BEYOND SECOND-ORDER INFORMATION: PAID LOGISTIC REGRESSION WITH NOISY FEATURES

Anonymous authors
Paper under double-blind review

ABSTRACT

We consider online binary prediction with paid stochastic features, where the learner can spend resources to obtain better observations of the underlying features. This setting is motivated by applications such as medical testing, where additional cost can provide more precise measurements of patient's features for predicting a binary outcome. At each round, a latent feature is drawn i.i.d., the response is generated through a logistic model, and the learner chooses how much to pay before observing a noisy version of the feature. The goal of learner is to jointly control prediction loss and acquisition cost. In contrast to the former result on the online linear regression with paid features, covariance information alone is no longer sufficient: two noise distributions with the same covariance may induce different optimal logistic losses, and lower variance can even lead to higher loss. Under a structural ordering in which larger payments yield better observations by removing independent noise, we obtain $\widetilde{O}(\sqrt{T})$ regret when the noise structure is known and $\widetilde{O}(T^{2/3})$ regret when it is unknown, with matching lower bounds up to logarithmic factors.

1 INTRODUCTION

Information rarely comes for free. In many real-world prediction problems, a decision maker must balance the quality of information used for prediction against the cost of acquiring it. Consider a clinic that predicts each patient's probability of specific health outcome, which is often binary. Before observing a patient's features, the clinic must decide how much to spend on measuring them. More accurate tests or repeated measurements can improve prediction, but incur more cost. A similar trade-off arises when an online platform predicts whether a customer purchase a product or not: reliable customer information may improve the prediction, but acquiring it requires resources.

We study this problem as online logistic regression with paid noisy features. At each round, an individual arrives independently from a fixed but unknown population. Before observing the individual's feature, the learner chooses a payment that determines the quality of feature measurement, together with a prediction rule to apply to the resulting observation. It then observes the noisy feature, predicts a probability, and observes a binary outcome. Over the horizon, the the learner aims to learn both how much to pay and how to predict from noisy features, with the goal of minimizing cumulative logistic loss plus a payment penalty.

Several lines of work study related problems. Classical errors-in-variables models study regression with corrupted covariates, while active or costly feature acquisition problem considers which features to purchase (Janisch et al., 2019). In online learning, Cesa-Bianchi et al. (2011) studied linear prediction with noisy features and known or unknown noise covariance, but without a cost-quality trade-off. Van Der Hoeven et al. (2023) studied binary classification with paid stochastic experts, where payments improve the reliability of binary expert predictions rather than the measurement of a continuous feature vector. The closest predecessor is Merlis et al. (2026), who introduced online linear regression with paid stochastic features under squared loss. They established regret rates of $\widetilde{O}(\sqrt{T})$ and $\widetilde{O}(T^{2/3})$ when the noise covariance profile is known and unknown, respectively, with matching lower bounds up to logarithmic factors.

Moving from linear to logistic prediction is natural for binary outcomes, but introduces new statistical challenge. Related extensions have been studied in online logistic regression (Hazan et al.,

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

2014) and bandits (Filippi et al., 2010; Faury et al., 2020; Abeille et al., 2021). In our setting, the nonlinearity of the logistic function and loss changes how how noisy feature affects prediction loss. Under squared loss, the contribution of independent mean-zero feature noise is determined by its covariance. Under logistic loss, however, we show that two noise distributions with the same covariance can produce different prediction loss, and a smaller variance can even yield a larger loss. Thus, covariance alone does not determine the prediction quality of a paid measurement. Moreover, noisy feature can also change which predictor works best: a predictor that is optimal for clean features may not be optimal for noisy measurements. The learner therefore needs a prediction rule adapted to its chosen measurement quality.

Our goal is to learn a payment and prediction rule that jointly minimizes expected prediction loss and cost across the population. We measure performance by expected cumulative regret relative to this optimal population loss. Regret captures the additional population loss incurred while learning the optimal trade-off from sequential observations. This leads us to the following questions:

How quickly can a learner approach the population-wise optimal trade-off between the prediction quality and acquisition cost?

Contributions. This paper makes the following contributions.

- **Second-order information is insufficient for paid logistic prediction (Proposition 1)** We first identify a fundamental distinction from the former paid linear regression setting. While the effect of additive feature noise under squared loss can be characterized through its covariance, this is no longer true under logistic loss. We construct noise distributions with identical covariance but different optimized logistic losses. More strikingly, a noise distribution with smaller variance can induce a strictly larger optimized prediction loss. Thus, second-order information alone is in general insufficient to characterize the improvement of feature quality in logistic prediction.
- **Order-optimal regret with a known degradation increment structure.** We introduce a structural ordering of the noise in which the lower payment yields a noisier observation by adding independent degradation noise (Assumption 2). When the corresponding distribution of degradation increment is known, an observation obtained at a higher payment can be transformed into statistically valid observations at lower payment levels (Proposition 4). Exploiting this one-sided information sharing, we develop a one-sided simulation-based grid elimination algorithm achieving $\widetilde{\mathcal{O}}(\sqrt{T})$ regret (Theorem 5). We prove a matching $\Omega(\sqrt{T})$ minimax lower bound (Theorem 8), showing that the horizon dependence is optimal up to logarithmic factors.
- **Order-optimal regret with an unknown noise profile.** When the payment-dependent noise profile is unknown, observations collected at one payment level cannot in general be reused at another. We develop an asynchronous grid lower confidence bound algorithm that directly learns the prediction–payment trade-off and achieves $\widetilde{\mathcal{O}}(T^{2/3})$ regret (Theorem 10). We complement this result with an $\Omega(T^{2/3})$ lower bound (Theorem 12), even for one-dimensional Gaussian noise with a decreasing variance profile.

Notation. For a vector $x \in \mathbb{R}^d$, $\|x\|_2$ denotes its Euclidean norm. For symmetric matrices A and B, we write $A \preceq B$ when $B - A$ is positive semidefinite. We write $U \overset{d}{=} V$ to denote equality in distribution. We use $\mathcal{O}(\cdot)$ for the standard asymptotic order notation and $\widetilde{\mathcal{O}}(\cdot)$ when logarithmic factors are ignored. For a positive integer T, let $[T] = \{1, 2, \cdots, T\}$ and $[T]_0 = [T] \cup \{0\}$. For $S > 0$, define the closed Euclidean ball in $\mathbb{R}^d$ by $\mathcal{B}_d(S) = \{x \in \mathbb{R}^d : \|x\|_2 \leq S\}$.

2 Problem Setting

Let $\theta^\star \in \mathcal{B}_d(S)$ be an unknown parameter, where the radius S is known. At every round $t \in [T]$, the environment (privately) draws a feature $X_t \overset{i.i.d.}{\sim} D_X$, and generates a binary response $Y_t \mid X_t \sim \text{Bernoulli}(\mu(X_t^\top \theta^\star))$, where $\mu(z) = (1 + \exp(-z))^{-1}$. The clean feature X_t is not revealed to the learner, and its distribution D_X is fixed across rounds but unknown. The learner selects a payment $c_t \in [0, 1]$ and chooses a linear predictor $\nu_t \in \mathcal{V} = \mathcal{B}_d(S)$, then observes the corrupted

feature $\widehat{X}_t = X_t + N_t(c_t)$, where $N_t(c_t) \sim D_N(c_t)$. For any payment c, the distribution $D_N(c)$ is mean-zero and independent of (X, Y) conditionally on the payment c and previous history.

Let $\mathcal{H}_{t-1} := \{(c_s, \nu_s, \widehat{X}_s, Y_s, \Xi_s)\}_{s=1}^{t-1}$ denote the learner's history before round t, where Ξ_s is the (possibly empty) collection of all additional quantities privately generated by the learner during round s. The learner determines c_t using only $\mathcal{H}_{t-1}$, and chooses the predictor ν_t before observing the noisy feature $\widehat{X}_t$. Precisely, ν_t is measurable with respect to $\sigma(\mathcal{H}_{t-1}, c_t)$. After observing $\widehat{X}_t$, the learner predicts the probability of the response to be 1 by $\mu(\widehat{X}_t^\top \nu_t)$. The response Y_t is then revealed to the learner.

The predicted probability is evaluated through the logistic negative log-likelihood, which is the canonical loss for probabilistic binary prediction. For $y \in \{0, 1\}$ and $z \in \mathbb{R}$, the logistic loss ℓ is

$$\ell(y, z) := -y \log \mu(z) - (1 - y) \log\big(1 - \mu(z)\big) = \log\big(1 + \exp(z)\big) - yz.$$

For a fixed payment c, let $\widehat{X}(c) = X + N(c)$ be a corrupted random feature at payment c. The *population loss* of the predictor ν and the cost c, measuring the expected logistic loss plus the payment penalty, is defined as

$$L(c, \nu) = \mathbb{E}\big[\ell(Y, \widehat{X}(c)^\top \nu)\big] + \lambda c.$$

where $\lambda > 0$ is a cost sensitivity parameter that determines the relative importance of prediction accuracy and payment. Here, the expectation is taken over the population of X, measurement noise $N(c)$, and outcome Y. For each payment c, we define the *optimized population loss* at and *optimal predictor* at payment c as

$$L^\star(c) = \min_{\nu \in \mathcal{V}} L(c, \nu), \qquad \nu^\star(c) \in \arg\min_{\nu \in \mathcal{V}} L(c, \nu).$$

Although the response follows the logistic model with parameter $\theta^\star$, the learner predicts from the noisy feature $\widehat{X}(c)$. Indeed, the conditional probability given this feature is $\mathbb{P}(Y = 1|\widehat{X}) = \mathbb{E}[\mu(X^\top \theta^\star)|\widehat{X}]$, which is generally not representable as $\mu(\widehat{X}^\top \nu)$ for any ν. Thus, $\theta^\star$ need not be the best logistic-linear predictor from the noisy feature (Kim et al., 2023; Stefanski & Carroll, 1985).

We define *(global) optimal population loss* by $L^\star = \inf_{c \in [0,1]} L^\star(c)$. This is the optimal expected loss per round over all fixed payments and predictors. Since each new feature is independent of the past, and the payment and predictor are chosen before observing the feature, the conditional expected loss at round t is $L(c_t, \nu_t) \geq L^\star$. We therefore compare the learner's expected cumulative population loss relative to $TL^\star$ and define the expected regret as

$$\mathrm{Reg}_T := \mathbb{E}\left[\sum_{t=1}^{T} \big\{L(c_t, \nu_t) - L^\star\big\}\right] = \mathbb{E}\left[\sum_{t=1}^{T} \big(\ell(Y_t, \widehat{X}_t^\top \nu_t) + \lambda c_t\big)\right] - TL^\star. \tag{1}$$

The second equality follows from the tower property; see Appendix B for proof.

Throughout, we adopt the following regularity assumption on the features and measurement noises.

Assumption 1 (Sub-Gaussian feature and noise)**.** *There exists a known constant $B > 0$ such that $\|\mathbb{E}[X]\|_2 \leq B$. There is a known constant $R \geq 1$ such that, for every payment $c \in [0, 1]$ and every $u \in \mathbb{R}^d$, $\mathbb{E}\big[e^{u^\top (X_t - \mathbb{E}[X])} \mid \mathcal{H}_{t-1}\big] \leq e^{\frac{R^2 \|u\|_2^2}{2}}$ and $\mathbb{E}\big[e^{u^\top N_t(c)} \mid \mathcal{H}_{t-1}, c\big] \leq e^{\frac{R^2 \|u\|_2^2}{2}}$.*

2.1 Beyond Covariance: Structure of Noise Profile

In the linear regression with squared loss (Merlis et al., 2026), the effect of additive measurement noise is characterized entirely by its covariance. For mean-zero noise N independent of (X, Y),

$$\mathbb{E}\big[((X + N)^\top \nu - X^\top \theta^\star)^2\big] = \mathbb{E}\big[\big(X^\top(\theta^\star - \nu)\big)^2\big] + \nu^\top \mathrm{Cov}(N) \nu.$$

Hence, for a fixed predictor ν, the additional squared loss caused by measurement noise is completely captured by its covariance $\mathrm{Cov}(N)$. On the other hand, **this reduction does not extend to the logistic loss.** To see this, we expand the logistic loss $\ell(y, z)$ as

$$\ell(y, z) = \log\big(1 + \exp(z)\big) - yz = \log 2 + \big(\tfrac{1}{2} - y\big) z + \tfrac{1}{8} z^2 - \tfrac{1}{192} z^4 + O(z^6).$$

Unlike the squared loss, logistic loss contains terms of order higher than 2. Therefore, covariance alone may not be sufficient to characterize the logistic loss. The following proposition formalizes the insufficiency of the covariance: covariance ordering does not imply the logistic loss ordering.

Proposition 1 (Insufficiency of covariance)**.** *There exists one-dimensional problem instances with mean-zero scalar noises $N(c)$ such that $\mathrm{Var}(N(c_1)) = \mathrm{Var}(N(c_2))$, but the two payments have different optimized logistic losses. Moreover, there exist other instances such that $\mathrm{Var}(N(c_1)) > \mathrm{Var}(N(c_2))$ while the optimized logistic loss at c_1 is strictly smaller than at c_2.*

The proof is deferred to Appendix C. Proposition 1 shows that covariance alone is insufficient to characterize the effect of measurement noise under logistic loss: even equal variances can lead to different losses, and variance ordering may not align with the loss ordering. Therefore, we impose the following distributional assumption on the noise across payment levels:

Assumption 2 (Ordered additive degradation)**.** *For every $0 \le c_1 \le c_2 \le 1$, there exists a mean-zero random vector ξ_{c_1,c_2}, independent of $(X, Y, N(c_2))$, such that*

$$N(c_1) \overset{d}{=} N(c_2) + \xi_{c_1,c_2}.$$

We call ξ_{c_1,c_2} a *degradation increment*, or simply *increment*. Assumption 2 means that for $c_1 \le c_2$ the noise $N(c_1)$ at c_1 can be obtained by adding an independent mean-zero *increment* ξ_{c_1,c_2} to the noise $N(c_2)$.

In general, Assumption 2 is stronger than covariance monotonicity in the paid linear case (Merlis et al., 2026). However, the two conditions are equivalent for Gaussian measurement noise. In particular, suppose $N(c) \sim \mathcal{N}(0, \Sigma(c))$ and $c_1 \le c_2$ implies $\Sigma(c_2) \preceq \Sigma(c_1)$, then Assumption 2 holds by taking $\xi_{c_1,c_2} \sim \mathcal{N}(0, \Sigma(c_1) - \Sigma(c_2))$ independently of $N(c_2)$.

Assumption 2 restores the natural ordering of loss. Since $\ell(y, z)$ is convex in z, for every $c_1 \le c_2$ and $\nu \in \mathcal{V}$, Jensen's inequality conditional on $X, Y, N(c_2)$ and the fact $\mathbb{E}[\xi_{c_1,c_2}] = 0$ yields

$$\mathbb{E}\Big[\ell\big(Y, \big(X+N(c_1)\big)^\top \nu\big)\Big] = \mathbb{E}\Big[\ell\big(Y, \big(X+N(c_2)+\xi_{c_1,c_2}\big)^\top \nu\big)\Big] \ge \mathbb{E}\Big[\ell\big(Y, \big(X+N(c_2)\big)^\top \nu\big)\Big]. \quad (2)$$

That is, the prediction loss is nonincreasing in payment c. Therefore, any increase in the population loss from c_1 to c_2 can only come from the additional payment, by at most $\lambda(c_2 - c_1)$. This directly gives the following one-sided regularity property of the population loss (see Appendix D for proof).

Proposition 2 (One-sided Lipschitzness)**.** *Under Assumption 2, for every $0 \le c_1 \le c_2 \le 1$ and every $\nu \in \mathcal{V}$, $L(c_2, \nu) \le L(c_1, \nu) + \lambda(c_2 - c_1)$. Consequently, $L^\star(c_2) \le L^\star(c_1) + \lambda(c_2 - c_1)$.*

We consider two information regimes according to whether the distributions of degradation increments ξ_{c_1,c_2} in Assumption 2 are available to the learner. In the *known-profile* setting, the learner knows the distribution of every degradation increment ξ_{c_1,c_2} ($c_1 < c_2$) and has access to an exact sampler from this distribution. In the *unknown-profile* setting, Assumption 2 is known to hold, but the increment distributions of increments ξ_{c_1,c_2} are unavailable to the learner.

3 Online Logistic Regression with Known Noise Profile

In this section, we consider the *known-profile* setting. For every pair of payments $c_1 \le c_2$, the learner knows the distribution of the increment ξ_{c_1,c_2} in Assumption 2 and has an exact sampler from this distribution. Thus, after paying c_2 and observing $\widehat{X}(c_2)$, the learner can independently draw ξ_{c_1,c_2}, then $\widehat{X}(c_2) + \xi_{c_1,c_2}$ has the same distribution as a noisy feature $\widehat{X}(c_1)$ at payment c_1.

To exploit this property, we first establish a concentration bound for the empirical logistic loss when all observations are collected at a fixed payment level c. We then show that observations at adaptively chosen higher payments can be used to generate simulated samples with exactly the same distribution as the observations at a fixed lower payment. These results form the statistical basis of the elimination algorithm introduced afterward.

3.1 Uniform concentration at a fixed payment

We begin with the simplest case in which the learner pays a fixed payment $c \in [0, 1]$. In this case, the observed pairs $(\widehat{X}_s, Y_s)_{s=1}^t$ are independent and identically distributed samples of $(\widehat{X}(c), Y)$. We therefore establish a uniform concentration of empirical loss to its population counterpart $L(c, \nu)$ at round t over all $\nu \in \mathcal{V}$, so that the bound remains valid for any data-dependent predictor.

Proposition 3 (Uniform concentration at a fixed payment)**.** *Fix $c \in [0,1]$, and let $(\widehat{X}_s(c), Y_s)_{s=1}^t$ be i.i.d. samples of $(\widehat{X}(c), Y)$. Under Assumption 1, there is a universal constant $C_0 > 0$ such that, for every fixed $t \geq 1$ and every $\delta \in (0,1)$,*

$$\mathbb{P}\left(\sup_{\nu \in \mathcal{V}} \left| \frac{1}{t}\sum_{s=1}^{t} \ell\Big(Y_s, \widehat{X}_s(c)^\top \nu\Big) + \lambda c - L(c,\nu) \right| \leq \gamma(t,\delta)\right) \geq 1-\delta,$$

where $\gamma(t,\delta) = C_0 S(B+R)\sqrt{(d+\log(1/\delta))/t}$.

Proof sketch. The centered logistic loss is a contraction in prediction score $\widehat{X}(c)^\top \nu$. By Rademacher symmetrization and contraction inequality, the uniform deviation of empirical loss can be controlled by Rademacher-symmetrized feature sum $\sum_{s=1}^t \varepsilon_s \widehat{X}_s(c)$, where $(\varepsilon_s)_{s=1}^t$ are Rademacher random variables. Under Assumption 1, this sum is a sub-Gaussian random vector. Therefore, a moment bound for the Euclidean norm of sub-Gaussian vector yields the desired uniform concentration bound. The full proof is given in Appendix E. □

Proposition 3 applies when all observations at round t are collected at the same payment c. However, in the online learning setting, observations may be collected at adaptively chosen payments. We next show that high-payment observations can be transformed into a *simulated observation* at a lower target payment, and derive a corresponding concentration bound for these simulated observations.

3.2 Validity of Simulated Lower-payment Observations

At round t, the learner selects c_t and observes $\widehat{X}_t = X_t + N_t(c_t)$. For any payment $c \leq c_t$, Assumption 2 gives $N(c) \overset{d}{=} N(c_t) + \xi_{c,c_t}$. Since the distribution of $\xi_{c,c_t}^{(t)}$ is known, the learner can sample ξ_{c,c_t} independently of $(\widehat{X}_t, Y_t)$ and construct a simulated feature $\widetilde{X}_t(c) := \widehat{X}_t + \xi_{c,c_t}^{(t)}$. Conditional on the history $\mathcal{H}_{t-1}$ and the current payment c_t, the pair of simulated observation satisfies $(\widetilde{X}_t(c), Y_t) \overset{d}{=} (\widehat{X}(c), Y)$. Thus, an observation collected at the payment c_t provides a valid simulated observation for every lower payment $c \leq c_t$.

We next consider a sequence of such simulated observations for a fixed target payment c. Define $\tau(c) = \inf\{t \geq 1 : c_t < c\}$ with $\tau(c) = \infty$ if the learner never selects a payment below c. If $\tau(c) > t$, $c_1, \ldots, c_t \geq c$, so all of the first t observations can be converted into a simulated observation at target payment c. Although the payments c_t may vary and be selected adaptively, the conditional distributional equivalence allows us to inherit the result of Proposition 3 as follows:

Proposition 4 (Concentration of simulated observations)**.** *Fix $c \in [0,1]$ and let $\tau(c) = \inf\{t \geq 1 : c_t < c\}$. For each $s < \tau(c)$, independently sample $\xi_{c,c_s}^{(s)}$ and construct a simulated feature as $\widetilde{X}_s(c) = \widehat{X}_s + \xi_{c,c_s}^{(s)}$. Then, under Assumption 1 and 2, for every fixed $t \geq 1$ and $\delta \in (0,1)$,*

$$\mathbb{P}\left(t < \tau(c), \quad \sup_{\nu \in \mathcal{V}} \left| \frac{1}{t}\sum_{s=1}^{t} \ell\Big(Y_s, \widetilde{X}_s(c)^\top \nu\Big) + \lambda c - L(c,\nu) \right| > \gamma(t,\delta)\right) \leq \delta.$$

Proof sketch. The proof completes the sequence after $\tau(c)$ with independent virtual observations drawn from the distribution of $(\widehat{X}(c), Y)$. Hence, each simulated observation pair before $\tau(c)$ also has this same conditional distribution regardless of preceding history. Therefore, the completed sequence is i.i.d. with the same distribution as $(\widehat{X}(c), Y)$. Since the event that $t < \tau(c)$ and the concentration bound fails for the simulated observations is contained in the failure event for the i.i.d. sequence above, Proposition 3 bounds this probability by δ. The full proof is given in Appendix F. □

Proposition 4 shows that, as long as the first t payments are all at least c, the simulated observations $(\widetilde{X}_s, Y_s)_{s=1}^t$ behave as if they have been collected directly at c, even when the actual payments c_t are chosen adaptively. This result is a key idea behind the known-profile algorithm: an observation collected at a sufficiently high payment can be reused to calculate the empirical loss at lower payment levels, with the same concentration rate as direct observations.

Algorithm 1 One-Sided Simulation-Based Grid Elimination

Require: Horizon T, grid parameter K, confidence level δ, known increment distributions.

$\mathcal{A} \leftarrow \{0, 1, \ldots, K\}$, $n \leftarrow 0$.
Set $\widehat{\nu}_{k,0} = 0$ for all $k \in \mathcal{A}$.
while $n < T$ **do**
  $b \leftarrow \min\{\max\{1, n\}, T - n\}$ and $j \leftarrow \max \mathcal{A}$.
  for $s = 1, \ldots, b$ **do**
    Pay $c^{(j)}$, and set $\nu \leftarrow \widehat{\nu}_{j,n}$.
    Observe $\widehat{X}$, predict $\mu(\widehat{X}^\top \nu)$, and observe Y.
    for each $k \in \mathcal{A}$ **do**
      if $k = j$ **then**
        $\widetilde{X}_k \leftarrow \widehat{X}$
      else
        Draw $\xi_{c^{(k)}, c^{(j)}}$ and set $\widetilde{X}_k \leftarrow \widehat{X} + \xi_{c^{(k)}, c^{(j)}}$.
      end if
      Append $(\widetilde{X}_k, Y)$ to the data stream of payment $c^{(k)}$.
    end for
  end for
  $n \leftarrow n + b$.
  if $n < T$ **then**
    for each $k \in \mathcal{A}$ **do**
      Compute $\widehat{\nu}_{k,n}$ and $\widehat{L}^\star_{k,n}$.
    end for
    $U \leftarrow \min_{k \in \mathcal{A}} \{\widehat{L}^\star_{k,n} + \gamma(n, \delta_T)\}$.
    $\mathcal{A} \leftarrow \{k \in \mathcal{A} : \widehat{L}^\star_{k,n} - \gamma(n, \delta_T) \le U\}$.
  end if
end while

3.3 One-Sided Simulation-Based Grid Elimination Algorithm

The algorithm discretizes the continuous payment interval $[0, 1]$ into the uniform grid $c^{(k)} = k/K$, $k \in [K]_0$, and maintains an active set $\mathcal{A} \subseteq [K]_0$ of payment indices that have not yet been ruled out as optimal. The algorithm works in sequential batches. At the beginning of each batch, it selects the largest active index $j = \max \mathcal{A}$ from the active set and keeps paying $c^{(j)}$ throughout the batch. Since every active index $k \in \mathcal{A}$ satisfies $c^{(k)} \le c^{(j)}$, each observation at $c^{(j)}$ can be converted into one valid simulated observation for every active payment $c^{(k)}$. Thus, all active payments accumulate observations at the same rate and shares a common sample count. Each batch ends when this sample count reaches the next power of 2 or T. Unless the batch terminates at T, empirical losses and predictors for each $k \in \mathcal{A}$ are updated, and we call the corresponding sample count a **checkpoint**. The pseudocode of algorithm is given in Algorithm 1.

Suppose a batch after checkpoint n begins with active set $\mathcal{A}$. At each round in the batch, the learner pays $c^{(j)}$ with $j = \max \mathcal{A}$ and uses the predictor $\hat{\nu}_{j,n}$ computed at the checkpoint. After observing $(\widehat{X}_s, Y_s)$ at round s, for every active $k < j$, the learner independently samples $\xi^{(s)}_{c^{(k)}, c^{(j)}}$ and constructs $\widetilde{X}_{k,s} = \widehat{X}_s + \xi^{(s)}_{c^{(k)}, c^{(j)}}$; while $\widetilde{X}_{j,s} = \widehat{X}_s$. The pair $(\widetilde{X}_{k,s}, Y_s)$ is then appended to the data stream associated with payment $c^{(k)}$.

At the end of the batch after checkpoint $n \ge 1$ that does not terminate at T, the new common sample count is $m = 2n$. For each active payment $c^{(k)}$, define the empirical losses and predictors as

$$\widehat{L}_{k,m}(\nu) := \frac{1}{m} \sum_{s=1}^{m} \ell\big(Y_s, \widetilde{X}_{k,s}^\top \nu\big) + \lambda c^{(k)}, \quad \widehat{\nu}_{k,m} \in \arg\min_{\nu \in \mathcal{V}} \widehat{L}_{k,m}(\nu);$$

and denote optimized empirical loss by $\widehat{L}^\star_{k,m} := \widehat{L}_{k,m}(\widehat{\nu}_{k,m})$. The algorithm compares the active payments using confidence intervals of optimized empirical losses. We set $\delta_T = \delta / \big((K + 1)\lceil \log_2 T \rceil\big)$ and define $U := \min_{k \in \mathcal{A}} \{\widehat{L}^\star_{k,m} + \gamma(m, \delta_T)\}$. A payment $c^{(k)}$ remains active if

$\widehat{L}^{\star}_{k,m} - \gamma(m, \delta_T) \le U$ and is eliminated otherwise. That is, $c^{(k)}$ is eliminated if its optimistic loss exceeds the smallest pessimistic loss among the active payments.

We next show that shared simulated observations, together with elimination, achieves $\widetilde{O}(\sqrt{T})$ regret.

Theorem 5 (Regret bound of Algorithm 1)**.** *Suppose Assumptions 1 and 2 hold and the degradation increment distributions are known to the learner. For every $T \ge 2$, $K \ge 1$, and $\delta \in (0,1)$, Algorithm 1 satisfies*

$$\mathrm{Reg}_T \le C_1 S(B+R)\sqrt{TD_T} + \frac{\lambda T}{K} + \log 2 + \lambda + \delta T\big[\log 2 + S(B+2R) + \lambda\big],$$

where $C_1 > 0$ is a universal constant and $D_T = d + \log(1/\delta_T)$.

Proof sketch. We define a good event $\mathcal{G}$ on which uniform concentration holds simultaneously for all grid payment and checkpoints. By Proposition 4 and a union bound, $\mathbb{P}(\mathcal{G}) \ge 1-\delta$. On $\mathcal{G}$, the best grid payment $c^{(k^\star)}$ is never eliminated, and every surviving payment $c^{(k)}$ satisfies $L(c^{(k)}, \widehat{\nu}_{k,n}) - L^\star(c^{(k^\star)}) = O(\gamma(n, \delta_T))$. Therefore, during the batch following checkpoint n, the instantaneous regret relative to $L^\star(c^{(k^\star)})$ is $O(\gamma(n, \delta_T))$. Since this batch has at most n rounds, regret during this batch is $O(n\gamma(n, \delta_T)) = O(S(B+R)\sqrt{nD_T})$. Since the sample count doubles after batches, summing over batches yields regret of $O(S(B+R)\sqrt{TD_T})$. The first round contributes at most $\log 2 + \lambda$. Finally, Proposition 2 implies $L^\star(c^{(k^\star)}) - L^\star \le \lambda/K$, so discretization contributes at most $\lambda T/K$ to the regret. On $\mathcal{G}^c$, the regret is controlled by the uniform per-round loss bound. Combining these terms gives the result. The full proof is given in Appendix G □

The leading terms in the regret bound are the statistical error $S(B+R)\sqrt{TD_T}$ and the discretization error $\lambda T/K$. Balancing these two terms suggests choosing $K \asymp \frac{\lambda\sqrt{T}}{S(B+R)\sqrt{D_T}}$. Taking $\delta = 1/T$ makes the contribution of failure event $O(1)$. Since D_T depends only logarithmically on T, K and $1/\delta$, $D_T = \widetilde{O}(d)$. Therefore, by treating S, B, R, and $\lambda > 0$ as fixed constants, the expected regret satisfies $\mathrm{Reg}_T = \widetilde{O}(\sqrt{dT})$.

Remark 6. *The simulated observations across different payments are generally dependent since they are generated from the same actual observation. This dependence does not affect the analysis since Proposition 4 is applied separately to each payment $c^{(k)}$, and their concentration events are combined only through a union bound, which does not require independence across payments.*

Remark 7. *A key advantage of the known-profile setting is that the statistical error does not scale polynomially with the number of grid points K. Since every observation collected at the largest active payment can be degraded to every active lower payment, all active payments are learned simultaneously. Consequently, increasing K primarily increases computation cost and only logarithmically affects the statistical error, while reducing the approximation error caused by discretization.*

We now show that our algorithm is minimax optimal in its dependence on T, up to logarithmic factor, by establishing a matching $\Omega(\sqrt{T})$ minimax lower bound for one-dimensional instances (see Appendix I.2 for full proof and construction).

Theorem 8 (Minimax lower bound in known-profile setting)**.** *There exist constants S, B, R, λ and $C_k > 0$ such that, for all sufficiently large T, every possibly randomized online algorithm incurs*

$$\mathrm{Reg}_T \ge C_k\sqrt{T}$$

on some one-dimensional instance with a known noise profile, where Assumptions 1 and 2 hold.

Remark 9. *In classical linear and logistic regression tasks, logarithmic regret in T is achievable in both settings (Vovk, 1997; Hazan et al., 2014). In the paid regression setting, Merlis et al. (2026) showed that introducing a payment–information trade-off fundamentally changes the picture even for linear regression, yielding an $\Omega(\sqrt{T})$ lower bound. Theorem 8 estalishes the same $\Omega(\sqrt{T})$ lower bound for paid logistic regression.*

4 Extension to Unknown Noise Profile

In the *unknown-profile* setting, Assumption 2 is known to hold, but neither the distribution of increment ξ_{c_1,c_2} nor its sampler is available to the learner. Therefore, an observation obtained at one

Algorithm 2 Asynchronous Grid Lower Confidence Bound

Require: Horizon T, grid parameter K, confidence level δ.

for $k = 0, 1, \ldots, K$ **do**
  Pay $c^{(k)}$, predict $1/2$ ($\nu = 0$), observe $(\widehat{X}, Y)$, and store the sample for $c^{(k)}$.
  Set $n(k) \leftarrow 1$, $m(k) \leftarrow 1$.
  Compute $\widehat{\nu}_{k,1}$ and $\widehat{L}^{\star}_{k,1}$.
end for
for $t = K+2, \ldots, T$ **do**
  Choose $k_t \in \arg\min_{k \in [K]_0} \{\widehat{L}^{\star}_{k,m(k)} - \gamma(m(k), \delta_T)\}$.
  Pay $c^{(k_t)}$, and set $\nu_t \leftarrow \widehat{\nu}_{k_t, m(k_t)}$.
  Observe $\widehat{X}_t$, predict $\mu(\widehat{X}_t^{\top} \nu_t)$, and observe Y_t
  Append $(\widehat{X}_t, Y_t)$ to the data stream of $c^{(k_t)}$ and set $n(k_t) \leftarrow n(k_t) + 1$.
  if $n(k_t) = 2m(k_t)$ **then**
    $m(k_t) \leftarrow 2m(k_t)$.
    Recompute $\widehat{\nu}_{k_t, m(k_t)}$ and $\widehat{L}^{\star}_{k_t, m(k_t)}$.
  end if
end for

payment cannot be converted into a valid observation at another payment. Unlike the known-profile setting in Section 3, each grid payment must be learned from its own observations.

4.1 Asynchronous Grid Lower Confidence Bound Algorithm

The algorithm uses the same grid as in Section 3. For each grid payment $c^{(k)} = k/K$, $k \in [K]_0$, let $n_t(k) := \sum_{s=1}^{t} \mathbb{1}[k_s = k]$ be the number of rounds the payment $c^{(k)}$ has been selected up to round t. The algorithm maintains a separate data stream for each grid and stores all actual observations at that payment. Consequently, the sample counts $n_t(k)$ evolve asynchronously across payments.

The algorithm first selects every grid payments once and uses $\nu_t = 0$. At round t, the empirical loss and minimizer for payment $c^{(k)}$ is recomputed whenever its sample count $n_t(k)$ reaches a power of 2. We call this sample count a **local checkpoint** of payment $c^{(k)}$ and denote its latest local checkpoint before round t by $m_t(k) := 2^{\lfloor \log_2 n_{t-1}(k) \rfloor}$, then $m_t(k) \le n_{t-1}(k) < 2m_t(k)$.

Let $(\widehat{X}_{k,s}, Y_{k,s})_{s=1}^{m}$ denote the observations collected at $c^{(k)}$ up to its local checkpoint m , then the empirical losses and predictors are defined as

$$\widehat{L}_{k,m}(\nu) := \frac{1}{m} \sum_{s=1}^{m} \ell\big(Y_{k,s}, \widehat{X}_{k,s}^{\top} \nu\big) + \lambda c^{(k)}, \quad \widehat{\nu}_{k,m} \in \arg\min_{\nu \in \mathcal{V}} \widehat{L}_{k,m}(\nu);$$

and denote optimized empirical losses by $\widehat{L}^{\star}_{k,m} := \widehat{L}_{k,m}(\widehat{\nu}_{k,m})$. At each round $t \ge K+2$, the algorithm optimistically selects $k_t \in \arg\min_{k \in [K]_0} \{\widehat{L}^{\star}_{k,m_t(k)} - \gamma(m_t(k), \delta_T)\}$, then pays $c^{(k_t)}$ and uses the predictor $\nu_t = \widehat{\nu}_{k_t, m_t(k_t)}$. After observing $(\widehat{X}_t, Y_t)$, the pair is appended to the data stream of payment $c^{(k_t)}$. The pseudocode is given in Algorithm 2.

We show that this asynchronous lower-confidence bound algorithm achieves $\widetilde{O}(T^{2/3})$ regret.

Theorem 10 (Regret bound of Algorithm 2)**.** *Suppose Assumptions 1 and 2 hold. For $T \ge 2$, integer $1 \le K < T$, and $\delta \in (0,1)$, the Algorithm 2 satisfies*

$$\mathrm{Reg}_T \le C_2 S(B+R)\sqrt{(K+1)TD_T} + \frac{\lambda T}{K} + (K+1)(\log 2 + \lambda) + \delta T\big[\log 2 + S(B+2R) + \lambda\big],$$

where $C_2 > 0$ is a universal constant and $D_T = d + \log(K+1) + \log\lceil \log_2 T \rceil + \log(1/\delta)$.

Proof sketch. We define a good event $\mathcal{G}$ on which uniform concentration holds simultaneously for all grid payments and all possible local checkpoints. On $\mathcal{G}$, the selection rule implies that the payment

$c^{(k_t)}$ and predictor $\widehat{\nu}_{k_t,m_t(k_t)}$ satisfy $L(c^{(k_t)}, \widehat{\nu}_{k_t,m_t(k_t)}) - L^\star(c^{(k^\star)}) = O(\gamma(m_t(k_t), \delta_T))$. We sum this regret separately for each payment over all possible batches. For each payment $c^{(k)}$, the cumulative regret is $O(S(B+R)\sqrt{n_T(k)D_T})$. Summing over all $K+1$ grid payments and Cauchy–Schwarz inequality yield the regret of $O(S(B+R)\sqrt{(K+1)TD_T})$. The remaining steps are the same as in the proof of Theorem 5. The full proof is given in Appendix H. □

The statistical error $S(B+R)\sqrt{KTD_T}$ and discretization error $\lambda T/K$ are balanced to $\lambda^{1/3}S^{2/3}(B+R)^{2/3}D_T^{1/3}T^{2/3}$ by setting $K \asymp \left(\frac{\lambda\sqrt{T}}{S(B+R)\sqrt{D_T}}\right)^{2/3}$. By taking $\delta = 1/T$ and treating S, B, R, and $\lambda > 0$ as fixed constants, we obtain $\mathrm{Reg}_T = \widetilde{O}(d^{1/3}T^{2/3})$.

Remark 11. *A key difference from the known-profile setting is that observations collected at one payment cannot be converted into valid observations at another payment, so each grid payment must be learned only from its own observations. This leads to the additional $\sqrt{K}$ dependence in the statistical error: increasing K improves the approximation over grid, but it increases the number of payments that must be learned separately. It also introduces an additional dependence of regret on λ. In the known-profile setting, the effect of λ can be offset by the choice of K without incurring a polynomial statistical cost. In the unknown-profile setting, however, increasing K also increases the statistical error polynomially, and this trade-off leaves the factor $\lambda^{1/3}$ in the regret bound.*

Without knowledge of the noise profile, attaining the same regret rate as in the known-profile setting is no longer possible. The following lower bound shows that the $\widetilde{O}(T^{2/3})$ regret achieved above is minimax optimal in its dependence on T, up to logarithmic factors.

Theorem 12 (Minimax lower bound in unknown-profile setting)**.** *There exist constants S, B, R, λ and $C_u > 0$ such that, for all sufficiently large T, every randomized online algorithm incurs*

$$\mathrm{Reg}_T \geq C_u T^{2/3}$$

on some one-dimensional instance with an unknown noise profile, where Assumptions 1 and 2 hold.

Remark 13. *The high-level construction is inspired by the unknown-profile lower bound of Merlis et al. (2026), but its realization under logistic loss requires an additional step. Unlike the squared-loss setting of Merlis et al. (2026), the optimized logistic loss has no such closed-form representation. We therefore first establish strict monotonicity and regularity of the scalar Gaussian logistic loss, together with a Lipschitz inverse, and use this inverse to construct noise-variance profiles that realize the desired flat baseline and localized loss perturbations.*

5 Conclusion

We studied online logistic regression with paid noisy features, where the learner pays to acquire higher–quality observations. Unlike the paid linear regression with squared loss, we showed that second-order information alone is insufficient to characterize the effect of measurement noise: even equal noise variance can yield different optimized logistic losses, and covariance ordering may not imply the loss ordering (Proposition 1). To recover the monotonicity of prediction loss in payment, we introduced an ordered additive degradation condition (Assumption 2) under which higher payments returns more informative observations.

Under this structure, we characterized the effect of knowing the noise profile. When the distributions of degradation increments are known, observations collected at higher payments can be exploited to simulate observations at lower payment levels (Proposition 4), enabling a simulation-based elimination algorithm to achieve $\widetilde{O}(\sqrt{dT})$ regret (Theorem 5). When the profile is unknown, such sample simulation is unavailable and each payment level must be learned separately, leading to $\widetilde{O}(T^{2/3})$ regret (Theorem 10). Matching lower bounds (Theorems 8 and 12) show that they are optimal in dependence on horizon T up to logarithmic factor.

Our results rely on the ordered additive degradation assumption (Assumption 2), which provides a tractable ordering of information quality but is generally stronger than covariance ordering (Merlis et al., 2026) . An imporatant direction of future work is to relax these structure, for example by considering weaker conditions or partially known noise profiles. Other natural extensions include generalized linear models, costly information acquisition in multi-armed bandits, strategic bandits, active learning, and preference learning.

AI USE STATEMENT

In this work, we used generative AI tools to help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, assist with translation. We have not used generative AI tools to interpret results, and generating synthetic data sets; cleaning and reformatting dataset; supporting qualitative and thematic data analysis are not applicable to this work. Additionally, we used generative AI tools to draft parts of a research paper, summarize or analyse existing literature, discover research topics or identify gaps, edit a research paper to improve readability, identify relevant literature, format references, suggest a structure for a research paper, propose a title or keywords for a research paper. We have reviewed all AI-assisted work: all LLM-generated mathematical statements and proofs are verified by at least two authors, all LLM-generated draft and expression are polished for readability, and all references suggested by LLM are manually checked. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

This work is theoretical and does not use external datasets or empirical experiments.

The problem setting and assumptions are fully specified in Section 2, including the sub-Gaussian condition in Assumption 1 and the ordered additive degradation structure in Assumption 2. Proposition 1 (see Appendix C for proof) motivates the latter by showing the insufficiency of covariance information, while Proposition 2 (Appendix D) is its direct result. For the known-profile setting (Section 3), Propositions 3 (Appendix E) and 4 (Appendix F) provide the concentration results underlying Algorithm 1, whose regret guarantee is stated in Theorem 5 (Appendix G). The unknown-profile procedure is given in Algorithm 2 (Section 4), with its regret guarantee stated in Theorem 10 (Appendix H). The lower bounds (Theorems 8 and 12) for them are proved in Appendices I.2 and I.3, respectively.

REFERENCES

Marc Abeille, Louis Faury, and Clément Calauzènes. Instance-wise minimax-optimal algorithms for logistic bandits. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 3691–3699, 2021.

Anish Agarwal, Devavrat Shah, Dennis Shen, and Dogyoon Song. On robustness of principal component regression. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Anish Agarwal, Keegan Harris, Justin Whitehouse, and Steven Z. Wu. Adaptive principal component regression with applications to panel data. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Ameya Anjarlekar, Rasoul Etesami, and R. Srikant. Online mechanism design for differentially private data acquisition. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, pp. 4500–4505. IEEE, 2024. doi: 10.1109/CDC56724.2024.10886292.

Katy S. Azoury and Manfred K. Warmuth. Relative loss bounds for on-line density estimation with the exponential family of distributions. *Machine Learning*, 43(3):211–246, 2001.

Alexandre Belloni, Mathieu Rosenbaum, and Alexandre B. Tsybakov. Linear and conic programming estimators in high-dimensional errors-in-variables models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):939–956, 2017. doi: 10.1111/rssb.12196.

Jeffrey S. Buzas. A note on corrected scores for logistic regression. *Statistics & Probability Letters*, 79(22):2351–2358, 2009. doi: 10.1016/j.spl.2009.08.005.

Zhiqiang Cao, Man Yu Wong, and Garvin H. L. Cheng. Logistic regression with correlated measurement error and misclassification in covariates. *Statistical Methods in Medical Research*, 32 (4):789–805, 2023. doi: 10.1177/09622802231154324.

Raymond J. Carroll and M. P. Wand. Semiparametric estimation in logistic measurement error models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):573–585, 1991. doi: 10.1111/j.2517-6161.1991.tb01846.x.

Raymond J. Carroll, Clifford H. Spiegelman, K. K. Gordon Lan, Kent T. Bailey, and Robert D. Abbott. On errors-in-variables for binary regression models. *Biometrika*, 71(1):19–25, 1984. doi: 10.1093/biomet/71.1.19.

Raymond J. Carroll, David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman & Hall/CRC, 2 edition, 2006.

Nicolò Cesa-Bianchi, Shai Shalev-Shwartz, and Ohad Shamir. Online learning of noisy data. *IEEE Transactions on Information Theory*, 57(12):7907–7931, 2011. doi: 10.1109/TIT.2011.2164053.

Yudong Chen and Constantine Caramanis. Noisy and missing data regression: Distribution-oblivious support recovery. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 383–391, 2013.

Abhirup Datta and Hui Zou. CoCoLasso for high-dimensional error-in-variables regression. *The Annals of Statistics*, 45(6):2400–2426, 2017. doi: 10.1214/16-AOS1527.

Alireza Fallah, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar. Optimal and differentially private data acquisition: Central and local mechanisms. *Operations Research*, 72(3): 1105–1123, 2024. doi: 10.1287/opre.2022.0014.

Louis Faury, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. Improved optimistic algorithms for logistic bandits. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 3052–3060, 2020.

Sarah Filippi, Olivier Cappé, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems*, volume 23, pp. 586–594, 2010.

Dylan J. Foster, Satyen Kale, Haipeng Luo, Mehryar Mohri, and Karthik Sridharan. Logistic regression: The importance of being improper. In *Proceedings of the 31st Conference on Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 167–208, 2018.

Wayne A. Fuller. *Measurement Error Models*. John Wiley & Sons, 2009. ISBN 9780470317334.

Arpita Ghosh and Aaron Roth. Selling privacy at auction. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, pp. 199–208, 2011. doi: 10.1145/1993574.1993605.

Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2–3):169–192, 2007. doi: 10.1007/s10994-007-5016-8.

Elad Hazan, Tomer Koren, and Kfir Y. Levy. Logistic regression: Tight bounds for stochastic and online optimization. In *Proceedings of the 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pp. 197–209, 2014.

Justin Hsu, Marco Gaboardi, Andreas Haeberlen, Sanjeev Khanna, Arjun Narayan, Benjamin C. Pierce, and Aaron Roth. Differential privacy: An economic method for choosing epsilon. In *2014 IEEE 27th Computer Security Foundations Symposium*, pp. 398–410. IEEE, 2014. doi: 10.1109/CSF.2014.35.

Jaromír Janisch, Tomáš Pevný, and Viliam Lisý. Classification with costly features using deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 3959–3966, 2019. doi: 10.1609/aaai.v33i01.33013959.

Abhishek Kaul and Hira L. Koul. Weighted ℓ_1-penalized corrected quantile regression for high dimensional measurement error models. *Journal of Multivariate Analysis*, 140:72–91, 2015. doi: 10.1016/j.jmva.2015.04.009.

Jung-Hun Kim, Se-Young Yun, Minchan Jeong, Junhyun Nam, Jinwoo Shin, and Richard Combes. Contextual linear bandits under noisy features: Towards bayesian oracles. In *International Conference on Artificial Intelligence and Statistics*, pp. 1624–1645. PMLR, 2023.

Jouni Kuha. Corrections for exposure measurement error in logistic regression models with an application to nutritional data. *Statistics in Medicine*, 13(11):1135–1148, 1994. doi: 10.1002/sim.4780131105.

Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*, volume 23. Springer, 1991.

Mushan Li and Yanyuan Ma. An update on measurement error modeling. *Annual Review of Statistics and Its Application*, 11:279–296, 2024. doi: 10.1146/annurev-statistics-040722-043616.

Po-Ling Loh and Martin J. Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. In *Advances in Neural Information Processing Systems*, volume 24, 2011.

Nadav Merlis, Kyoungseok Jang, and Nicolò Cesa-Bianchi. Online linear regression with paid stochastic features. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 24370–24378, 2026. doi: 10.1609/aaai.v40i29.39618.

Tsuyoshi Nakamura. Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77(1):127–137, 1990. doi: 10.1093/biomet/77.1.127.

Kobbi Nissim, Claudio Orlandi, and Rann Smorodinsky. Privacy-aware mechanism design. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 774–789, 2012. doi: 10.1145/2229012.2229073.

Kobbi Nissim, Salil Vadhan, and David Xiao. Redrawing the boundaries on purchasing data from privacy-sensitive individuals. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, pp. 411–422, 2014. doi: 10.1145/2554797.2554835.

Hyung Park and Seonjoo Lee. Logistic regression error-in-covariate models for longitudinal high-dimensional covariates. *Stat*, 8(1):e246, 2019. doi: 10.1002/sta4.246.

Sophia Rabe-Hesketh, Andrew Pickles, and Anders Skrondal. Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation. *Statistical Modelling*, 3(3):215–232, 2003. doi: 10.1191/1471082X03st056oa.

Mathieu Rosenbaum and Alexandre B. Tsybakov. Sparse recovery under matrix uncertainty. *The Annals of Statistics*, 38(5):2620–2651, 2010. doi: 10.1214/10-AOS793.

Mathieu Rosenbaum and Alexandre B. Tsybakov. Improved matrix uncertainty selector. In *From Probability to Statistics and Back: High-Dimensional Models and Processes—A Festschrift in Honor of Jon A. Wellner*, volume 9 of *IMS Collections*, pp. 276–290. Institute of Mathematical Statistics, 2013. doi: 10.1214/12-IMSCOLL920.

Bernard Rosner, Walter C. Willett, and Donna Spiegelman. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*, 8(9):1051–1069, 1989. doi: 10.1002/sim.4780080905.

Maytal Saar-Tsechansky, Prem Melville, and Foster Provost. Active feature-value acquisition. *Management Science*, 55(4):664–684, 2009. doi: 10.1287/mnsc.1080.0952.

Gil I. Shamir. Logistic regression regret: What's the catch? In *Proceedings of the 33rd Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 3296–3319, 2020.

Xiao Song and Ching-Yun Wang. GMM nonparametric correction methods for logistic regression with error contaminated covariates and partially observed instrumental variables. *Scandinavian Journal of Statistics*, 46(3):898–919, 2019. doi: 10.1111/sjos.12364.

Øystein Sørensen, Kristoffer Herland Hellton, Arnoldo Frigessi, and Magne Thoresen. Covariate selection in high-dimensional generalized linear models with measurement error. *Journal of Computational and Graphical Statistics*, 27(4):739–749, 2018. doi: 10.1080/10618600.2018.1425626.

Leonard A Stefanski and Raymond J Carroll. Covariate measurement error in logistic regression. *The annals of statistics*, pp. 1335–1351, 1985.

Leonard A. Stefanski and Raymond J. Carroll. Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, 74(4):703–716, 1987. doi: 10.1093/biomet/74.4.703.

Yan Shuo Tan and Roman Vershynin. Polynomial time and sample complexity for non-gaussian component analysis: Spectral methods. In *Conference On Learning Theory*, pp. 498–534. PMLR, 2018.

Dirk Van Der Hoeven, Ciara Pike-Burke, Hao Qiu, and Nicolò Cesa-Bianchi. Trading-off payments and accuracy in online classification with paid stochastic experts. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 34809–34830, 2023.

Kabir Aladin Verchand and Andrea Montanari. High-dimensional logistic regression with missing data: Imputation, regularization, and universality. *arXiv preprint arXiv:2410.01093*, 2024.

Volodya Vovk. Competitive on-line linear regression. In *Advances in Neural Information Processing Systems*, volume 10, pp. 364–370, 1997.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

C. Y. Wang and Suojin Wang. Semiparametric methods in logistic regression with measurement error. *Statistica Sinica*, 7(4):1103–1120, 1997.

Weina Wang, Lei Ying, and Junshan Zhang. The value of privacy: Strategic data subjects, incentive mechanisms, and fundamental limits. *ACM Transactions on Economics and Computation*, 6(2): 8:1–8:26, 2018. doi: 10.1145/3232863.

Kai Zheng, Tianle Cai, Weiran Huang, Zhenguo Li, and Liwei Wang. Locally differentially private (contextual) bandits learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

CONTENTS

1 Introduction **1**

2 Problem Setting **2**

2.1 Beyond Covariance: Structure of Noise Profile 3

3 Online Logistic Regression with Known Noise Profile **4**

3.1 Uniform concentration at a fixed payment 4

3.2 Validity of Simulated Lower-payment Observations 5

3.3 One-Sided Simulation-Based Grid Elimination Algorithm 6

4 Extension to Unknown Noise Profile **7**

4.1 Asynchronous Grid Lower Confidence Bound Algorithm 8

5 Conclusion **9**

A Related Work **15**

B Proof of Eq. (1) **16**

C Proof of Proposition 1 **17**

D Proof of Proposition 2 **18**

E Proof of Proposition 3 **19**

E.1 Supporting Lemmas 21

F Proof of Proposition 4 **22**

G Proof of Theorem 5 **23**

H Proof of Theorem 10 **27**

I Proof of lower bounds **30**

I.1 Regularity of the scalar Gaussian logistic loss 30

I.2 Proof of Theorem 8: known profile 32

I.3 Proof of Theorem 12: unknown profile 34

APPENDIX

A RELATED WORK

Errors-in-variables models. Errors-in-variables (EIV) models have a long history in statistics. Much of the modern theory for linear EIV focuses on parameter estimation in high-dimensional settings under structural assumptions such as sparsity or low rank (Rosenbaum & Tsybakov, 2010; Loh & Wainwright, 2011; Chen & Caramanis, 2013; Rosenbaum & Tsybakov, 2013; Kaul & Koul, 2015; Belloni et al., 2017; Datta & Zou, 2017); see Fuller (2009); Carroll et al. (2006); Li & Ma (2024) for general treatments and recent reviews. Related work has also considered adaptively selected covariates and prediction with corrupted or missing designs (Agarwal et al., 2019; 2023).

More directly related to our setting is the literature on nonlinear EIV models, and in particular binary and logistic regression with mismeasured covariates. Early work showed that covariate measurement error can substantially distort binary-response models and lead to biased logistic-regression estimates (Carroll et al., 1984; Stefanski & Carroll, 1985). This motivated conditional- and corrected-score methods for logistic and, more generally, generalized linear measurement-error models (Stefanski & Carroll, 1987; Nakamura, 1990), as well as likelihood, calibration, and semiparametric approaches (Rosner et al., 1989; Carroll & Wand, 1991; Kuha, 1994; Wang & Wang, 1997). Later work developed nonparametric and approximate corrected-score methods (Rabe-Hesketh et al., 2003; Buzas, 2009), and extended measurement-error corrections to high-dimensional generalized linear and logistic models (Sørensen et al., 2018; Park & Lee, 2019). More recent work studies logistic models with instrumental variables, correlated measurement errors, or high-dimensional noisy and missing covariates (Song & Wang, 2019; Cao et al., 2023; Verchand & Montanari, 2024). These works primarily aim to recover or perform inference on the regression relationship defined by the latent clean covariates. Our objective is different: prediction is performed using the corrupted feature itself, whose quality is controlled through a payment, and hence the optimal logistic-linear predictor need not coincide with the parameter generating the clean responses.

Online learning with noisy and costly information. Noisy or corrupted inputs have also been studied in online learning, including online prediction with randomly corrupted examples (Cesa-Bianchi et al., 2011) and contextual bandits with noisy or privatized contexts (Zheng et al., 2020; Kim et al., 2023). A particularly relevant starting point for the explicit cost–information trade-off is Van Der Hoeven et al. (2023), who study online binary classification with paid stochastic experts, where larger payments improve the reliability of the observed experts. Their setting differs from ours in that information is provided by finitely many binary experts with independently corrupted predictions, whereas we observe a continuous noisy feature vector and perform logistic-linear prediction.

The closest predecessor of our work is Merlis et al. (2026), who introduced online linear regression with paid stochastic features. Their model studies essentially the same interaction between feature quality and acquisition cost under squared loss, and establishes optimal horizon dependences of $\widetilde{\Theta}(\sqrt{T})$ and $\widetilde{\Theta}(T^{2/3})$ when the payment–noise profile is known and unknown, respectively. Moving from squared to logistic loss, however, fundamentally changes how feature noise affects prediction: in the linear case its effect is characterized by the noise covariance, whereas for logistic loss covariance alone is insufficient and even covariance ordering need not preserve risk ordering. This motivates the stronger distributional degradation structure used in our analysis and leads to different information-sharing arguments, despite the same optimal dependence on the horizon.

Privacy, incentives, and data acquisition. The possibility of paying for more informative data is closely connected to mechanism design for privacy-sensitive data acquisition. A substantial literature studies how an analyst should compensate individuals for privacy loss while achieving a desired estimation accuracy or respecting a fixed acquisition budget (Ghosh & Roth, 2011; Nissim et al., 2012; 2014; Wang et al., 2018; Hsu et al., 2014; Fallah et al., 2024). In these models, agents may have heterogeneous and privately known valuations for privacy, so the central issue is often incentive compatibility and the joint design of payments and privacy guarantees. Recent work has also considered sequential mechanisms for acquiring differentially private data specifically for training logistic regression models (Anjarlekar et al., 2024). Our setting abstracts away strategic reporting and mechanism-design constraints: the payment deterministically indexes a noise distribution, and

the learner instead faces the statistical problem of choosing the information quality that optimally balances predictive loss and acquisition cost at each round.

Cost-sensitive and active feature acquisition. Our formulation is also related to the literature on costly or active feature acquisition, where a learner decides which information to obtain while trading predictive performance against acquisition cost. Early work on active feature-value acquisition selects missing feature values according to their expected value for improving a predictive model (Saar-Tsechansky et al., 2009). More recent approaches formulate classification with costly features as a sequential decision problem and learn instance-dependent acquisition policies, for example using reinforcement learning (Janisch et al., 2019). These models generally decide which coordinates or tests to acquire, rather than how accurately a common latent feature vector should be measured. Moreover, much of this literature focuses on batch prediction or learned acquisition policies, whereas our objective is an online regret guarantee relative to the optimal fixed payment–predictor pair. Nevertheless, both settings capture the same broad principle that predictive information has an endogenous acquisition cost.

Online logistic regression and generalized linear prediction. Finally, our work builds on the literature on online logistic regression and sequential prediction with log loss. Relative-loss guarantees for exponential-family prediction already appear in early work on online density estimation (Azoury & Warmuth, 2001), while logarithmic regret for suitable curved losses follows from second-order methods for online convex optimization (Hazan et al., 2007). For logistic regression, Hazan et al. (2014) studied tight stochastic and online optimization bounds and highlighted the unfavorable dependence of proper algorithms on the radius of the parameter set. Foster et al. (2018) showed that improper prediction can substantially improve this dependence by exploiting the mixability of logistic loss, and Shamir (2020) further characterized minimax logistic-regression regret across different norm constraints and dimensional regimes. Logistic models have also been extensively studied in the bandit setting, beginning with generalized linear bandits (Filippi et al., 2010) and followed by work identifying logistic-specific curvature and instance-dependent phenomena (Faury et al., 2020; Abeille et al., 2021). These works assume that the covariates entering the logistic predictor are observed at a fixed quality. Our setting introduces an additional online decision—how much to pay for the quality of those covariates—and shows that this acquisition decision fundamentally changes the attainable dependence on the horizon.

B Proof of Eq. (1)

Since c_t and ν_t are chosen before observing $(\widehat{X}_t, Y_t)$,

$$\mathbb{E}\left[\ell(Y_t, \widehat{X}_t^\top \nu_t) + \lambda c_t \middle| \mathcal{H}_{t-1}, c_t, \nu_t\right] = L(c_t, \nu_t).$$

Hence, by the tower property,

$$\mathbb{E}\left[\ell(Y_t, \widehat{X}_t^\top \nu_t) + \lambda c_t\right] = \mathbb{E}\left[L(c_t, \nu_t)\right],$$

and therefore,

$$\mathrm{Reg}_T = \mathbb{E}\left[\sum_{t=1}^{T} \left(\ell(Y_t, \widehat{X}_t^\top \nu_t) + \lambda c_t\right)\right] - TL^\star = \mathbb{E}\left[\sum_{t=1}^{T} \{L(c_t, \nu_t) - L^\star\}\right].$$

C Proof of Proposition 1

Proof. Let $d = 1$, $X = 1$, and $0 < \theta^\star < S$, then $Y \sim \text{Bernoulli}(\mu(\theta^\star))$. For any $z \in \mathbb{R}$,

$$\mathbb{E}_Y\big[\ell(Y,z)\big] = \mathbb{E}_Y\left[\log\big(1+\exp(z)\big) - Yz\right] = \log\big(1+\exp(z)\big) - \mu(\theta^\star)z.$$

Note that $\mathbb{E}[\ell(Y,z)]$ is convex and $\theta^\star$ is its unique minimizer. Its direct results are

$$\mathbb{E}_Y\big[\ell(Y,\theta^\star)\big] < \mathbb{E}_Y\big[\ell(Y,0)\big] = \log 2,$$

and for $z \le 0$

$$\mathbb{E}_Y\big[\ell(Y,z)\big] \ge \mathbb{E}_Y\big[\ell(Y,0)\big] = \log 2.$$

Without loss of generality, take $c_1 = 0$ and $c_2 = 1$, and for $a \ge \sqrt{2}$ and $b \ge 1$, define

$$N(c_1) = \begin{cases} a, & \text{w.p. } 1/a^2 \\ -a, & \text{w.p. } 1/a^2 \\ 0, & \text{w.p. } 1 - 2/a^2 \end{cases}$$

and

$$N(c_2) = \begin{cases} b, & \text{w.p. } 1/2 \\ -b, & \text{w.p. } 1/2 \end{cases}.$$

Both noises have mean zero with $\text{Var}(N(c_1)) = 2$ and $\text{Var}(N(c_2)) = b^2$.

For $N(c_1)$, since $\theta^\star \in \mathcal{V} = [-S, S]$,

$$\begin{aligned}
\min_{\nu\in\mathcal{V}} \mathbb{E}\left[\ell\big(Y,\widehat{X}(c_1)\nu\big)\right] &\le \mathbb{E}\left[\ell\big(Y,\widehat{X}(c_1)\theta^\star\big)\right] \\
&= \mathbb{E}\left[\ell\big(Y,\big(1+N(c_1)\big)\theta^\star\big)\right] \\
&= \left(1-\frac{2}{a^2}\right)\mathbb{E}\left[\ell\big(Y,\theta^\star\big)\right] + \frac{1}{a^2}\mathbb{E}\left[\ell\big(Y,(1+a)\theta^\star\big)\right] \\
&\quad + \frac{1}{a^2}\mathbb{E}\left[\ell\big(Y,(1-a)\theta^\star\big)\right].
\end{aligned}$$

Since $\ell(y,z)$ is 1-Lipschitz in z,

$$\mathbb{E}\left[\ell\big(Y,(1\pm a)\theta^\star\big)\right] \le \mathbb{E}\left[\ell\big(Y,\theta^\star\big)\right] + a\theta^\star.$$

Therefore,

$$\min_{\nu\in\mathcal{V}} \mathbb{E}\left[\ell\big(Y,\widehat{X}(c_1)\nu\big)\right] \le \mathbb{E}\left[\ell\big(Y,\theta^\star\big)\right] + \frac{2\theta^\star}{a}.$$

For $N(c_2)$, for every $\nu \in \mathcal{V}$,

$$\mathbb{E}\left[\ell\big(Y,\widehat{X}(c_2)\nu\big)\right] = \frac{1}{2}\mathbb{E}\left[\ell\big(Y,(1+b)\nu\big)\right] + \frac{1}{2}\mathbb{E}\left[\ell\big(Y,(1-b)\nu\big)\right].$$

For any ν Since $b \ge 1$, one of $(1+b)\nu$ and $(1-b)\nu$ is nonpositive. One is larger than $\log 2$, and the other is larger than $\mathbb{E}[\ell(Y,\theta^\star)]$. Therefore, by taking minimum over $\nu \in \mathcal{V}$,

$$\min_{\nu\in\mathcal{V}} \mathbb{E}\left[\ell\big(Y,\widehat{X}(c_2)\nu\big)\right] \ge \frac{1}{2}\big(\log 2 + \mathbb{E}\left[\ell(Y,\theta^\star)\right]\big).$$

Now choose a sufficiently large so that

$$a > \frac{4\theta^\star}{\log 2 - \mathbb{E}\big[\ell(Y,\theta^\star)\big]},$$

then

$$\min_{\nu\in\mathcal{V}} \mathbb{E}\left[\ell\big(Y,\widehat{X}(c_1)\nu\big)\right] < \min_{\nu\in\mathcal{V}} \mathbb{E}\left[\ell\big(Y,\widehat{X}(c_2)\nu\big)\right].$$

Finally, setting $b = \sqrt{2}$ gives $\text{Var}(N(c_1)) = \text{Var}(N(c_2)) = 2$, while setting $b = 1$ gives $\text{Var}(N(c_1)) = 2 > 1 = \text{Var}(N(c_2))$. These show that even equal variance can yield different optimized logistic losses, and variance ordering does not imply the logistic loss ordering. □

D PROOF OF PROPOSITION 2

Proof. Under Assumption 2, Eq. (2) gives, for every $c_1 \leq c_2$ and every $\nu \in \mathcal{V}$

$$\mathbb{E}\Big[\ell\big(Y, \big(X + N(c_1)\big)^\top \nu\big)\Big] \geq \mathbb{E}\Big[\ell\big(Y, \big(X + N(c_2)\big)^\top \nu\big)\Big].$$

Thus, for a fixed $\nu \in \mathcal{V}$

$$\begin{aligned} L(c_2, \nu) &= \mathbb{E}\Big[\ell\big(Y, \big(X + N(c_2)\big)^\top \nu\big)\Big] + \lambda c_2 \\ &\leq \mathbb{E}\Big[\ell\big(Y, \big(X + N(c_1)\big)^\top \nu\big)\Big] + \lambda c_2 \\ &= L(c_1, \nu) + \lambda(c_2 - c_1). \end{aligned}$$

Using this and letting $\nu^\star(c_1) \in \arg\min_{\nu \in \mathcal{V}} L(c_1, \nu)$ and $\nu^\star(c_2) \in \arg\min_{\nu \in \mathcal{V}} L(c_2, \nu)$, we finally obtain

$$\begin{aligned} L^\star(c_2) &= L\big(c_2, \nu^\star(c_2)\big) \\ &\leq L\big(c_2, \nu^\star(c_1)\big) \\ &\leq L\big(c_1, \nu^\star(c_1)\big) + \lambda(c_2 - c_1) \\ &= L^\star(c_1) + \lambda(c_2 - c_1). \end{aligned}$$

□

E Proof of Proposition 3

Proof. For a random variable Z and $p \geq 1$, we denote its L_p-norm as

$$\|Z\|_{L_p} := \left(\mathbb{E}\big[|Z|^p\big]\right)^{1/p}.$$

Define a random variable

$$Z_t := \sup_{\nu \in \mathcal{V}} \left| \frac{1}{t} \sum_{s=1}^{t} \ell\Big(Y_s, \widehat{X}_s(c)^\top \nu\Big) - \mathbb{E}\left[\ell\left(Y, \widehat{X}(c)^\top \nu\right)\right] \right|.$$

Define $\phi(y, z) := \ell(y, z) - \log 2$, then ϕ is 1-Lipschitz in z and $\phi(y, 0) = 0$. Therefore, Z_t can be rewritten as:

$$Z_t = \sup_{\nu \in \mathcal{V}} \left| \frac{1}{t} \sum_{s=1}^{t} \phi\Big(Y_s, \widehat{X}_s(c)^\top \nu\Big) - \mathbb{E}\left[\phi\left(Y, \widehat{X}(c)^\top \nu\right)\right] \right|.$$

Letting $W_s = (Y_s, \widehat{X}_s(c))$, $\mathcal{F} = \{f_\nu : \nu \in \mathcal{V}\}$, where $f_\nu(y, x) = \phi(y, x^\top \nu)$, and $\Phi(x) = x^p$ with $p \geq 2$, Lemma 14 gives

$$\|Z_t\|_{L_p} \leq 2 \left\| \sup_{\nu \in \mathcal{V}} \left| \frac{1}{t} \sum_{s=1}^{t} \varepsilon_s \phi\left(Y_s, \widehat{X}_s(c)^\top \nu\right) \right| \right\|_{L_p},$$

where ε_s's are independent Rademacher random variables.

Now, conditional on observations $\mathcal{D}_t := \{(\widehat{X}_s(c), Y_s)\}_{s=1}^t$, let $f_s(z) := \phi(Y_s, z)$ and $A := \{(\widehat{X}_1(c)^\top \nu, \cdots, \widehat{X}_t(c)^\top \nu) : \nu \in \mathcal{V}\}$. Since $\phi(y, z)$ is 1-Lipschitz in z and $f_s(0) = \phi(Y_s, 0) = 0$, f_s is a contraction. Moreover, conditional on $\mathcal{D}_t$, A is bounded because $|\widehat{X}_s(c)^\top \nu| \leq S\|\widehat{X}_s(c)\|_2 < \infty$ for finite s. By taking $\Phi(x) = x^p$, we can apply Lemma 15 as

$$\begin{aligned}
\mathbb{E}_\varepsilon \left[\left(\frac{1}{2} \sup_{\nu \in \mathcal{V}} \left| \frac{1}{t} \sum_{s=1}^{t} \varepsilon_s \phi\Big(Y_s, \widehat{X}_s(c)^\top \nu\Big) \right| \right)^p \middle| \mathcal{D}_t \right] &\leq \mathbb{E}_\varepsilon \left[\left(\sup_{\nu \in \mathcal{V}} \left| \frac{1}{t} \sum_{s=1}^{t} \varepsilon_s \widehat{X}_s(c)^\top \nu \right| \right)^p \middle| \mathcal{D}_t \right] \\
&= \mathbb{E}_\varepsilon \left[\left(\frac{S}{t} \left\| \sum_{s=1}^{t} \varepsilon_s \widehat{X}_s(c) \right\|_2 \right)^p \middle| \mathcal{D}_t \right],
\end{aligned}$$

where the equation holds because $\sup_{\nu \in \mathcal{V}} |v^\top \nu| = S\|v\|_2$. Taking expectation over $\mathcal{D}_t$, followed by taking the p-th root, gives

$$\frac{1}{2} \left\| \sup_{\nu \in \mathcal{V}} \left| \frac{1}{t} \sum_{s=1}^{t} \varepsilon_s \phi\Big(Y_s, \widehat{X}_s(c)^\top \nu\Big) \right| \right\|_{L_p} \leq \frac{S}{t} \left(\mathbb{E}\left[\left\| \sum_{s=1}^{t} \varepsilon_s \widehat{X}_s(c) \right\|_2^p \right] \right)^{1/p}$$

Combining the above results, we finally get

$$\|Z_t\|_{L_p} \leq \frac{4S}{t} \left(\mathbb{E}\left[\left\| \sum_{s=1}^{t} \varepsilon_s \widehat{X}_s(c) \right\|_2^p \right] \right)^{1/p}.$$

Define

$$V_t := \sum_{s=1}^{t} \varepsilon_s \widehat{X}_s(c),$$

then $\mathbb{E}[V_t] = 0$. To apply Lemma 16, we first show that V_t is sub-Gaussian random vector. Let $m := \mathbb{E}[X]$, then the observed feature can be decomposed as $\widehat{X}(c) = m + \big(\widehat{X}(c) - m\big)$. By Assumption 1, $\|m\|_2 \leq B$, and $\big(\widehat{X}(c) - m\big) = (X - m) + N(c)$ is $2R^2$-sub-Gaussian due to independence of $(X - m)$ and $N(c)$. Since ε has values $+1$ or -1 with equal probability,

$$\mathbb{E}_{\varepsilon, X, N(c)} \left[\exp\left(u^\top \big(\varepsilon \widehat{X}(c)\big) \right) \right] = \frac{1}{2} \mathbb{E}_{X, N(c)} \left[\exp\left(u^\top \widehat{X}(c) \right) \right] + \frac{1}{2} \mathbb{E}_{X, N(c)} \left[\exp\left(-u^\top \widehat{X}(c) \right) \right].$$

For the first term,

$$\begin{aligned}\mathbb{E}_{X,N(c)}\left[\exp\left(u^\top \widehat{X}(c)\right)\right] &= \mathbb{E}_{X,N(c)}\left[\exp\left(u^\top\left(\left(\widehat{X}(c)-m\right)+m\right)\right)\right] \\ &\le \exp(u^\top m)\exp\left(R^2\|u\|_2^2\right),\end{aligned}$$

and similarly,

$$\mathbb{E}_{X,N(c)}\left[\exp\left(-u^\top \widehat{X}(c)\right)\right] \le \exp\left(-u^\top m\right)\exp\left(R^2\|u\|_2^2\right).$$

Therefore, since $\cosh(z) \le \exp(z^2/2)$ and $\|m\|_2 \le B$, we finally get

$$\begin{aligned}\mathbb{E}_{\varepsilon,X,N(c)}\left[\exp\left(u^\top\left(\varepsilon\widehat{X}(c)\right)\right)\right] &\le \frac{1}{2}\left(\exp\left(u^\top m\right)+\exp\left(-u^\top m\right)\right)\exp\left(R^2\|u\|_2^2\right) \\ &= \cosh\left(u^\top m\right)\exp\left(R^2\|u\|_2^2\right) \\ &\le \exp\left(R^2\|u\|_2^2+\frac{\left(u^\top m\right)^2}{2}\right) \\ &\le \exp\left(\frac{(B^2+2R^2)\|u\|_2^2}{2}\right),\end{aligned}$$

that is, $\varepsilon\widehat{X}(c)$ is (B^2+2R^2)-sub-Gaussian. Since the pairs $\left(\varepsilon_s, \widehat{X}_s(c)\right)$, $s=1,\cdots,t$, are independent across s, it follows that V_t is $t(B^2+2R^2)$-sub-Gaussian. Therefore Lemma 16 implies, for every $p \ge 2$

$$\begin{aligned}\left(\mathbb{E}\left[\left\|\sum_{s=1}^t \varepsilon_s \widehat{X}_s(c)\right\|_2^p\right]\right)^{1/p} &\le C\sqrt{t(B^2+2R^2)(d+p)} \\ &\le \sqrt{2}C(B+R)\sqrt{t(d+p)},\end{aligned}$$

where the second inequality is from $\sqrt{B^2+2R^2} \le B+\sqrt{2}R \le \sqrt{2}(B+R)$. Absorbing all constants into the universal constant C and combining this with the above result, we obtain

$$\|Z_t\|_{L_p} \le CS(B+R)\sqrt{\frac{d+p}{t}}. \tag{3}$$

Now, choose $p = 2+\log\frac{1}{\delta}$ so that $p \ge 2$, then by Markov's inequality,

$$\begin{aligned}\mathbb{P}\left(Z_t > e\|Z_t\|_{L_p}\right) &= \mathbb{P}\left(Z_t^p > e^p\|Z_t\|_{L_p}^p\right) \\ &\le \frac{\mathbb{E}[Z_t^p]}{e^p\|Z_t\|_{L_p}^p} \\ &= e^{-p} = \frac{\delta}{e^2} \le \delta.\end{aligned}$$

Moreover, $d+p = d+2+\log\frac{1}{\delta} \le 3(d+\log\frac{1}{\delta})$ for $d \ge 1$ and $\delta \in (0,1)$. Therefore, by Eq. (3),

$$e\|Z_t\|_{L_p} \le C_0 S(B+R)\sqrt{\frac{d+\log(1/\delta)}{t}}$$

for another universal constant $C_0 > 0$. This implies

$$\mathbb{P}\left(Z_t > C_0 S(B+R)\sqrt{\frac{d+\log(1/\delta)}{t}}\right) \le \delta,$$

and together with the fact

$$Z_t = \sup_{\nu\in\mathcal{V}}\left|\frac{1}{t}\sum_{s=1}^t \ell\left(Y_s, \widehat{X}_s(c)^\top\nu\right)+\lambda c - L(c,\nu)\right|,$$

we get the desired result.

□

E.1 SUPPORTING LEMMAS

Lemma 14 (Proposition 4.11 of Wainwright (2019)). *Let $W_1, \cdots, W_n$ be i.i.d. copies of random element W in a measurable space $\mathcal{W}$, and let $\mathcal{F}$ be a class of measurable functions $f : \mathcal{W} \to \mathbb{R}$. Let $\varepsilon_1, \cdots, \varepsilon_n$ be independent Rademacher random variables independent $W_1, \cdots, W_n$. Then, for any convex non-decreasing function $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$,*

$$\mathbb{E}_W \left[\Phi \left(\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{s=1}^{n} f(W_s) - \mathbb{E}\big[f(W)\big] \right| \right) \right] \leq \mathbb{E}_{W,\varepsilon} \left[\Phi \left(2 \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{s=1}^{n} \varepsilon_s f(W_s) \right| \right) \right].$$

Lemma 15 (Theorem 4.12 of Ledoux & Talagrand (1991)). *Let $A \subset \mathbb{R}^n$ be bounded, and let $f_s : \mathbb{R} \to \mathbb{R}$, $s = 1, \cdots, n$, satisfy $f_s(0) = 0$ and $|f_s(x) - f_s(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}$. Then, for any convex non-decreasing function $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$,*

$$\mathbb{E}_\varepsilon \left[\Phi \left(\frac{1}{2} \sup_{(x_1, \cdots, x_n) \in A} \left| \frac{1}{n} \sum_{s=1}^{n} \varepsilon_s f_s(x_s) \right| \right) \right] \leq \mathbb{E}_\varepsilon \left[\Phi \left(\sup_{(x_1, \cdots, x_n) \in A} \left| \frac{1}{n} \sum_{s=1}^{n} \varepsilon_s x_s \right| \right) \right].$$

Lemma 16 (Theorem 32 of Tan & Vershynin (2018)). *Let $V \in \mathbb{R}^d$ be mean-zero σ^2-sub-Gaussian random vector satisfying*

$$\mathbb{E}\left[\exp\left(u^\top V\right)\right] \leq \exp\left(\frac{\sigma^2 \|u\|_2^2}{2}\right)$$

for all $u \in \mathbb{R}^d$. Then for every $p \geq 2$, there exists a universal constant $C > 0$ such that

$$\left(\mathbb{E}\big[\|V\|_2^p\big]\right)^{1/p} \leq C\sigma\sqrt{d+p}.$$

F PROOF OF PROPOSITION 4

Proof. Fix a target payment $c \in [0,1]$ and a sample count $t \geq 1$. For every $s < \tau(c)$, we have $c_s \geq c$, so the learner can generate simulated feature

$$\widetilde{X}_s(c) = \widehat{X}_s + \xi^{(s)}_{c,c_s}.$$

Conditional on $\mathcal{H}_{s-1}$ and c_s, the increment $\xi^{(s)}_{c,c_s}$ is independent of $(\widehat{X}_s, Y_s)$ and by Assumption 2,

$$(\widetilde{X}_s(c), Y_s) \overset{d}{=} (\widehat{X}(c), Y).$$

Moreover, the conditional distribution of $(\widetilde{X}_s(c), Y_s)$ is the same fixed distribution regardless of the history. Therefore, for $s < \tau(c)$, the pair $(\widetilde{X}_s(c), Y_s)$ is independent of the preceding history and has the same distribution as $(\widehat{X}(c), Y)$.

To apply Proposition 3, we complete the sequence to length t with virtual observations from $\tau(c)$, only for the purpose of proof. For $s < \tau(c)$, set

$$\big(\overline{X}_s(c), Y_s\big) = \big(\widetilde{X}_s(c), Y_s\big),$$

and for $s \geq \tau(c)$, generate a virtual observation $(\overline{X}_s(c), Y_s)$ independently from the distribution of $(\widehat{X}(c), Y)$, and independently of preceding observations as well.

For every $s = 1, \cdots, t$, the completed pair $\overline{X}_s(c), Y_s)$ has the same distribution to $(\widehat{X}(c), Y)$ and is independent of the previous observations. If $s < \tau(c)$, it follows from the equivalence of conditional distributions above, while if $s \geq \tau(c)$, it holds by construction. Therefore, the complete sequence $\big((\overline{X}_s(c), Y_s)\big)_{s=1}^{t}$ is an i.i.d. sequence with the same distribution as $(\widehat{X}(c), Y)$.

Now, we can apply Proposition 3 as

$$\mathbb{P}\left(\sup_{\nu \in \mathcal{V}} \left| \frac{1}{t} \sum_{s=1}^{t} \ell\big(Y_s, \overline{X}_s(c)^\top \nu\big) + \lambda c - L(c,\nu) \right| > \gamma(t,\delta) \right) \leq \delta.$$

On the event $t < \tau(c)$, there is no virtual observation in the sequence, i.e. $\overline{X}_s(c) = \widetilde{X}_s(c)$ for all $s = 1, \cdots, t$. Consequently,

$$\begin{aligned}
&\left\{ t < \tau(c), \quad \sup_{\nu \in \mathcal{V}} \left| \frac{1}{t} \sum_{s=1}^{t} \ell\Big(Y_s, \widetilde{X}_s(c)^\top \nu\Big) + \lambda c - L(c,\nu) \right| > \gamma(t,\delta) \right\} \\
&\quad = \left\{ t < \tau(c), \quad \sup_{\nu \in \mathcal{V}} \left| \frac{1}{t} \sum_{s=1}^{t} \ell\big(Y_s, \overline{X}_s(c)^\top \nu\big) + \lambda c - L(c,\nu) \right| > \gamma(t,\delta) \right\} \\
&\quad \subseteq \left\{ \sup_{\nu \in \mathcal{V}} \left| \frac{1}{t} \sum_{s=1}^{t} \ell\big(Y_s, \overline{X}_s(c)^\top \nu\big) + \lambda c - L(c,\nu) \right| > \gamma(t,\delta) \right\}.
\end{aligned}$$

Combining this inclusion with the concentration result above, we obtain

$$\mathbb{P}\left(t < \tau(c), \quad \sup_{\nu \in \mathcal{V}} \left| \frac{1}{t} \sum_{s=1}^{t} \ell\Big(Y_s, \widetilde{X}_s(c)^\top \nu\Big) + \lambda c - L(c,\nu) \right| > \gamma(t,\delta) \right) \leq \delta.$$

Equivalently, with probability at least $1 - \delta$,

$$t < \tau(c) \quad \Longrightarrow \quad \sup_{\nu \in \mathcal{V}} \left| \frac{1}{t} \sum_{s=1}^{t} \ell\Big(Y_s, \widetilde{X}_s(c)^\top \nu\Big) + \lambda c - L(c,\nu) \right| \leq \gamma(t,\delta)$$

since $\{A \Rightarrow B\}^c = A \cap B^c$.

□

G PROOF OF THEOREM 5

Proof. Let $\mathcal{N}_T = \{1, 2, 4, 8, \cdots\} \cap [T-1]$ denote the set of checkpoints, which is the sample count afters immediately after a batch ends and empirical estimates are computed, but before elimination is performed, then $|\mathcal{N}_T| = \lceil \log_2 T \rceil$. For every grid index $k \in [K]_0$ and checkpoint $n \in \mathcal{N}_T$, define the failure event as

$$\mathcal{F}_{k,n} := \left\{ n < \tau(c^{(k)}), \quad \sup_{\nu \in \mathcal{V}} \left| \widehat{L}_{k,n}(\nu) - L(c^{(k)}, \nu) \right| > \gamma(n, \delta_T) \right\},$$

that is, on $\mathcal{F}_{k,n}$, the first n payments are all at least $c^{(k)}$, so n valid (simulated) observations at payment $c^{(k)}$ are available, but the concentration bound fails. By Proposition 4, for every k and n,

$$\mathbb{P}(\mathcal{F}_{k,n}) \le \delta_T.$$

For analysis, we keep generating simulated observations at payment $c^{(k)}$ as long as $s < \tau(c^{(k)})$ regardless of whether $c^{(k)}$ is active at checkpoint n or not. This is only a proof-side construction and does not affect the algorithm.

Define the good event over the whole procedure as

$$\mathcal{G} := \bigcap_{k \in [K]_0} \bigcap_{n \in \mathcal{N}_T} \mathcal{F}_{k,n}^c,$$

then by union bound,

$$\mathbb{P}\left(\mathcal{G}^c\right) = \mathbb{P}\left(\bigcup_{k \in [K]_0} \bigcup_{n \in \mathcal{N}_T} \mathcal{F}_{k,n} \right) \le \sum_{k \in [K]_0} \sum_{n \in \mathcal{N}_T} \mathbb{P}\left(\mathcal{F}_{k,n}\right) \le \delta.$$

From now on, we work on the good event $\mathcal{G}$.

Suppose a payment $c^{(k)}$ is active at checkpoint n, then the payment made in the first n rounds are at least $c^{(k)}$ by the selection rule. Therefore, if $c^{(k)}$ is active at checkpoint n, then $n < \tau(c^{(k)})$, which implies on $\mathcal{G}$,

$$\sup_{\nu \in \mathcal{V}} \left| \widehat{L}_{k,n}(\nu) - L(c^{(k)}, \nu) \right| \le \gamma(n, \delta_T). \tag{4}$$

For simplicity, write $\gamma_n := \gamma(n, \delta_T)$. Let $\nu_k^\star = \nu^\star(c^{(k)}) \in \arg\min_\nu L(c^{(k)}, \nu)$ so that $L^\star(c^{(k)}) = L(c^{(k)}, \nu_k^\star)$, and recall that

$$\widehat{\nu}_{k,n} \in \arg\min_{\nu} \widehat{L}_{k,n}(\nu), \qquad \widehat{L}_{k,n}^\star = \widehat{L}_{k,n}(\widehat{\nu}_{k,n}) \le \widehat{L}_{k,n}(\nu_k^\star).$$

By optimality of $\widehat{\nu}_{k,n}$ and Eq. (4),

$$\widehat{L}_{k,n}^\star = \widehat{L}_{k,n}(\widehat{\nu}_{k,n}) \le \widehat{L}_{k,n}(\nu_k^\star) \le L(c^{(k)}, \nu_k^\star) + \gamma_n = L^\star(c^{(k)}) + \gamma_n,$$

and conversely,

$$\widehat{L}_{k,n}^\star = \widehat{L}_{k,n}(\widehat{\nu}_{k,n}) \ge L(c^{(k)}, \widehat{\nu}_{k,n}) - \gamma_n \ge L^\star(c^{(k)}) - \gamma_n,$$

Therefore, for every active payment $c^{(k)}$ at checkpoint n,

$$\left| \widehat{L}_{k,n}^\star - L^\star(c^{(k)}) \right| \le \gamma_n, \tag{5}$$

that is, for every active payment $c^{(k)}$ at checkpoint n, its empirical optimal loss and population optimal loss differ by at most γ_n. In addition, again by Eqs. (4) and (5),

$$L(c^{(k)}, \widehat{\nu}_{k,n}) \le \widehat{L}_{k,n}(\widehat{\nu}_{k,n}) + \gamma_n \le \widehat{L}_{k,n}(\nu_k^\star) + \gamma_n \le L^\star(c^{(k)}) + 2\gamma_n. \tag{6}$$

Let $\mathcal{A}$ be an active set immediately before elimination at checkpoint n and let $U_n = \min_{k \in \mathcal{A}} \{ \widehat{L}_{k,n}^\star + \gamma_n \}$ be the threshold, where the index k is eliminated if $\widehat{L}_{k,n}^\star - \gamma_n > U_n$. Let $k^\star \in \arg\min_{k \in [K]_0} L^\star(c^{(k)})$ be the best grid payment index.

We show by induction over checkpoints that $k^\star$ is never eliminated under good event $\mathcal{G}$. It is active initially. Suppose it is active before elimination at checkpoint n. On the good event, by Eq. (5) and the optimality of $k^\star$, for every $k \in \mathcal{A}$,

$$\widehat{L}^\star_{k^\star,n} - \gamma_n \le L^\star(c^{(k^\star)}) \le L^\star(c^{(k)}) \le \widehat{L}^\star_{k,n} + \gamma_n.$$

Taking minimum of r.h.s. over $k \in \mathcal{A}$, we get

$$\widehat{L}^\star_{k^\star,n} - \gamma_n \le U_n,$$

which means that $k^\star$ remains active after checkpoint n. Therefore, by induction, $k^\star$ is never eliminated.

Suppose k remains active after elimination at checkpoint n, which means that

$$\widehat{L}^\star_{k,n} - \gamma_n \le U_n.$$

Since $k^\star$ is active at checkpoint n, the definition of U_n implies

$$U_n \le \widehat{L}^\star_{k^\star,n} + \gamma_n.$$

Combining these two inequalities,

$$\widehat{L}^\star_{k,n} \le \widehat{L}^\star_{k^\star,n} + 2\gamma_n.$$

Applying Eq. (5) to both k and $k^\star$,

$$L^\star(c^{(k)}) \le L^\star(c^{(k^\star)}) + 4\gamma_n.$$

that is, for every surviving payment $c^{(k)}$, its optimal population loss lies within $4\gamma_n$ of the best grid payment. Finally, combining this with Eq. (6), for every surviving payment $c^{(k)}$,

$$L(c^{(k)}, \widehat{\nu}_{k,n}) \le L^\star(c^{(k^\star)}) + 6\gamma_n. \tag{7}$$

Let $\mathcal{A}'$ be an active set after elimination, and let $k' = \max \mathcal{A}'$. During the next batch, the algorithm pays $c^{(k')}$, and uses the predictor $\widehat{\nu}_{k',n}$. Since k' survives after elimination, Eq. (7) gives

$$L(c^{(k')}, \widehat{\nu}_{k',n}) \le L^\star(c^{(k^\star)}) + 6\gamma_n.$$

For each checkpoint $n \in \mathcal{N}_T$, define the set of rounds in the batch after checkpoint n as

$$B_n = \big\{n+1, n+2, \cdots, \min\{2n, T\}\big\},$$

then $|B_n| \le n$, with equality except possibly for the last batch. Therefore, cumulative regret during the batch is

$$\sum_{t \in B_n} \big[L(c_t, \nu_t) - L^\star(c^{(k^\star)})\big] = |B_n| \times \big[L(c^{(k')}, \widehat{\nu}_{k',n}) - L^\star(c^{(k^\star)})\big] \le 6n\gamma_n.$$

Since $[T] = \{1\} \cup \bigcup_{n \in \mathcal{N}_T} B_n$ with disjoint batches B_n, the cumulative regret relative to the best grid payment is decomposed as:

$$\sum_{t=1}^{T} \big[L(c_t, \nu_t) - L^\star(c^{(k^\star)})\big] = L(c_1, \nu_1) - L^\star(c^{(k^\star)}) + \sum_{n \in \mathcal{N}_T} \sum_{t \in B_n} \big[L(c_t, \nu_t) - L^\star(c^{(k^\star)})\big].$$

For the first round, the algorithm pays $c_1 = 1$ and uses $\nu_1 = 0$. Since $\ell(Y, 0) = \log 2$ regardless of $Y \in \{0, 1\}$, $L(1, 0) = \log 2 + \lambda$. Moreover, since $L^\star(c^{(k^\star)}) \ge 0$, $L(c_1, \nu_1) - L^\star(c^{(k^\star)}) \le \log 2 + \lambda$. Combining the above results,

$$\begin{aligned}\sum_{t=1}^{T} \big[L(c_t, \nu_t) - L^\star(c^{(k^\star)})\big] &\le \log 2 + \lambda + \sum_{n \in \mathcal{N}_T} \sum_{t \in B_n} \big[L(c_t, \nu_t) - L^\star(c^{(k^\star)})\big] \\ &\le \log 2 + \lambda + 6 \sum_{n \in \mathcal{N}_T} n\gamma_n.\end{aligned}$$

Recall that

$$\delta_T = \frac{\delta}{(K+1)\lceil \log_2 T \rceil}, \quad \gamma_n = \gamma(n, \delta_T) = C_0 S(B+R)\sqrt{\frac{d + \log(1/\delta_T)}{n}},$$

and $D_T = d + \log(1/\delta_T)$, then

$$\begin{aligned}\sum_{n \in \mathcal{N}_T} n\gamma_n &= \sum_{n \in \mathcal{N}_T} C_0 S(B+R)\sqrt{nD_T} \\ &= C_0 S(B+R)\sqrt{D_T} \sum_{n \in \mathcal{N}_T} \sqrt{n}.\end{aligned}$$

Let $p = \lceil \log_2 T \rceil$ Since $\mathcal{N}_T = \{2^r : r = 0, \ldots, p-1\}$, we have

$$\sum_{n \in \mathcal{N}_T} \sqrt{n} = \sum_{r=0}^{p-1} \sqrt{2^r} = \frac{2^{p/2} - 1}{\sqrt{2} - 1} \le \frac{\sqrt{2T} - 1}{\sqrt{2} - 1} \le \frac{\sqrt{2}}{\sqrt{2} - 1}\sqrt{T}. \tag{8}$$

Substituting this, for some universal constant $C_1 > 0$,

$$\sum_{t=1}^{T} \left[L(c_t, \nu_t) - L^\star(c^{(k^\star)}) \right] \le C_1 S(B+R)\sqrt{TD_T} + \log 2 + \lambda.$$

It remains to compare the best grid loss $L^\star(c^{(k^\star)})$ with the optimal loss $L^\star$ over the continuous interval $[0, 1]$. For any $c \in [0, 1]$, we compare this with the grid point $c^{(k)} = \lceil Kc \rceil / K$, then

$$c \le c^{(k)} \le c + 1/K.$$

Since $c^{(k)} \ge c$, Proposition 2 implies that

$$L^\star(c^{(k)}) \le L^\star(c) + \lambda(c^{(k)} - c) \le L^\star(c) + \frac{\lambda}{K}.$$

Since $k^\star$ is the best grid index, $L^\star(c^{(k^\star)}) \le L^\star(c^{(k)})$ for every $k \in [K]_0$, and thus

$$L^\star(c^{(k^\star)}) \le L^\star(c) + \frac{\lambda}{K}$$

for any $c \in [0, 1]$. By taking infimum over $c \in [0, 1]$, we get

$$L^\star(c^{(k^\star)}) \le L^\star + \frac{\lambda}{K}. \tag{9}$$

Finally, since this holds on $\mathcal{G}$, with probability at least $1 - \delta$,

$$\begin{aligned}\sum_{t=1}^{T} \left[L(c_t, \nu_t) - L^\star \right] &= \sum_{t=1}^{T} \left[L(c_t, \nu_t) - L^\star(c^{(k^\star)}) \right] + \sum_{t=1}^{T} \left[L^\star(c^{(k^\star)}) - L^\star \right] \\ &\le C_1 S(B+R)\sqrt{TD_T} + \frac{\lambda T}{K} + \log 2 + \lambda.\end{aligned}$$

For the expected regret Reg_T, let $R_T = \sum_{t=1}^{T} \left[L(c_t, \nu_t) - L^\star \right]$, then $\mathrm{Reg}_T = \mathbb{E}[R_T]$ by Eq. (1). We decompose the expectation according to the good event $\mathcal{G}$ and bad event $\mathcal{G}^c$. On $\mathcal{G}$, the previous bound applies. We now bound the regret R_T on $\mathcal{G}^c$.

For every $y \in \{0, 1\}$ and $z \in \mathbb{R}$,

$$\ell(y, z) \le \log 2 + |z|$$

due to its 1-Lipschitzness in z and the fact that $\ell(y, 0) = \log 2$. Let $m := \mathbb{E}[X]$, then for any $c \in [0, 1]$ and $\nu \in \mathcal{V}$,

$$\begin{aligned}L(c, \nu) &= \mathbb{E}\left[\ell\big(Y, \widehat{X}(c)^\top \nu\big) \right] + \lambda c \\ &\le \log 2 + \mathbb{E}\left[\left| \widehat{X}(c)^\top \nu \right| \right] + \lambda \\ &\le \log 2 + \mathbb{E}\left[\left| m^\top \nu \right| \right] + \mathbb{E}\left[\left| (X - m)^\top \nu \right| \right] + \mathbb{E}\left[\left| N(c)^\top \nu \right| \right] + \lambda \\ &\le \log 2 + SB + \sqrt{\mathbb{E}\left[\left| (X - m)^\top \nu \right|^2 \right]} + \sqrt{\mathbb{E}\left[\left| N(c)^\top \nu \right|^2 \right]} + \lambda.\end{aligned}$$

Moreover, by Assumption 1, for any fixed $\nu \in \mathcal{V}$, both $(X - m)^\top \nu$ and $N(c)^\top \nu$ are mean-zero $R^2\|\nu\|_2^2$-sub-Gaussian random variables, hence their variances are bounded by $R^2\|\nu\|_2^2$. Therefore, together with the fact $L^\star \geq 0$,

$$\begin{aligned} L(c,\nu) - L^\star &\leq \log 2 + SB + \sqrt{\mathbb{E}\left[\left|(X-m)^\top \nu\right|^2\right]} + \sqrt{\mathbb{E}\left[\left|N(c)^\top \nu\right|^2\right]} + \lambda \\ &\leq \log 2 + SB + 2R\|\nu\|_2 + \lambda \\ &\leq \log 2 + S(B + 2R) + \lambda. \end{aligned}$$

Consequently, on every realization on $\mathcal{G}^c$,

$$R_T \leq T\big[\log 2 + S(B + 2R) + \lambda\big].$$

Finally, using $\mathbb{P}(\mathcal{G}^c) \leq \delta$, we obtain

$$\begin{aligned} \mathrm{Reg}_T &= \mathbb{E}[R_T \mid \mathcal{G}] \times \mathbb{P}(\mathcal{G}) + \mathbb{E}[R_T \mid \mathcal{G}^c] \times \mathbb{P}(\mathcal{G}^c) \\ &\leq C_1 S(B+R)\sqrt{TD_T} + \frac{\lambda T}{K} + \log 2 + \lambda \\ &\quad + \delta T\big[\log 2 + S(B + 2R) + \lambda\big]. \end{aligned}$$

□

H Proof of Theorem 10

Proof. Let $\mathcal{M}_T := \{1, 2, 4, 8, \cdots\} \cap [T-1]$ be the set of possible local checkpoints, then $|\mathcal{M}_T| = \lceil \log_2 T \rceil$.

For each grid $k \in [K]_0$ and its local checkpoint $m \in \mathcal{M}_T$, define the local good event as

$$\mathcal{G}_{k,m} := \left\{ \sup_{\nu \in \mathcal{V}} \left| \widehat{L}_{k,m}(\nu) - L(c^{(k)}, \nu) \right| \le \gamma(m, \delta_T) \right\}.$$

For analysis, for each grid payment $c^{(k)}$, consider an infinite i.i.d. sequence $(\widehat{X}_{k,s}, Y_{k,s})_{s \ge 1}$. When the algorithm selects $c^{(k)}$, it observes the next unobserved pair in order. This makes $\widehat{L}_{k,m}$ well-defined for every $k \in [K]_0$ and $m \in \mathcal{M}_T$ even if the algorithm never actually reaches m pulls of $c^{(k)}$. This is only a proof-side construction and does not affect the algorithm.

Although the rounds at which $c^{(k)}$ is paid are adaptive, the observations $(\widehat{X}_{k,s}, Y_{k,s})_{s \ge 1}$ collected at payment $c^{(k)}$ are i.i.d. samples of $(\widehat{X}(c^{(k)}), Y)$. Indeed, conditional on history and on selecting $c^{(k)}$, the current observation has this fixed distribution independently of the preceding history. Therefore, Proposition 3 gives

$$\mathbb{P}(\mathcal{G}_{k,m}) \ge 1 - \delta_T.$$

Define a good event as

$$\mathcal{G} := \bigcap_{k \in [K]_0} \bigcap_{m \in \mathcal{M}_T} \mathcal{G}_{k,m},$$

then by union bound,

$$\mathbb{P}(\mathcal{G}^c) = \mathbb{P}\left(\bigcup_{k \in [K]_0} \bigcup_{m \in \mathcal{M}_T} \mathcal{G}_{k,m}^c \right) \le \sum_{k \in [K]_0} \sum_{m \in \mathcal{M}_T} \mathbb{P}\left(\mathcal{G}_{k,m}^c\right) \le \delta.$$

We work on the good event $\mathcal{G}$.

For simplicity write $\gamma_m := \gamma(m, \delta_T)$. As in the known-profile proof (Appendix G), uniform concentration implies

$$\left| \widehat{L}^\star_{k,m} - L^\star(c^{(k)}) \right| \le \gamma_m \tag{10}$$

and

$$L(c^{(k)}, \widehat{\nu}_{k,m}) \le \widehat{L}^\star_{k,m} + \gamma_m \le L^\star(c^{(k)}) + 2\gamma_m.$$

Let $k^\star \in \arg\min_{k \in [K]_0} L^\star(c^{(k)})$ be the best grid payment index. At each round t after initialization, i.e., $t \ge K + 2$, by the selection rule,

$$\begin{aligned} \widehat{L}^\star_{k_t, m_t(k_t)} - \gamma_{m_t(k_t)} &\le \widehat{L}^\star_{k^\star, m_t(k^\star)} - \gamma_{m_t(k^\star)} \\ &\le L^\star(c^{(k^\star)}). \qquad \text{(by Eq. (10))} \end{aligned}$$

By the concentration in the set $\mathcal{G}_{k_t, m_t(k_t)}$, we know

$$L(c^{(k_t)}, \widehat{\nu}_{k_t, m_t(k_t)}) \le \widehat{L}_{k_t, m_t(k_t)}(\widehat{\nu}_{k_t, m_t(k_t)}) + \gamma_{m_t(k_t)} = \widehat{L}^\star_{k_t, m_t(k_t)} + \gamma_{m_t(k_t)}.$$

Combining the above inequalities, together with $\nu_t = \widehat{\nu}_{k_t, m_t(k_t)}$

$$L(c^{(k_t)}, \nu_t) \le L^\star(c^{(k^\star)}) + 2\gamma_{m_t(k_t)}. \tag{11}$$

For a fixed grid payment $c^{(k)}$, it is selected $n_T(k)$ times over T. Since every payment is selected once in the initialization, we have $n_T(k) \ge 1$ almost surely, and since exactly one payment is selected at each round,

$$\sum_{k \in [K]_0} n_T(k) = T.$$

For each payment $c^{(k)}$, let $\mathcal{M}_T(k) := \{2^r : r = 0, 1, \cdots, \lceil \log_2 n_T(k) \rceil - 1\}$ be the set of local checkpoints after which $c^{(k)}$ is selected at least once.

For each $m \in \mathcal{M}_T(k)$, let $B_{k,m} := \{t \geq K+2 : k_t = k, m_t(k) = m\}$ be the set of rounds at which $c^{(k)}$ is selected and estimator computed at checkpoint m is used. This estimator is used until the count doubles, so $|B_{k,m}| \leq m$.

For $t \in B_{k,m}$, $k_t = k$, $m_t(k) = m$, and hence by Eq. (11),

$$L(c_t, \nu_t) - L^\star(c^{(k^\star)}) \leq 2\gamma_m.$$

Recall that $\gamma_m = \gamma(m, \delta_T) = C_0 S(B+R)\sqrt{D_T/m}$. By partitioning the rounds after initialization by the sets $B_{k,m}$,

$$\begin{aligned}
\sum_{t=K+2}^{T} \Big[L(c_t, \nu_t) - L^\star(c^{(k^\star)})\Big] &= \sum_{k\in[K]_0} \sum_{m\in\mathcal{M}_T(k)} \sum_{t\in B_{k,m}} \Big[L(c^{(k_t)}, \nu_t) - L^\star(c^{(k^\star)})\Big] \\
&\leq \sum_{k\in[K]_0} \sum_{m\in\mathcal{M}_T(k)} \sum_{t\in B_{k,m}} 2\gamma_m \\
&\leq \sum_{k\in[K]_0} \sum_{m\in\mathcal{M}_T(k)} 2m\gamma_m \\
&\leq 2C_0 S(B+R)\sqrt{D_T} \sum_{k\in[K]_0} \sum_{m\in\mathcal{M}_T(k)} \sqrt{m}.
\end{aligned}$$

By the same geometric-sum argument as in Eq. (8),

$$\sum_{m\in\mathcal{M}_T(k)} \sqrt{m} \leq \frac{\sqrt{2}}{\sqrt{2}-1}\sqrt{n_T(k)}.$$

Substituting this into the above decomposition, we obtain

$$\begin{aligned}
\sum_{t=K+2}^{T} \Big[L(c_t, \nu_t) - L^\star(c^{(k^\star)})\Big] &\leq C_2 S(B+R)\sqrt{D_T} \sum_{k\in[K]_0} \sqrt{n_T(k)} \\
&\leq C_2 S(B+R)\sqrt{D_T(K+1)\sum_{k\in[K]_0} n_T(k)} \quad \text{(Cauchy–Schwarz)} \\
&= C_2 S(B+R)\sqrt{(K+1)TD_T},
\end{aligned}$$

where C_2 is a universal constant. It remains to bound the regret during $K+1$ initialization rounds. During initialization, every payment is selected once and the algorithm uses $\nu = 0$. Since

$$L(c^{(k)}, 0) = \log 2 + \lambda c^{(k)} \leq \log 2 + \lambda$$

and $L^\star(c^{(k^\star)}) \geq 0$,

$$\sum_{t=1}^{K+1} \Big[L(c_t, \nu_t) - L^\star(c^{(k^\star)})\Big] \leq (K+1)(\log 2 + \lambda).$$

Combining them,

$$\sum_{t=1}^{T} \Big[L(c_t, \nu_t) - L^\star(c^{(k^\star)})\Big] \leq C_2 S(B+R)\sqrt{(K+1)TD_T} + (K+1)(\log 2 + \lambda).$$

Finally, using Eq. (9),

$$\begin{aligned}
\sum_{t=1}^{T} \Big[L(c_t, \nu_t) - L^\star\Big] \leq C_2 S(B+R)\sqrt{(K+1)TD_T} + \frac{\lambda T}{K} \\
+ (K+1)(\log 2 + \lambda)
\end{aligned}$$

on the event $\mathcal{G}$, which holds with probability at least $1 - \delta$.

By the same argument in Appendix G, we get

$$\mathrm{Reg}_T \le C_2 S(B+R)\sqrt{(K+1)TD_T} + \frac{\lambda T}{K} + (K+1)(\log 2 + \lambda) + \delta T\big[\log 2 + S(B+2R) + \lambda\big],$$

which proves the theorem. □

I PROOF OF LOWER BOUNDS

In this section, we establish the lower bounds for both the known- and unknown-profile settings. Both constructions are carried out in a one-dimensional Gaussian model, which already captures the statistical difficulty responsible for the respective regret rates. We first collect several regularity properties of the optimized scalar Gaussian logistic loss in Section I.1 that will be used throughout the constructions. We then prove the known-profile lower bound in Section I.2, followed by the unknown-profile lower bound in Section I.3.

For an environment M and a possibly randomized policy π, write

$$\mathrm{Reg}_T^{\pi}(M) := \mathbb{E}_M^{\pi}\left[\sum_{t=1}^{T}\big(L_M(c_t,\nu_t) - L_M^{\star}\big)\right]$$

for expected regret, where $L_M(c,\nu)$ is the population loss including payment and $L_M^{\star} = \inf_{c\in[0,1],\,|\nu|\le S} L_M(c,\nu)$. We index the environment by its clean-feature variance a in the known-profile construction and by its noise-variance profile s in the unknown-profile construction. Any additional revelation of X_t occurs after the prediction and response at round t; the payment and predictor are still selected before observing the current corrupted feature.

I.1 REGULARITY OF THE SCALAR GAUSSIAN LOGISTIC LOSS

Fix $S > 0$ and $0 < \theta \le S$. For $a > 0$ and $s \ge 0$, let

$$X \sim \mathcal{N}(0,a), \qquad N_s \sim \mathcal{N}(0,s),$$

independently, and let $Y \mid X \sim \mathrm{Ber}(\mu(\theta X))$. Define

$$m(a) := \mathbb{E}[XY] = \mathbb{E}[X\mu(\theta X)].$$

By Stein's identity,

$$m(a) = a\theta\,\mathbb{E}\big[\mu'(\theta\sqrt{a}Z)\big] > 0, \qquad Z \sim \mathcal{N}(0,1). \tag{12}$$

For $q \ge 0$, define

$$\Psi(q) := \mathbb{E}_{Z\sim\mathcal{N}(0,1)}\left[\log(1+\exp(\sqrt{q}Z))\right].$$

The prediction loss excluding payment is defined by

$$F(a,s,\nu) := \Psi((a+s)\nu^2) - m(a)\nu, \qquad |\nu| \le S,$$

and optimal prediction loss is defined by

$$H(a,s) := \min_{|\nu|\le S} F(a,s,\nu).$$

The following lemma provides the basic regularity properties needed by both lower-bound constructions. For the known-profile lower bound, the strict increase of $H(a,s)$ in the noise variance together with differentiability in a will allow us to identify a point at which two payment regions are locally tied and exchange optimality under a small perturbation of the clean-feature variance. For the unknown-profile lower bound, strict monotonicity in s will allow us to prescribe a desired optimized-loss profile and map it back to a Gaussian noise-variance profile.

Lemma 17 (Regularity of the constrained scalar Gaussian loss)**.** *For every $a > 0$ and $s \ge 0$:*

1. *the map $\nu \mapsto F(a,s,\nu)$ is strictly convex on $[-S,S]$;*
2. *the constrained minimizer*
$$\nu_\star(a,s) := \arg\min_{|\nu|\le S} F(a,s,\nu)$$
is unique and satisfies $0 < \nu_\star(a,s) \le S$;
3. *$(a,s) \mapsto \nu_\star(a,s)$ is continuous on $(0,\infty)\times[0,\infty)$;*
4. *H is continuously differentiable on $(0,\infty)\times[0,\infty)$, and*
$$\partial_s H(a,s) = \nu_\star(a,s)^2\Psi'\big((a+s)\nu_\star(a,s)^2\big) > 0. \tag{13}$$

In particular, the conclusion does not require the constraint $|\nu| \le S$ to be inactive.

Proof. Let $W \sim \mathcal{N}(0, a+s)$. Then

$$\Psi((a+s)\nu^2) = \mathbb{E}[\log(1+e^{\nu W})].$$

Differentiating twice in ν gives

$$\partial^2_{\nu\nu} F(a,s,\nu) = \mathbb{E}\left[\mu'(\nu W) W^2\right].$$

Because $a+s>0$, W is nondegenerate, and $\mu'(z)>0$ for every finite z. Hence the right-hand side is strictly positive for every ν, proving strict convexity and uniqueness of the constrained minimizer.

At $\nu = 0$,

$$\partial_\nu F(a,s,0) = -m(a) < 0$$

by (12). Thus the objective strictly decreases when moving from zero in the positive direction, so its unique minimizer cannot be zero or negative. Hence

$$0 < \nu_\star(a,s) \le S.$$

The boundary value $\nu_\star(a,s) = S$ is allowed.

The function F is jointly continuous in (a,s,ν), while $[-S,S]$ is compact. Berge's maximum theorem and uniqueness of the minimizer imply continuity of $(a,s) \mapsto \nu_\star(a,s)$.

The map $a \mapsto m(a)$ is continuously differentiable on $(0,\infty)$. This follows from (12) by differentiation under the Gaussian expectation, using boundedness of the relevant derivatives of the logistic function. Moreover, the Gaussian heat equation gives

$$\Psi'(q) = \frac{1}{2}\mathbb{E}[\mu'(\sqrt{q}Z)], \qquad q \ge 0,$$

with the continuous extension $\Psi'(0) = 1/8$. Thus F is continuously differentiable in (a,s).

Since the constrained minimizer is unique, Danskin's theorem yields

$$\partial_s H(a,s) = \partial_s F(a,s,\nu_\star(a,s)),$$

which is exactly (13). Both factors in (13) are strictly positive. Continuity of the derivative follows from continuity of $\nu_\star$ and of the derivatives of F. The same argument applies to the a derivative, proving $H \in C^1$. □

For the unknown-profile construction, we need a quantitative version of this monotonicity: a small perturbation of the optimized logistic loss should induce only a controlled perturbation of the corresponding noise variance. The next corollary provides a uniformly Lipschitz inverse for $s \mapsto H(a,s)$ on a compact variance interval. It will be used both to keep the constructed variance profiles uniformly Lipschitz and to control the KL divergence between nearby profiles.

Corollary 18 (Uniform inverse regularity)**.** *Fix $a > 0$ and $0 < s_{\min} < s_{\max} < \infty$, and write $H_a(s) := H(a,s)$. Then*

$$H_a : [s_{\min}, s_{\max}] \longrightarrow [H_a(s_{\min}), H_a(s_{\max})]$$

is a continuously differentiable, strictly increasing bijection. Furthermore,

$$g := \min_{s\in[s_{\min},s_{\max}]} H_a'(s) > 0, \tag{14}$$

and

$$|H_a^{-1}(u) - H_a^{-1}(v)| \le \frac{1}{g}|u-v|. \tag{15}$$

Proof. Lemma 17 gives continuity and strict positivity of H_a' on the compact interval. Hence (14) holds. Strict monotonicity gives invertibility, and the mean-value theorem gives (15). □

Both lower-bound arguments compare nearby centered Gaussian distributions. We therefore record the following elementary quadratic bound on their KL divergence.

Lemma 19 (Quadratic KL bound for centered Gaussians)**.** *Let $u, v \geq \underline{s} > 0$. Then*

$$D_{\mathrm{KL}}(\mathcal{N}(0,u) \,\|\, \mathcal{N}(0,v)) \leq \frac{(u-v)^2}{4\underline{s}^{\,2}}.$$

Proof. For centered Gaussians,

$$D_{\mathrm{KL}}(\mathcal{N}(0,u) \,\|\, \mathcal{N}(0,v)) = \frac{1}{2}\left(\frac{u}{v} - 1 - \log\frac{u}{v}\right).$$

For fixed v, the value and first derivative of the right-hand side vanish at $u = v$, while its second derivative with respect to u equals $1/(2u^2) \leq 1/(2\underline{s}^{\,2})$. Taylor's theorem therefore gives the claimed bound. □

I.2 Proof of Theorem 8: known profile

Proof sketch. We construct a known step-shaped Gaussian noise profile with two relevant payment regions, $c < c_0$ and $c \geq c_0$. We first choose a clean-feature variance a_0 at which the optimized population loss of the two regions are tied. Perturbing the variance to $a_- = a_0 - \varepsilon$ and $a_+ = a_0 + \varepsilon$ makes opposite regions optimal, with a loss gap of order ε for choosing the wrong region. Even if the clean feature X_t is revealed after every round, the two instances remain statistically hard to distinguish: their transcript KL divergence is of order $T\varepsilon^2$. Taking $\varepsilon \asymp T^{-1/2}$ keeps the two instances close while forcing a linear number of wrong-region decisions in at least one of them, which yields $\Omega(\sqrt{T})$ regret.

Fix $s_0 > 0$ and define

$$G(a) := H(a, s_0) - H(a, 0), \qquad a > 0.$$

Lemma 20 (Existence of a transverse crossing)**.** *There exists $a_0 > 0$ such that*

$$G(a_0) > 0, \qquad G'(a_0) > 0.$$

Proof. For every $a > 0$, Lemma 17 implies that $s \mapsto H(a, s)$ is strictly increasing. Hence

$$G(a) > 0 \qquad \text{for all } a > 0.$$

The function G has a continuous extension to $a = 0$. Indeed, $m(a) \to 0$ as $a \downarrow 0$, and for fixed $s \geq 0$ the functions $F(a, s, \nu)$ converge uniformly over $|\nu| \leq S$ to

$$F(0, s, \nu) = \Psi(s\nu^2).$$

Since Ψ is nondecreasing and $\nu = 0$ is feasible,

$$H(0, s) = \min_{|\nu| \leq S} \Psi(s\nu^2) = \Psi(0) = \log 2$$

for every $s \geq 0$. Therefore

$$G(0) = 0.$$

Fix any $A > 0$. Since $G(A) > 0$ and $G(a) \to 0$ as $a \downarrow 0$, there exists $a_1 \in (0, A)$ such that $G(a_1) < G(A)$. By Lemma 17, G is continuously differentiable on $(0, \infty)$. The mean-value theorem therefore gives some $a_0 \in (a_1, A)$ for which

$$G'(a_0) = \frac{G(A) - G(a_1)}{A - a_1} > 0.$$

□

Theorem 21 (Known Gaussian step-profile lower bound)**.** *Fix any $S > 0$. There exist fixed constants*

$$\theta \in (0, S], \qquad a_0 > 0, \qquad s_0 > 0, \qquad c_0 \in (0, 1), \qquad \lambda > 0,$$

a completely known nonincreasing Gaussian variance profile

$$s(c) = s_0 \mathbf{1}\{c < c_0\}, \tag{16}$$

and constants $0 < \rho < a_0$ and $c > 0$ such that, for all sufficiently large T and every possibly randomized learning policy π,

$$\inf_{\pi} \sup_{a \in [a_0-\rho, a_0+\rho]} \mathrm{Reg}_T^{\pi}(a) \geq c\sqrt{T}. \tag{17}$$

The result remains valid when θ, S, λ, a_0, ρ and the intire noise profile are known to the learner, while the variance a remains unknown, and X_t is additionally revealed after every round.

Proof. Choose $0 < \theta \leq S$ and fix any $s_0 > 0$. Let a_0 be supplied by Lemma 20. Set

$$c_0 := \frac{1}{2}, \qquad \lambda := \frac{G(a_0)}{c_0} = 2G(a_0) > 0.$$

Consider the known step profile (16). Since this Gaussian variance profile is nonincreasing in the payment, it admits the required ordered additive degradation assumption (Assumption 2): for $c_1 \leq c_2$, the lower-payment noise can be represented as the higher-payment noise plus an independent centered Gaussian increment. Thus the profile satisfies the ordered additive degradation assumption, and the distributions of the Gaussian increments can be revealed to the learner.

For a fixed clean-feature variance a, all payments $c < c_0$ have the same prediction loss $H(a, s_0)$, so the best payment in that region is $c = 0$, with optimized population loss $H(a, s_0)$. Likewise, all payments $c \geq c_0$ have prediction loss $H(a, 0)$, so the best payment in that region is $c = c_0$, with optimized population loss $H(a, 0) + \lambda c_0$.

Define

$$D(a) := H(a, s_0) - H(a, 0) - \lambda c_0 = G(a) - G(a_0).$$

Then

$$D(a_0) = 0, \qquad D'(a_0) = G'(a_0) > 0.$$

By continuity of D', there exist $\rho, \gamma > 0$, with $\rho < a_0$, such that

$$D'(a) \geq \gamma \qquad \text{whenever } |a - a_0| \leq \rho. \tag{18}$$

All Gaussian instances in this interval satisfy Assumption 1 with any fixed $B > 0$ and

$$R \geq \max\{1, \sqrt{a_0 + \rho}, \sqrt{s_0}\}.$$

These bounds, as well as θ, S, λ and the noise profile, are independent of T.

Fix any possibly randomized policy π. For $0 < \varepsilon \leq \rho$, define

$$a_+ := a_0 + \varepsilon, \qquad a_- := a_0 - \varepsilon.$$

Integrating (18) gives

$$D(a_+) \geq \gamma\varepsilon, \qquad D(a_-) \leq -\gamma\varepsilon. \tag{19}$$

Thus in the a_+ instance the region $c \geq c_0$ is optimal, and every choice $c < c_0$ incurs expected excess loss at least $\gamma\varepsilon$. In the a_- instance the region $c < c_0$ is optimal, and every choice $c \geq c_0$ incurs expected excess loss at least $\gamma\varepsilon$. These statements remain true for arbitrary predictors because each payment-region comparison uses the prediction loss minimized over $|\nu| \leq S$.

We again strengthen the learner by revealing X_t after each round. Let P_+^T and P_-^T denote the enhanced transcript laws. Conditional on X_t and the chosen payment, the label channel and the measurement-noise channel are identical in the two instances. The clean feature is fresh and independent of the past, so the chain rule gives

$$D_{\mathrm{KL}}(P_+^T \| P_-^T) = T\, D_{\mathrm{KL}}(\mathcal{N}(0, a_+) \,\|\, \mathcal{N}(0, a_-)).$$

For $\varepsilon \leq a_0/2$, both variances are at least $a_0/2$. By Lemma 19 with $u = a_+$, $v = a_-$, and $\underline{s} = a_0/2$,

$$D_{\mathrm{KL}}(\mathcal{N}(0, a_+) \,\|\, \mathcal{N}(0, a_-)) \leq \frac{4\varepsilon^2}{a_0^2}.$$

Choose

$$\varepsilon = \frac{\eta}{\sqrt{T}},$$

where $\eta > 0$ is a fixed constant small enough that $4\eta^2/a_0^2 \le 1/8$. For sufficiently large T, we also have $\varepsilon \le \min\{\rho, a_0/2\}$. Then

$$D_{\mathrm{KL}}(P_+^T \| P_-^T) \le \frac{1}{8}, \qquad \mathrm{TV}(P_+^T, P_-^T) \le \frac{1}{4}.$$

Let

$$M_T := \sum_{t=1}^{T} \mathbf{1}\{c_t \ge c_0\}.$$

In the a_+ instance the number of wrong-region decisions is $T - M_T$, whereas in the a_- instance it is M_T. Since $0 \le M_T \le T$,

$$\begin{aligned}\mathbb{E}_+[T - M_T] + \mathbb{E}_-[M_T] &= T - \big(\mathbb{E}_+[M_T] - \mathbb{E}_-[M_T]\big) \\ &\ge T\big(1 - \mathrm{TV}(P_+^T, P_-^T)\big) \\ &\ge \frac{3T}{4}.\end{aligned}$$

Therefore at least one of the two instances has expected wrong-region count at least $3T/8$. By (19),

$$\begin{aligned}\max\{\mathrm{Reg}_T^\pi(a_+), \mathrm{Reg}_T^\pi(a_-)\} &\ge \frac{3\gamma}{8} T\varepsilon \\ &= \frac{3\gamma\eta}{8}\sqrt{T}.\end{aligned}$$

Since π was arbitrary and $a_\pm \in [a_0 - \rho, a_0 + \rho]$, this proves (17). The same fixed model bounds prove Theorem 21. □

Remark 22 (Source of the known-profile lower bound)**.** *The known-profile construction leaves only the clean-feature variance $a \in \{a_-, a_+\}$ unknown. The measurement profile, payment cost, structural logistic parameter, and predictor radius may all be revealed, and the clean feature may be observed after each round. Thus the $\sqrt{T}$ lower bound arises from the difficulty of distinguishing two nearby instances whose optimal payment regions are different, rather than from estimating the logistic parameter or recovering the clean feature.*

I.3 Proof of Theorem 12: unknown profile

Proof sketch. We first use the inverse regularity of the scalar logistic loss to construct a baseline Gaussian variance profile whose optimized population loss is exactly flat over the entire payment interval. We then create K alternative profiles, each obtained by inserting a small improvement of depth $\delta_K \asymp 1/K$ inside one of K disjoint payment regions. Under alternative k, only observations collected inside the hidden region A_k are informative for distinguishing it from the flat baseline. A pigeonhole argument identifies a region that is visited only $O(T/K)$ times under the baseline, and the resulting transcript KL divergence is of order T/K^3. Choosing $K \asymp T^{1/3}$ keeps this divergence bounded, so the learner cannot reliably localize the improved region and must spend a constant fraction of the horizon outside it. Since each such round incurs a gap of order $1/K$, the resulting regret is $\Omega(T/K) = \Omega(T^{2/3})$. The constructed Gaussian profiles are coupled through Brownian motion so that they satisfy the required ordered additive degradation assumption.

Theorem 23 (Unknown decreasing Lipschitz Gaussian profile)**.** *Fix any $S > 0$. There exist fixed constants*

$$a > 0, \quad \theta \in (0, S], \quad \lambda > 0, \quad 0 < s_{\min} < s_{\max} < \infty,$$

and $L, c > 0$, independent of T, such that the following holds. Let

$$\mathcal{S}_L := \{s : [0,1] \to [s_{\min}, s_{\max}] : s \text{ is nonincreasing and } L\text{-Lipschitz}\}.$$

For every sufficiently large T and every possibly randomized learning policy π,

$$\sup_{s \in \mathcal{S}_L} \mathrm{Reg}_T^\pi(s) \ge cT^{2/3}. \tag{20}$$

The conclusion remains true if $a, \theta, S, \lambda, s_{\min}, s_{\max}$ are revealed to the learner and the clean feature X_t is additionally revealed after every round. Every Gaussian noise profile in this class satisfies Assumption 2. In particular, one may take $s_{\min} = 1$ and $s_{\max} = 2$.

Proof. Fix any $a > 0$, $0 < \theta \le S$, and $0 < s_{\min} < s_{\max} < \infty$. Write $H = H_a$, where $H_a(s) = H(a, s)$, and define

$$C := H(s_{\max}), \qquad \lambda := H(s_{\max}) - H(s_{\min}) > 0.$$

All profiles constructed below satisfy Assumption 1 with any fixed $B > 0$ and

$$R \ge \max\{1, \sqrt{a}, \sqrt{s_{\max}}\}.$$

The payment weight λ and these model bounds are fixed independently of T. Define the baseline variance profile

$$s_0(c) := H^{-1}(C - \lambda c), \qquad c \in [0, 1].$$

By construction,

$$H(s_0(c)) + \lambda c = C \qquad \text{for every } c \in [0, 1].$$

Thus the baseline optimized population loss is flat over the entire payment interval.

Let $K \ge 1$ be an integer. Partition $J = [1/4, 3/4]$ into K cells of length $1/(2K)$, and let z_k be the midpoint of cell k. Define

$$\phi_k(c) := \big(1 - 4K|c - z_k|\big)_+, \qquad k \in [K],$$

and let

$$A_k := \{c \in [0, 1] : \phi_k(c) > 0\}.$$

The sets $A_1, \dots, A_K$ are pairwise disjoint, $0 \le \phi_k \le 1$, $\phi_k(z_k) = 1$, and ϕ_k is $4K$-Lipschitz.

Set

$$\alpha := \frac{\lambda}{16}, \qquad \delta_K := \frac{\alpha}{K},$$

and define

$$q_k(c) := C - \lambda c - \delta_K \phi_k(c), \qquad s_k(c) := H^{-1}(q_k(c)). \tag{21}$$

We verify that s_k is well defined and belongs to a common decreasing Lipschitz class. If $c \notin A_k$, then $q_k(c) = C - \lambda c$ lies in the domain of H^{-1}. If $c \in A_k \subset [1/4, 3/4]$, then

$$C - \lambda c \ge H(s_{\min}) + \frac{\lambda}{4},$$

while $\delta_K \le \lambda/16$. Hence

$$q_k(c) \ge H(s_{\min}) + \frac{3\lambda}{16} > H(s_{\min}),$$

and clearly $q_k(c) \le H(s_{\max})$. Therefore $s_k(c)$ is well defined and remains in $[s_{\min}, s_{\max}]$.

For $c_1 < c_2$, using the $4K$-Lipschitz property of ϕ_k,

$$\begin{aligned} q_k(c_2) - q_k(c_1) &= -\lambda(c_2 - c_1) - \delta_K\big(\phi_k(c_2) - \phi_k(c_1)\big) \\ &\le (-\lambda + 4K\delta_K)\,(c_2 - c_1) \\ &= -\frac{3\lambda}{4}(c_2 - c_1) < 0. \end{aligned}$$

Thus q_k is decreasing, and H^{-1} is increasing, $s_k(\cdot) = H^{-1}(q_k(\cdot))$ is decreasing. Moreover,

$$|q_k(c_2) - q_k(c_1)| \le \frac{5\lambda}{4}|c_2 - c_1|,$$

so Corollary 18 gives

$$\mathrm{Lip}(s_k) \le \frac{5\lambda}{4g} =: L,$$

where L is independent of k, K, and T. The baseline profile s_0 also belongs to this class.

We next verify Assumption 2. For each round t, let B_t be a standard Brownian motion with these motions independent across rounds and independent of all clean features and labels, and define

$$N_t(c) := B_t(s_k(c)).$$

Then $N_t(c) \sim \mathcal{N}(0, s_k(c))$. If $c_1 \le c_2$, then $s_k(c_1) \ge s_k(c_2)$ and

$$N_t(c_1) = N_t(c_2) + \big(B_t(s_k(c_1)) - B_t(s_k(c_2))\big),$$

where the increment is independent of $N_t(c_2)$ and has variance $s_k(c_1) - s_k(c_2)$.

Write $L_k(c, \nu) := F(a, s_k(c), \nu) + \lambda c$ and $L_k^\star(c) := \min_{|\nu| \le S} L_k(c, \nu)$. By (21), the optimized population loss in alternative k is exactly

$$L_k^\star(c) = H(s_k(c)) + \lambda c = C - \delta_K \phi_k(c).$$

Hence

$$L_k^\star := \min_{c \in [0,1]} L_k^\star(c) = C - \delta_K,$$

attained at z_k, and for every $c \notin A_k$,

$$L_k^\star(c) - L_k^\star = \delta_K. \tag{22}$$

Fix any possibly randomized policy π. It remains to show that the learner cannot identify the hidden region A_k quickly enough. We work under the stronger feedback model in which X_t is revealed after each round. Let P_0^T and P_k^T denote the laws of the complete enhanced transcript under the baseline and alternative k, respectively. Conditional on the past and on the chosen payment c_t, the clean-feature law and the conditional label law are identical under P_0^T and P_k^T. Once X_t is revealed, the noise realization $N_t(c_t) = \widehat{X}_t - X_t$ is observed. Therefore the one-step conditional KL divergence equals

$$D_{\mathrm{KL}}(\mathcal{N}(0, s_0(c_t)) \,\|\, \mathcal{N}(0, s_k(c_t))).$$

By Lemma 19, for any $u, v \ge s_{\min}$,

$$D_{\mathrm{KL}}(\mathcal{N}(0, u) \,\|\, \mathcal{N}(0, v)) \le \frac{(u - v)^2}{4 s_{\min}^2}.$$

Together with (15),

$$|s_0(c) - s_k(c)| \le \frac{\delta_K}{g} \phi_k(c).$$

Thus

$$D_{\mathrm{KL}}(\mathcal{N}(0, s_0(c)) \,\|\, \mathcal{N}(0, s_k(c))) \le C_1 \delta_K^2 \mathbf{1}\{c \in A_k\}, \qquad C_1 := \frac{1}{4 s_{\min}^2 g^2}.$$

Define

$$N_k(T) := \sum_{t=1}^{T} \mathbf{1}\{c_t \in A_k\}.$$

The chain rule for relative entropy applied to the adaptive transcript gives

$$D_{\mathrm{KL}}(P_0^T \| P_k^T) \le C_1 \delta_K^2 \mathbb{E}_0[N_k(T)].$$

No additional term arises from adaptive action selection because, conditional on a realized history, the same randomized decision rule is used in both environments.

Since the A_k are pairwise disjoint,

$$\sum_{k=1}^{K} N_k(T) \le T.$$

Hence there exists $k^\star \in [K]$ such that

$$\mathbb{E}_0[N_{k^\star}(T)] \le \frac{T}{K}.$$

For this index,

$$D_{\mathrm{KL}}(P_0^T \| P_{k^\star}^T) \le C_1 \alpha^2 \frac{T}{K^3}.$$

Choose

$$K = \left\lceil \kappa T^{1/3} \right\rceil$$

with a fixed κ sufficiently large that $C_1\alpha^2/\kappa^3 \le 1/8$. Then

$$D_{\mathrm{KL}}(P_0^T \| P_{k^\star}^T) \le \frac{1}{8},$$

and Pinsker's inequality yields

$$\mathrm{TV}(P_0^T, P_{k^\star}^T) \le \frac{1}{4}.$$

Because $0 \le N_{k^\star}(T) \le T$,

$$\mathbb{E}_{k^\star}[N_{k^\star}(T)] \le \mathbb{E}_0[N_{k^\star}(T)] + T\mathrm{TV}(P_0^T, P_{k^\star}^T) \le \frac{T}{K} + \frac{T}{4}.$$

For sufficiently large T, $K \ge 4$, so

$$\mathbb{E}_{k^\star}[N_{k^\star}(T)] \le \frac{T}{2}.$$

For every selected predictor ν_t,

$$L_{k^\star}(c_t, \nu_t) \ge L_{k^\star}^\star(c_t).$$

Therefore (22) implies

$$\begin{aligned}\mathrm{Reg}_T^\pi(s_{k^\star}) &\ge \delta_K \, \mathbb{E}_{k^\star}[T - N_{k^\star}(T)] \\ &\ge \frac{\delta_K T}{2} = \frac{\alpha T}{2K} = \Omega(T^{2/3}).\end{aligned}$$

Since π is arbitrary, this proves (20). Taking $s_{\min} = 1$ and $s_{\max} = 2$ proves Theorem 23. □

Remark 24 (Source of the lower bound)**.** *Theorem 23 continues to hold when all fixed model parameters are known and the clean feature is revealed after every round. Thus the $T^{2/3}$ obstruction is caused by localization of an unknown payment profile, rather than by structural logistic-parameter estimation or by post hoc separation of the clean feature from measurement error.*