\; T^{-\beta/(m+3\beta)}$$

gives

$$T\epsilon \;\asymp\; T^{\frac{m+2\beta}{m+3\beta}}, \qquad \epsilon^{-m/\beta-2} \;\asymp\; T^{\frac{m+2\beta}{m+3\beta}},$$

up to logarithmic factors, which is precisely the claimed scaling.

High-probability events and why they are the right ones. Because our feedback is binary and our actions are adaptive, the proof is organized around a single global good event $\mathcal{E}$ on which all confidence statements used by the algorithm are simultaneously valid. Concretely, $\mathcal{E}$ is the intersection of three types of statements.

First, for valuation calibration, we require that for every active node u and every time t at which u is used for optimization, the node-level estimator is uniformly accurate:

$$\sup_{x \in \mathrm{cell}(u)} |\hat{g}(u) - g(x)| \;\le\; c_1 \sqrt{\frac{\log(T)}{n(u) \vee 1}} + L_g \,\mathrm{rad}(u)^{\beta}. \tag{8}$$

The right-hand side matches the statistical uncertainty from bounded one-bit probing and the deterministic Hölder bias within the cell. Importantly, (8) is *time-uniform* and *node-uniform*, so that once a cell is declared calibrated, it remains safe to use its estimate as contexts revisit that region.

Second, for pooled demand estimation, we require concentration of the empirical purchase rates at each increment δ_k, but centered at the *correct* conditional mean given adaptive valuation error. Writing the purchase indicator on a round that plays increment δ_k as o_s, we have

$$\mathbb{E}[o_s \mid \mathcal{F}_{s-1}, x_s, \delta_k] \;=\; D\big(\delta_k + \hat{g}_s(x_s) - g(x_s)\big),$$

so the natural martingale difference is $o_s - \mathbb{E}[o_s \mid \cdot]$. On $\mathcal{E}$ we require that for all k and all t,

$$\left|\hat{D}_t(\delta_k) - D(\delta_k)\right| \;\le\; c_2 \sqrt{\frac{\log(T)}{N_t(k) \vee 1}} \;+\; L_\xi \epsilon_t, \tag{9}$$

where ϵ_t upper bounds the valuation error on rounds contributing to $N_t(k)$. The second term is not an artifact: it is the explicit price of using $\hat{g}_t$ to implement increments. The Lipschitz property of D makes this bias linear in the valuation error; without Lipschitzness, pooling across contexts would generally break.

Third, for elimination correctness, we require that the revenue confidence bounds computed from (8)–(9) simultaneously bracket the true expected revenue for every increment that remains feasible. This ensures that elimination never discards an increment that could be optimal at the relevant valuation levels, and that when an increment is eliminated, it is genuinely dominated within the current tolerance. Formally, $\mathcal{E}$ implies that for every t and every feasible k,

$$\mathrm{LCB}_t(k) \;\le\; (g(x_t) + \delta_k) D(\delta_k) \;\le\; \mathrm{UCB}_t(k),$$

with the understanding that feasibility is enforced so that $g(x_t) + \delta_k \in [0, B_y]$ up to the same tolerance.

How parameters enter and what must be known. The algorithm uses (B_y, L_ξ) to scale probing ranges and demand confidence widths, and (L_g, β) to decide when further calibration within a cell is no longer cost-effective relative to refinement. Knowledge of m is not required to run the procedure;

m enters only in the analysis through doubling-covering bounds that control how many cells are activated at a given radius along the realized context sequence. If (L_g, β) are unknown, one can replace the refinement rule by a more conservative schedule (or run parallel instances over a grid of candidate smoothness levels and aggregate via a standard doubling trick), at the cost of additional logarithmic factors; the rate exponent is driven primarily by the bias–variance balancing and thus remains the same in typical adaptive variants.

Interpreting the exponent $\frac{m+2\beta}{m+3\beta}$. Two limiting cases are helpful. When m is small (contexts effectively lie on a low-dimensional manifold or trajectory), the exponent approaches $2/3$ as $m \to 0$:

$$\lim_{m\to 0} \frac{m+2\beta}{m+3\beta} = \frac{2}{3},$$

reflecting that even with essentially no geometric complexity, we still face one-bit feedback and must learn a demand curve sufficiently well to choose near-optimal increments. At the other extreme, as $m \to \infty$ (or, more realistically, as the realized sequence activates many fine neighborhoods), the exponent approaches 1, indicating that calibration dominates and sublinear regret becomes harder because we must essentially relearn valuations in too many distinct places.

The role of smoothness is equally transparent. Larger β (smoother g) improves the exponent by making coarse cells informative for larger regions, which reduces the number of calibration samples needed before optimization becomes reliable. In particular, with $\beta = 1$ (Lipschitz valuations) the exponent becomes $(m+2)/(m+3)$, matching the familiar pattern in nonparametric bandits where one pays m degrees of geometric freedom plus the statistical cost of estimation.

What hides in the logarithms. The $\tilde{O}(\cdot)$ notation conceals three sources of polylogarithmic factors. The first is time-uniform concentration (union bounds over $t \le T$) for both calibration statistics and pooled demand estimates. The second is uniformity over increments k (a union bound over $|\mathcal{K}| \asymp \epsilon^{-1}$ grid points). The third is uniformity over active tree nodes; here the doubling structure is crucial, because the number of nodes at scale $\mathrm{rad}(u) \approx r$ grows like r^{-m} rather than depending on an ambient dimension. Taken together, these effects contribute multiplicative factors of order $\log(T)\log(1/\epsilon)$ (and metric-dependent constants), which are secondary to the polynomial dependence on T captured by the exponent.

Economic interpretation and limitations. From a pricing perspective, the theorem quantifies a concrete tradeoff: we must spend rounds to *calibrate*

what different contexts mean in terms of valuation, but once calibrated, every subsequent interaction refines a *single* demand curve shared across the market. The intrinsic-dimension dependence formalizes the idea that what matters for learning is not how contexts are represented, but how many meaningfully different valuation neighborhoods appear. At the same time, the rate does not escape the fundamental informational constraint of binary feedback: even with perfect geometric structure, identifying near-optimal prices requires enough experimentation to pin down purchase probabilities, and this necessity is what keeps the exponent bounded below by $2/3$ in the best geometric case. The next section shows how these ingredients combine in the proof via an explicit regret decomposition and a coupling argument that controls demand-learning bias induced by imperfect valuation estimates.

6 Proof sketch: why intrinsic geometry and one-dimensional pooling suffice

We sketch the argument behind the bound by isolating the two distinct learning problems that are intertwined by binary feedback: (i) *local valuation calibration*—we must infer $g(x_t)$ well enough to know what it means to post a given *increment* above (or below) the valuation level—and (ii) *global demand learning over increments*—once increments are implemented correctly, the purchase process depends on $\delta = p - g(x)$ only through the context-invariant demand curve $D(\delta)$. Intrinsic-VAPE is designed so that these two tasks interact in only one direction: valuation error can bias demand observations, but the algorithm allocates explicit rounds to reduce this bias, and then reuses the (nearly) unbiased demand information everywhere.

Step 0: a single "good event" controls all adaptivity. Because prices and cells are chosen adaptively, we organize the analysis around a global high-probability event $\mathcal{E}$ on which *all* confidence statements needed by the algorithm hold simultaneously for all $t \le T$, all active tree nodes u, and all increments $\delta_k \in \mathcal{K}$. Concretely, $\mathcal{E}$ is the intersection of: (i) node-uniform valuation accuracy bounds of the form

$$\sup_{x \in \mathrm{cell}(u)} |\hat{g}(u) - g(x)| \;\le\; \mathrm{rad}_g(u) \;\equiv\; c\sqrt{\frac{\log T}{n(u) \vee 1}} + L_g\,\mathrm{rad}(u)^{\beta}, \tag{10}$$

(ii) increment-uniform demand concentration bounds with a bias term driven by valuation error (formalized below), and (iii) correctness of the revenue confidence bounds used for elimination (i.e., the true revenue at every feasible increment lies within the algorithm's LCB/UCB). Standard union bounds over times, nodes, and increments (together with Freedman/Hoeffding inequalities for bounded outcomes) give $\mathbb{P}(\mathcal{E}) \ge 1 - \tilde{O}(T^{-1})$. We then bound regret on $\mathcal{E}$ and add a negligible $O(B_y)$ contribution from $\mathcal{E}^c$.

Step 1: decompose regret into valuation-approximation rounds and price-elimination rounds. Intrinsic-VAPE alternates between two modes. In *calibration* mode, it posts randomized prices within the currently active cell to estimate g to the accuracy required at that cell's scale; these rounds may be locally suboptimal but are counted explicitly. In *optimization* mode, it posts a price of the form

$$p_t = \hat{g}_t(x_t) + \delta_{k_t},$$

where the increment index k_t is chosen by a global elimination routine that uses pooled data at each δ_k.

This yields a transparent decomposition. Let $\mathcal{C}$ denote the set of calibration rounds and $\mathcal{O}$ the set of optimization rounds. Then

$$R_T \;=\; \sum_{t\in\mathcal{C}} \Big(\pi(x_t,p^*(x_t)) - \pi(x_t,p_t)\Big) \;+\; \sum_{t\in\mathcal{O}} \Big(\pi(x_t,p^*(x_t)) - \pi(x_t,p_t)\Big). \quad (11)$$

On $\mathcal{C}$, we simply upper bound per-round regret by $O(B_y)$ and focus on controlling $|\mathcal{C}|$ via intrinsic covering arguments; this is the TVA term. On $\mathcal{O}$, regret is driven by whether the elimination routine has already removed clearly suboptimal increments; this is the price-elimination term.

Step 2: control valuation error by tree confidence radii (intrinsic zooming). The valuation module maintains a hierarchical partition tree $\mathcal{T}$ adapted to the doubling metric. Each node u corresponds to a region $\mathrm{cell}(u)$ with radius proxy $\mathrm{rad}(u)$. The key design choice is the *stopping rule*: we stop refining a cell once its valuation uncertainty is no larger than the granularity at which the optimization module operates. Operationally, a node u is declared *calibrated* once $\mathrm{rad}_g(u) \lesssim \epsilon$, where ϵ is the target uniform valuation accuracy (typically matched to the increment grid step).

The bound (10) has two components with distinct economic interpretations. The term $L_g\,\mathrm{rad}(u)^\beta$ is a deterministic *heterogeneity* penalty: even with infinite data, if we treat the cell as having a single valuation level, we incur a mismatch across heterogeneous contexts within the cell. The term $\sqrt{\log T/n(u)}$ is the statistical penalty from one-bit feedback. In calibration mode we choose probing prices so that the induced binary outcomes yield an (approximately) unbiased signal for the local valuation level; then $n(u) = \tilde{O}(\epsilon^{-2})$ samples suffice to drive the statistical term down to $O(\epsilon)$.

To bound the *total* number of calibration rounds, we sum the sample requirements across all nodes that become active along the realized context sequence. Here doubling geometry enters: the number of disjoint cells of radius r that can be activated in a metric space of doubling dimension m scales like $O(r^{-m})$ (up to constants depending on the doubling cover). Since we refine until $\mathrm{rad}(u) \approx (\epsilon/L_g)^{1/\beta}$, the number of active cells at the finest

relevant scale is $\tilde{O}(\epsilon^{-m/\beta})$. Multiplying by $\tilde{O}(\epsilon^{-2})$ samples per cell yields the TVA bound

$$\mathrm{TVA}(\epsilon) \;=\; \tilde{O}\Big(\epsilon^{-m/\beta}\epsilon^{-2}\Big)\,,$$

which is precisely the intrinsic-dimension replacement for ambient $\epsilon^{-d/\beta}$ coverings.

Step 3: a coupling lemma converts valuation error into controlled demand-estimation bias. The demand module pools observations across contexts at each increment δ_k. The subtlety is that we never play the *true* increment relative to $g(x_t)$; we play it relative to $\hat{g}_t(x_t)$. If we set $p_t = \hat{g}_t(x_t) + \delta_k$, then the purchase indicator satisfies

$$o_t = \mathbf{1}\{p_t \le g(x_t) + \xi_t\} = \mathbf{1}\{\xi_t \ge \delta_k + (\hat{g}_t(x_t) - g(x_t))\}.$$

Hence the conditional mean is shifted:

$$\mathbb{E}[o_t \mid \mathcal{F}_{t-1}, x_t, \delta_k] \;=\; D(\delta_k + e_t)\,, \qquad e_t \equiv \hat{g}_t(x_t) - g(x_t). \tag{12}$$

If we were to ignore e_t, pooling would mix observations from different effective increments and generally break identification. The key observation is that D is L_ξ-Lipschitz, so small valuation errors only cause small bias:

$$|D(\delta_k + e_t) - D(\delta_k)| \le L_\xi |e_t|.$$

On the event $\mathcal{E}$ we ensure $|e_t| \le \epsilon_t$ whenever increment δ_k is used, with ϵ_t controlled by the cell's calibration radius. Define the pooled estimator $\hat{D}_t(\delta_k)$ as the empirical average of purchase indicators over rounds up to t in which increment k was played. Then we decompose

$$\hat{D}_t(\delta_k) - D(\delta_k) = \underbrace{\big(\hat{D}_t(\delta_k) - \bar{D}_t(\delta_k)\big)}_{\text{martingale noise}} + \underbrace{\big(\bar{D}_t(\delta_k) - D(\delta_k)\big)}_{\text{bias from } e_t},$$

where $\bar{D}_t(\delta_k)$ is the average of the shifted means $D(\delta_k + e_s)$ over those rounds. The first term is controlled by Freedman or Hoeffding for bounded martingale differences, giving

$$\Big|\hat{D}_t(\delta_k) - \bar{D}_t(\delta_k)\Big| \;\le\; c\sqrt{\frac{\log T}{N_t(k) \vee 1}} \quad \text{uniformly in } t, k.$$

The second term is bounded by Lipschitzness:

$$\big|\bar{D}_t(\delta_k) - D(\delta_k)\big| \;\le\; \frac{1}{N_t(k) \vee 1} \sum_{s \le t:\, k_s = k} L_\xi |e_s| \;\le\; L_\xi\, \epsilon_t,$$

where ϵ_t can be taken as an upper bound on the valuation error over all contributing rounds (or, more carefully, a maximal error across the relevant cells). This yields the coupling statement used by the algorithm:

$$\left|\hat{D}_t(\delta_k) - D(\delta_k)\right| \;\le\; c\sqrt{\frac{\log T}{N_t(k) \vee 1}} + L_\xi\, \epsilon_t, \tag{13}$$

uniformly over t and k on $\mathcal{E}$. Economically, (13) formalizes the idea that inaccurate "context calibration" does not destroy cross-context pooling; it merely acts like an additional noise term whose magnitude is proportional to how miscalibrated the valuation scale is.

Step 4: successive elimination over increments and a counting argument. In optimization mode, the algorithm maintains an active set of increments and eliminates those whose *upper* confidence bound on revenue is below the *lower* confidence bound of some competitor. Using the structural decomposition $\pi(x, g(x)+\delta) = (g(x)+\delta)D(\delta)$, the only context dependence enters through $g(x)$, while $D(\delta)$ is common. On $\mathcal{E}$, we have a calibrated estimate $\hat{g}_t(x_t)$ with error at most ϵ, and an estimate of $D(\delta_k)$ satisfying (13). Combining these yields revenue confidence intervals of width on the order of

$$\mathrm{width}_t(k) \;\approx\; B_y\left(c\sqrt{\frac{\log T}{N_t(k)}} + L_\xi\epsilon\right) \;+\; \epsilon,$$

where the final ϵ captures the effect of replacing $g(x_t)$ by $\hat{g}_t(x_t)$ in the multiplicative term $(g+\delta_k)$. The elimination rule ensures that as soon as $\mathrm{width}_t(k)$ is smaller than the suboptimality gap (up to the grid resolution), increment k is removed.

The counting step is standard in elimination analyses: each increment k must be sampled until its confidence radius drops below a target threshold. Since the stochastic part scales as $1/\sqrt{N_t(k)}$, this yields a sample requirement of order $\tilde{O}(\epsilon^{-2})$ per relevant scale, and summing across the $|\mathcal{K}| \asymp \epsilon^{-1}$ grid points while noting that many are eliminated quickly produces an aggregate contribution of order $\tilde{O}(\epsilon^{-2})$ to regret (up to polylog factors). Intuitively, despite having ϵ^{-1} candidate increments, one-dimensional structure plus elimination prevents paying ϵ^{-3}: the algorithm concentrates samples on the handful of increments near the optimum, and the rest are discarded after $O(\log T)$ evidence.

Step 5: put the pieces together. On $\mathcal{E}$, calibration rounds contribute at most $O(B_y)\cdot\mathrm{TVA}(\epsilon)$, while optimization rounds incur at most $O(\epsilon)$ per round from discretization and residual bias, plus the elimination-driven learning cost summarized above. Plugging these bounds into (11) yields the tunable form

$$R_T \;\le\; \tilde{O}\Big(T\epsilon + \epsilon^{-2} + \epsilon^{-m/\beta}\epsilon^{-2}\Big),$$

and optimizing ϵ recovers the intrinsic-dimension rate. The conceptual takeaway is that intrinsic geometry governs *how many* distinct valuation neighborhoods must be calibrated, while global pooling and elimination ensure that learning demand over increments remains a one-dimensional statistical task even under adversarial context sequences.

7 Extensions and limitations: beyond a fixed metric and fully shared demand

The analysis above is deliberately modular: one component learns the context-dependent level $g(x)$ at a resolution dictated by the geometry of $(\mathcal{X}, \rho)$, while a second component treats demand over increments $D(\delta)$ as a one-dimensional object that can be pooled globally. This modularity makes it straightforward to adapt the framework to settings that practitioners often face—where geometry is inherited from an ambient embedding, where the "right" metric is itself learned, or where pooling across all contexts is too strong an assumption. We discuss three natural extensions and then summarize the main limitations that remain.

7.1 (a) Manifold structure in an ambient space

A common modeling choice is that contexts are observed as vectors in $\mathbb{R}^d$ (customer covariates, product embeddings, text/image features), yet the effective variability lies on or near a lower-dimensional surface. Formally, suppose $\mathcal{X} \subset \mathbb{R}^d$ lies on a compact, smooth m_0-dimensional submanifold $\mathcal{M}$ with bounded curvature and reach, and take ρ to be either the ambient Euclidean metric or the geodesic distance on $\mathcal{M}$. Under standard regularity conditions, $(\mathcal{M}, \rho)$ has doubling dimension $m = \Theta(m_0)$ (up to constants depending on curvature and diameter). In this case our regret rate becomes

$$R_T \;\le\; \tilde{O}\left(T^{\frac{m_0+2\beta}{m_0+3\beta}}\right),$$

which formalizes a simple economic message: the difficulty of valuation calibration is driven by the *intrinsic* degrees of freedom in the population of contexts, not by the number of raw features recorded by the platform.

Two practical subtleties matter here. First, the choice between Euclidean and geodesic distance affects constants and, at fine scales, the effective Hölder smoothness of g. If g is β-Hölder with respect to geodesic distance but we build the tree using Euclidean distance, the distortion is typically controlled on compact manifolds, so the same rates continue to hold with adjusted constants. Second, manifold structure motivates a concrete implementation of the tree: instead of an abstract hierarchical net, one can build covers using approximate nearest-neighbor routines in $\mathbb{R}^d$, refining only around observed

contexts. This is attractive empirically because it avoids constructing global covers of a high-dimensional ambient space while still exploiting the intrinsic covering growth.

From a policy perspective, the manifold view provides a rationale for investing in richer embeddings (large d) without necessarily incurring a statistical penalty, provided the embedding is "well-behaved" and collapses onto a low-dimensional locus for the relevant traffic. It also clarifies when such optimism is unwarranted: if the embedding scatters contexts in many independent directions (high intrinsic dimension), then the valuation calibration burden necessarily grows.

7.2 (b) Learned representations and online metric refinement

In many applications the platform does not begin with a trusted metric ρ. Instead, it has raw covariates $z \in \mathcal{Z}$ and a learned representation $\phi : \mathcal{Z} \to \mathbb{R}^q$ (or a learned similarity kernel) that is updated over time. This raises a conceptual challenge: our guarantees rely on Hölder smoothness of g with respect to the metric used to build the adaptive partition, yet the platform may only have access to an evolving proxy ρ_t.

One way to extend the framework is to posit a *latent* metric space $(\mathcal{X}, \rho^\star)$ on which g is (L_g, β)-Hölder, and assume we observe a sequence of data-dependent metrics ρ_t that converge to $\rho^\star$ in the sense that

$$\sup_{x,x' \in \mathcal{X}} \left|\rho_t(x,x') - \rho^\star(x,x')\right| \;\le\; \eta_t, \qquad \eta_t \downarrow 0.$$

If the partition tree is constructed using ρ_t, then the heterogeneity term in the valuation radius effectively becomes

$$L_g \operatorname{rad}_{\rho^\star}(u)^\beta \;\lesssim\; L_g\big(\operatorname{rad}_{\rho_t}(u) + \eta_t\big)^\beta,$$

so metric error enters additively at the scale level. Economically, this says that representation error behaves like additional within-cell heterogeneity: even if we probe many prices, we cannot extrapolate valuations across contexts that the learned metric mistakenly deems "close."

Algorithmically, this suggests a two-timescale design. On the fast timescale, Intrinsic-VAPE proceeds as before using the current metric ρ_t to decide whether a cell is calibrated and which cell to refine. On the slow timescale, the platform updates ρ_t using auxiliary objectives (supervised learning from downstream outcomes, self-supervised similarity learning, or metric learning informed by observed purchase signals). The analysis would then separate (i) regret due to statistical uncertainty conditional on ρ_t from (ii) regret due to representation drift. A sufficient condition for preserving the headline rate is that η_t decreases quickly enough that the induced extra bias is dominated

by the target valuation accuracy ϵ at the time when fine-scale refinement is needed.

We emphasize a limitation: making this extension fully rigorous typically requires numerical assumptions that go beyond the bandit model (e.g., stability of representation updates, bounded sensitivity of ϕ_t to single observations, or explicit regularization). Without such conditions, the tree can "chase" a moving metric, repeatedly recalibrating regions that shift under ρ_t. In practice, a conservative implementation would freeze the representation over long epochs, rebuild or adjust the tree only between epochs, and treat representation learning as an outer loop.

7.3 (c) Partial pooling across segments: relaxing fully shared demand

The baseline model assumes that the increment demand curve $D(\delta)$ is common across all contexts. This is a powerful pooling device, but it can be too strong in environments with heterogeneous price sensitivity: different product categories, customer segments, geographies, or traffic sources can exhibit meaningfully different noise distributions. A natural compromise is *partial pooling*: we retain the structural idea that each segment has a one-dimensional demand curve over increments, but we allow a small family of curves indexed by a segment label $s \in \{1, \dots, S\}$.

One simple formulation is

$$y_t \;=\; g(x_t) + \xi_t^{(s_t)}, \qquad D_s(\delta) = \mathbb{P}(\xi^{(s)} \geq \delta),$$

where the seller observes s_t along with x_t. Then we can run the same increment-elimination routine within each segment, obtaining regret bounds that replace the global ϵ^{-2} demand-learning cost by roughly S copies (up to the fact that some segments may be rare). The economic tradeoff is transparent: more segmentation reduces model bias (wrongly pooling unlike buyers) but increases variance (fewer observations per segment).

A more flexible alternative is to impose structure across segments, such as a parametric family $D_s(\delta) = D(\delta; \theta_s)$ with θ_s lying in a low-dimensional set, or a hierarchical prior that shrinks θ_s toward a global mean. In that case, the elimination step can be replaced by a multi-task confidence procedure that shares information across segments while still guarding against negative transfer. Importantly, the valuation-calibration module need not change: it still targets $g(x)$ locally in $(\mathcal{X}, \rho)$, while the demand side becomes "one-dimensional per segment." From the standpoint of deployment, this extension is attractive because segmentation is often dictated by business constraints (fairness considerations, regulation, category management), and partial pooling provides a principled way to trade statistical efficiency against robustness.

7.4 Limitations and open directions

Several assumptions are doing real work, and relaxing them would require new ideas.

Binary feedback and bounded noise. One-bit outcomes make calibration costly: the ϵ^{-2} sample requirement per active region is intrinsic to estimating a mean from Bernoulli data. If richer feedback is available (e.g., noisy willingness-to-pay signals, multiple units, or censored quantities), the valuation module could be substantially faster. The boundedness assumptions ($|\xi_t| \le B_\xi$, $|g(x)| \le B_g$) mainly simplify concentration and ensure prices lie in a compact interval; heavy-tailed valuations would call for robust estimators and typically slower rates.

Lipschitz demand over increments. The coupling argument hinges on D being L_ξ-Lipschitz, converting valuation error into controlled bias. If D has sharp jumps (e.g., point masses in valuations), small calibration errors can cause large demand shifts, and pooling becomes fragile. In practice, this is precisely the regime where very fine price discrimination is difficult; our theory makes that difficulty explicit.

Stationarity and context invariance. We treat g and the noise distribution as time-invariant. Many markets exhibit drift (seasonality, learning buyers, competitive shocks). A pragmatic extension is to use sliding-window or discounted estimators for both $\hat{g}$ and $\hat{D}$, but proving tight regret bounds would require a dynamic-regret formulation with variation budgets. Likewise, the assumption that D is common across contexts is a modeling choice; partial pooling addresses some heterogeneity, but fully context-dependent demand would forfeit the one-dimensional reduction and reintroduce the curse of dimensionality.

Tuning and constants. While the rate depends on m and β in a clean way, implementation requires choices of grid resolution ϵ, confidence levels, and refinement thresholds. Adaptive tuning (e.g., doubling tricks over ϵ) is possible, but constants can be large when L_g or L_ξ is large, reflecting genuine economic hardness: steep demand curves and highly irregular valuations amplify the cost of miscalibration.

These extensions clarify where the framework is most useful: settings with strong cross-context commonality in price response and low intrinsic geometric complexity in valuations. The next section turns to simulation designs that make these tradeoffs visible and benchmark Intrinsic-VAPE against fixed-cover baselines.

8 Simulations and empirical protocol

We use simulations for two complementary goals. First, we want to verify that the adaptive partitioning in Intrinsic-VAPE behaves as the theory suggests: refinement should concentrate on the subset of the context space actually visited, and the number of active regions should scale with the *intrinsic* covering growth rather than an ambient dimension. Second, we want to translate the regret decomposition into operational diagnostics that are meaningful in pricing deployments—calibration error for the valuation module, and bias/variance tradeoffs for the pooled demand estimator over increments.

Common experimental scaffold. Across all environments, we fix a horizon T (typically between 2×10^4 and 2×10^5), and we simulate rounds using the same timing: observe x_t, post p_t, draw $y_t = g(x_t) + \xi_t$, observe the purchase indicator $o_t = \mathbf{1}\{p_t \le y_t\}$, and record revenue $r_t = p_t o_t$. Since the focus is static regret, we benchmark performance against the oracle policy that knows both g and D, i.e.,

$$p^\star(x) \in \arg\max_{p \in [0, B_y]} p\, D(p - g(x)), \qquad R_T = \sum_{t=1}^{T} \big(\pi(x_t, p^\star(x_t)) - \pi(x_t, p_t)\big).$$

To isolate algorithmic effects from numerical edge cases, we use bounded noise with a Lipschitz CDF; a convenient default is $\xi \sim \mathrm{Unif}[-B_\xi, B_\xi]$, for which $D(\delta) = \mathbb{P}(\xi \ge \delta)$ is piecewise linear and $L_\xi = 1/(2B_\xi)$. This choice makes it easy to compute the oracle revenue curve and to check that observed differences are not artifacts of poorly conditioned demand estimation.

Environment class I: synthetic doubling-metric contexts without an ambient embedding. To stress-test the "intrinsic geometry" claims directly, we generate contexts in spaces where the natural access model is distance queries rather than coordinates. Concretely, we sample a finite set $\{x^{(1)}, \dots, x^{(N)}\}$ and define ρ via shortest-path distance on a random geometric graph built on $[0,1]^{m_0}$ (with m_0 small), then expose to the algorithm only $(x_t, \rho(x_t, \cdot))$ rather than the latent coordinates. With standard connectivity parameters, the resulting metric has doubling dimension $m = \Theta(m_0)$ up to constants, while still allowing irregular local density (a feature that matters for adaptive refinement). We choose g to be β-Hölder with respect to ρ by constructing it as a smooth function on the latent coordinates and then composing with the graph embedding; in practice we take

$$g(x) = B_g \cdot \Big(2\sigma(h(z(x))) - 1\Big),$$

where $z(x) \in [0,1]^{m_0}$ is the latent coordinate, h is a low-frequency Fourier series, and σ is the logistic map to enforce boundedness. We vary β by control-

ling the smoothness of h (for $\beta < 1$, we additionally apply a coordinate-wise transformation $u \mapsto u^\beta$ before h, which yields a controlled Hölder exponent at small scales).

In this class, Intrinsic-VAPE builds its hierarchical net using only ρ, while the main baseline is a fixed-cover version of VAPE that partitions the space at a pre-chosen resolution (equivalently, a fixed-radius cover constructed once at the beginning). Because fixed covers require specifying the scale ex ante, we report performance both at an "oracle-tuned" scale (chosen to minimize regret in hindsight over a grid of candidate scales) and at a practically tuned scale (chosen by a doubling trick over the resolution parameter). This comparison isolates the value of adaptivity even when one is generous to the fixed-cover approach.

Environment class II: embedding-like contexts with low-dimensional structure in high ambient dimension. To mirror modern pricing systems that rely on large embeddings, we also simulate contexts as vectors in $\mathbb{R}^d$ with $d \in \{20, 50, 100\}$, while restricting the data-generating variation to an m_0-dimensional factor. A simple and transparent construction is

$$x = Au + \sigma_{\text{nuis}}\,\varepsilon, \qquad u \sim \text{Unif}([0,1]^{m_0}),\ \varepsilon \sim \mathcal{N}(0, I_d),$$

where $A \in \mathbb{R}^{d \times m_0}$ has orthonormal columns and σ_{nuis} controls how much irrelevant variation contaminates distances. We then define ρ as Euclidean distance in $\mathbb{R}^d$ (the common practitioner choice) and set $g(x) = \tilde{g}(u)$ to depend only on the intrinsic coordinates. When σ_{nuis} is small, Euclidean distances remain approximately bi-Lipschitz to the intrinsic distances on the locus, and the tree should refine at a rate governed by m_0 rather than d; when σ_{nuis} is large, this becomes a diagnostic for representation quality, since "nearby" points in ρ are no longer economically similar.

In this setting we compare (i) Intrinsic-VAPE using Euclidean distance; (ii) Intrinsic-VAPE using an approximate geodesic distance computed from a k-NN graph (to mimic manifold-aware similarity); and (iii) fixed-cover VAPE implemented as an axis-aligned grid in $\mathbb{R}^d$ (which becomes infeasible quickly as d grows, forcing coarse resolutions). The last baseline is intentionally harsh: its purpose is to visualize the curse of dimensionality that motivates intrinsic adaptation, rather than to serve as a practical competitor in very high dimension.

Baselines and ablations. Beyond fixed-cover VAPE, we include two diagnostic variants that help interpret where gains come from. The first is a "no pooling" ablation: we keep the same adaptive tree for valuation calibration but estimate a separate demand curve $\hat{D}_u(\cdot)$ within each active cell u. This removes the one-dimensional global sharing and should degrade sharply

as the number of active cells grows. The second is a "no refinement" ablation: we keep global pooling over increments but replace the tree module by a global valuation estimator that does not localize (e.g., a single $\hat{g}$ fit with the same one-bit probing logic). This variant is useful when contexts are nearly homogeneous; it should fail when g varies meaningfully across $\mathcal{X}$.

Implementation details and reporting choices. To keep the comparison faithful, we align algorithms on the same increment grid $\mathcal{K}$ and the same confidence level (e.g., $\delta = 1/T^2$), and we enforce the same price bounds $[0, B_y]$. For Intrinsic-VAPE, we implement the tree as a cover tree / navigating net variant that supports insertion of new contexts online and returns the path of active nodes covering x_t; empirically, this makes per-round cost close to logarithmic in the number of visited contexts. For the fixed-cover baselines, we precompute the cover (or grid) and map each x_t to a cell by nearest center.

We report not only regret but also module-level diagnostics that correspond to the theoretical decomposition:

1. *Calibration frequency:* the fraction of rounds spent in valuation-approximation mode.

2. *Empirical valuation error:* for simulated data where g is known, we compute $|\hat{g}_t(x_t) - g(x_t)|$ and also a cell-level proxy $\sup_{x \in \mathrm{cell}(u)} |\hat{g}(x) - g(x)|$ by Monte Carlo sampling within the cell (where applicable).

3. *Demand estimation error:* $\max_{\delta_k \in \mathcal{K}} |\hat{D}_t(\delta_k) - D(\delta_k)|$ evaluated at checkpoints.

4. *Active tree size:* number of active nodes, and distribution of radii $\mathrm{rad}(u)$ among active nodes.

These quantities let us check whether observed regret differences are driven by discretization, by demand learning variance, or by miscalibration bias leaking into pooled demand estimates.

Sensitivity to intrinsic dimension m_0. We vary $m_0 \in \{1, 2, 4, 8\}$ while holding fixed d (in embedding-like contexts) and holding fixed the noise distribution and bounds. The main predicted pattern is a smooth deterioration in the log-log slope of R_T versus T, consistent with the exponent $(m+2\beta)/(m+3\beta)$. In practice, finite-sample slopes are noisy, so we use two complementary summaries: (i) for each (m_0, β) we fit a line to $\log R_T$ against $\log T$ over a range of T and report the fitted slope; (ii) we plot the normalized regret $R_T / T^{(m_0+2\beta)/(m_0+3\beta)}$ to see whether curves flatten as T grows. The most informative diagnostic tends to be active tree size: Intrinsic-VAPE should increase the number of active cells roughly like a covering number at

the finest explored scale, while fixed-cover methods either explode (if fine) or incur persistent discretization regret (if coarse).

Sensitivity to smoothness β and heterogeneity scale L_g. We vary $\beta \in \{1, 0.75, 0.5\}$ and separately scale L_g by multiplicative factors. Holding all else fixed, decreasing β should cause the tree to refine more aggressively to achieve the same valuation bias control; empirically this appears as (a) a larger share of calibration rounds, and (b) smaller typical $\mathrm{rad}(u)$ among active nodes. Because demand learning is pooled, the demand estimation curves $\hat{D}_t(\cdot)$ should look similar across β once calibration error is controlled; when they do not, it is evidence that valuation error is dominating the coupling term. Increasing L_g has a similar qualitative effect but is primarily visible in constants: the same regret slope with an upward shift, and a delayed transition from coarse to fine cells.

Embedding distortion and "representation quality" stress tests. To connect back to practice, we interpret σ_{nuis} (or, alternatively, random feature rotations that scramble intrinsic neighborhoods) as a proxy for representation error. As σ_{nuis} grows, Euclidean distance becomes less informative, and Intrinsic-VAPE will either (i) over-refine because it cannot pool effectively within a cell that now contains heterogeneous intrinsic types, or (ii) under-refine because true neighbors are far in the learned metric, reducing reuse. Both effects increase calibration cost and ultimately regret. Using k-NN geodesic distances partially mitigates this when the intrinsic locus remains locally connected, and the comparison between Euclidean and geodesic versions provides an empirical analogue of the metric-misspecification discussion in the extensions section: representation error behaves like extra within-cell heterogeneity.

Takeaways we expect to be robust. Across these designs, the qualitative lessons are stable. Intrinsic-VAPE tends to dominate fixed-cover baselines when (i) contexts have low intrinsic dimension relative to their ambient description, and (ii) pooling demand over increments is approximately correct. The no-pooling ablation typically performs competitively only in the easiest regimes (very small m_0 and very smooth g), highlighting that the one-dimensional demand reduction is not merely a technical convenience but a primary source of sample efficiency. Conversely, when representation quality is poor (large σ_{nuis}) or when g is highly irregular (small β, large L_g), Intrinsic-VAPE still degrades gracefully but the constant factors become visible: more rounds are spent calibrating valuations, and demand learning slows because valuation error induces bias. In this sense, the simulations operationalize the core tradeoff: we can share aggressively across contexts only to the extent that we can certify that the valuation level $g(x)$ has been locally calibrated

at the granularity implied by the metric geometry.

9 Conclusion: implications for 2026 pricing systems and open problems

Our central message is that *geometry* is an economically meaningful primitive in contextual pricing with limited feedback. When the seller only observes a purchase indicator, learning must proceed by carefully chosen perturbations of price, and the value of those perturbations hinges on whether we can treat nearby contexts as approximately substitutable. The main theorem formalizes this idea in a way that is aligned with how modern pricing systems are built: contexts are often represented as embeddings, similarity is encoded by a metric, and the relevant complexity is not the ambient dimension of the representation but the *intrinsic* growth of neighborhoods under that metric. Intrinsic-VAPE makes this tradeoff explicit: we gain from global pooling of the demand curve over increments, but we can only pool safely to the extent that we can locally calibrate valuation levels $g(x)$ at the granularity dictated by the metric structure.

Implications for 2026 pricing stacks: learning should be representation-aware. Many deployed pricing pipelines in 2026 rely on high-dimensional user–item or user–market embeddings learned from historical interactions. In such stacks, "contextual pricing" is frequently implemented by regressing purchase probability on price and features, or by doing local bandit exploration in embedding space. Our results suggest a more diagnostic posture. If the metric induced by the embedding is well-aligned with economic substitutability, then an adaptive partition that refines only where the platform actually operates can deliver the same qualitative benefit as hand-designed segmentations, but with (i) explicit exploration guarantees and (ii) a complexity bound governed by an intrinsic dimension proxy rather than the raw embedding dimension. Conversely, if embedding distances are distorted (because nuisance variation dominates Euclidean neighborhoods, or because the embedding collapses relevant heterogeneity), then the algorithm will reveal this via operational metrics: it will either spend too many rounds calibrating valuations (over-refinement) or fail to reuse information effectively (under-refinement). In this sense, the learning problem itself becomes a stress test for representation quality.

Operational interpretation of the regret decomposition. The decomposition behind Intrinsic-VAPE has a natural engineering reading. The term that scales like $T\epsilon$ corresponds to *discretization and approximation*: how much revenue we lose by restricting attention to a finite increment grid, and by acting on an approximate valuation estimate. The terms that scale

like ϵ^{-2} and the tree complexity term reflect *statistical effort*: how many random-price probes are needed to (i) pin down the local valuation level using one-bit outcomes and (ii) estimate a one-dimensional demand curve over increments with enough precision to eliminate suboptimal prices. In practice, these pieces map onto familiar monitoring objects: a calibration budget (fraction of exploration rounds), a stability/bias check for pooled demand estimates, and a notion of "active segmentation" (how many distinct regions the system effectively treats differently). This mapping matters because it tells a platform where to invest: better similarity metrics reduce the calibration burden; better instrumentation (e.g., richer feedback than binary purchase) reduces the ϵ^{-2} cost; and tighter operational constraints (e.g., price-change limits) effectively impose a minimum feasible ϵ.

A limitation that is also a design principle: pooling is only as good as invariances. A strong assumption in the model is that demand over increments, $D(\delta)$, is context-invariant. The virtue of this assumption is conceptual clarity: it isolates one dimension of variation (valuation level $g(x)$) that can be locally calibrated, and another dimension (the noise law) that can be learned globally. But it also highlights a practical design principle: *global pooling is not a free lunch.* In real markets, the distribution of idiosyncratic shocks may vary by geography, device type, or acquisition channel, so the appropriate object to pool may be a low-dimensional family $\{D_\theta\}_{\theta\in\Theta}$ rather than a single curve. One way to read Intrinsic-VAPE is as a template: whenever we can identify a context-invariant (or slowly varying) component of the purchase process, we should pool it aggressively, and then spend calibration effort on the remaining heterogeneity. The converse is also true: if invariance fails, then pooling introduces systematic bias that no amount of exploration can remove. This is why it is valuable that the coupling argument makes the bias channel explicit through Lipschitz demand: it tells us what to measure (how much valuation error leaks into demand estimation) and what to relax (introduce partially pooled, cluster-specific demand curves) when the invariance is not credible.

From theory to practice: why intrinsic dimension is a useful knob. The intrinsic dimension m is not merely a technical artifact; it provides a language for capacity planning in pricing systems. In segment-based pricing, teams often debate how many segments to maintain, which segments to split, and how to justify the resulting complexity. The doubling dimension offers a principled proxy: if the context metric has small doubling dimension on the support of traffic, then a system can afford finer personalization because the number of "distinct" neighborhoods grows moderately as resolution increases. If m is large, then aggressive personalization is sample-inefficient under binary feedback, and the platform should either accept coarser policies

or invest in richer signals (e.g., auction-style feedback, counterfactual logging that improves variance, or additional covariates that render the metric more informative). Put differently, our bound suggests an empirical workflow: estimate an intrinsic dimension proxy on the observed context stream, then choose an exploration budget and a refinement policy consistent with that complexity.

Open problem I: lower bounds in intrinsic dimension for one-bit pricing. While our upper bound replaces ambient dimension by intrinsic dimension, a complete story requires matching lower bounds. A natural conjecture is that the exponent $\frac{m+2\beta}{m+3\beta}$ is minimax-optimal (up to logarithms) over doubling-metric context spaces and (L_g, β)-Hölder valuations, under bounded one-bit feedback. Proving this would require constructing hard instances that simultaneously (i) force local estimation of g at scale ϵ in about $\epsilon^{-m/\beta}$ distinct regions and (ii) force demand learning at the canonical ϵ^{-2} rate from binary outcomes. Technically, this is subtle because the difficulty is shared between two coupled unknowns (g and D), and because the adversary can choose the context sequence. A sharp lower bound would clarify which part of the regret is information-theoretic and which part is an artifact of algorithm design, and it would also guide whether further improvements should target demand estimation (e.g., adaptive discretizations over δ) or valuation calibration (e.g., more efficient probing schemes).

Open problem II: drift and nonstationarity (dynamic regret). Static regret is the right starting point, but pricing systems increasingly operate in environments with demand shocks, seasonality, and policy-driven discontinuities. The model already separates a stable demand curve over increments from a context-dependent valuation level; this suggests several drift regimes worth formalizing. One regime is *slow drift in* $g_t(x)$: valuations shift over time due to trend or competition, but remain Hölder in x at each t. Another regime is *drift in* D_t: the idiosyncratic noise law changes with macro conditions. In both cases, the key question is how to trade off continual recalibration against exploitation, and how intrinsic dimension interacts with a variation budget (e.g., $\sum_t \sup_x |g_{t+1}(x) - g_t(x)|$ or a Wasserstein drift of the noise distribution). Algorithmically, one expects windowed estimators and tree nodes with "expiration" or time-decay; theoretically, one would like dynamic regret bounds that interpolate between the intrinsic-dimension rate in stationary settings and the unavoidable tracking cost under rapid drift. From a deployment perspective, such results would support principled resets and monitoring thresholds: when drift accelerates, the system should admit higher exploration and temporarily coarsen personalization to avoid compounding bias.

Open problem III: fairness, constraints, and welfare objectives. Pricing is increasingly constrained by fairness and compliance requirements. In our framework, constraints can enter in at least three distinct ways. First, one can constrain the *policy class*, e.g., enforce Lipschitz or bounded-variation prices across similar contexts: $|p(x)-p(x')| \le L_p\rho(x,x')$. This kind of constraint aligns naturally with metric structure and may reduce effective complexity, but it also changes the benchmark and can create nontrivial tension with revenue maximization. Second, one can impose *group fairness* constraints, such as bounding price dispersion across protected groups conditional on observable context, or imposing constraints on acceptance rates. With binary feedback, enforcing these constraints online while learning g and D is nontrivial: naive exploration can violate constraints even if the long-run policy is feasible. Third, one can broaden the objective beyond revenue to include consumer surplus or error costs (e.g., penalties for overpricing relative to latent willingness to pay). Each direction raises new questions about what can be learned from one-bit signals and how to certify constraint satisfaction with high probability. A promising approach is to integrate constraint-aware confidence sets into the elimination step, so that candidate increments are pruned not only by revenue but also by worst-case constraint violation within the confidence region.

Further directions: richer feedback, multiple products, and strategic behavior. Two additional extensions are likely to matter in near-term practice. The first is to relax the one-bit feedback assumption by incorporating partial refund signals, dwell time, or add-to-cart events; even modestly richer feedback can change the ϵ^{-2} scaling and may reduce the calibration burden. The second is to move from single-product pricing to multi-product or inventory-constrained settings, where the "increment pooling" insight may still apply but the relevant demand object becomes multi-dimensional. Finally, strategic buyer behavior (delayed purchases, reference price effects, or learning by consumers) complicates the interpretation of $D(\delta)$ as an exogenous noise law; understanding whether intrinsic-geometry methods remain valid under such endogeneity is an open and practically important question.

Closing perspective. We view the model as illuminating a specific and consequential tradeoff: personalization is valuable, but in binary-feedback pricing it is only as feasible as the system's ability to *certify local calibration* and to exploit *structural invariances* for pooling. The intrinsic-dimension lens turns a vague intuition—that "only a few directions matter"—into a quantitative complexity measure that predicts when adaptive refinement is worth its cost. At the same time, the open problems above emphasize that the next generation of pricing theory must engage with the realities of drift, constraints, and representation error. Progress on these fronts would not

only sharpen our understanding of what is learnable, but also provide the kind of interpretable diagnostics that modern pricing systems need to be both profitable and accountable.