Improved Algorithms for Online Classification with Surrogate Losses

Abed Razawy
Mathematical Institute, Leiden University
O.Razawy@math.leidenuniv.nl

Valentina Masarotto
Mathematical Institute, Leiden University
V.Masarotto@math.leidenuniv.nl

Dirk van der Hoeven
Mathematical Institute, Leiden University
Dirk@dirkvanderhoeven.com

Abstract

We study online multiclass classification with surrogate losses. We identify and exploit a structural property of margin-based classifiers: on many rounds predictions are *stable* in the sense that after an update of the parameters the algorithms do not change the predicted label on the current example. By exploiting this stability we improve upon the state of the art in several ways. We provide an improved surrogate regret bound for the perceptron, develop a parameter-free version of GAPTRON, develop improved results for the delayed feedback setting, and extend our results to the batch setting. These results are complimented by empirical evaluations.

1 Introduction

Online multiclass classification is a sequential setting where in each round $t = 1, \dots, T$ the learner maintains parameters $\boldsymbol{W}_t \in \mathcal{W} \subseteq \mathbb{R}^{K \times d}$, receives an input vector $\boldsymbol{x}_t \in \mathbb{R}^d$, predicts a label $y'_t \in [K] = \{1, \dots, K\}$ (or $y'_t \in \{-1, 1\}$ for binary prediction), and suffers loss $\mathbb{1}[y_t \neq y'_t]$, where $y_t \in [K]$ (or $y_t \in \{-1, 1\}$). In the full information setting the learner observes y_t, and in the bandit setting only $\mathbb{1}[y_t \neq y'_t]$. Both the bandit and full information feedback settings are widely applicable, from spam filtering [Hazan et al., 2016] to advertising and recommender systems [Kakade et al., 2008], and have been extensively studied.

In many applications, feedback arrives with a delay. For example, an advertisement must be shown before a user clicks or ignores it, and by the time the outcome is observed several more ads have already been served. We model this through a general delayed feedback framework: at each round t the learner receives $\boldsymbol{x}_t$, predicts y'_t, and receives feedback for the rounds $s \in \mathcal{F}_t$, where $\mathcal{F}_t = \{s : s + d_s = t\}$ where $d_s \in \mathbb{N}$ is the delay for round s. For feedback that arrives after round T we set $d_s = T - s$ without loss of generality. We write $n_t = \{s \in [t-1] : s \notin \bigcup_{r \leq t} \mathcal{F}_r\}$ for the rounds whose feedback has not yet arrived by the beginning of round t, and $D = \max_t |n_t|$ for the maximum number of pending feedbacks.

We provide algorithms with guarantees of the form $\mathbb{E}[M_T] \leq \sum_{t=1}^T \ell_t(\boldsymbol{U}) + \mathbb{E}[\mathcal{R}_T]$, where $M_T = \sum_{t=1}^T \mathbb{1}[y_t \neq y'_t]$, $\ell_t : \mathbb{R}^{K \times d} \to \mathbb{R}_+$ is a surrogate loss (e.g. multiclass hinge or logistic), $\boldsymbol{U} \in \mathbb{R}^{K \times d}$ is the comparator, and $\mathcal{R}_T$ is the surrogate regret and the expectation is over the learners' randomness.

Our improvements over prior work rest on a single observation about *stability*. Stability for us means that the algorithm's prediction on $\boldsymbol{x}_t$ would not change even if the pending updates from round t (and from any unobserved past rounds) had already been applied. Stable rounds therefore admit a free one-step lookahead in the analysis. We explain this principle in Section 2 by applying to the analysis

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

of the Perceptron [Rosenblatt, 1958, Novikoff, 1962]. Here we reveal an unexpected robustness not observed in prior work: there exist a scenario where prior bounds would be $O(\|\boldsymbol{U}\|X\sqrt{T})$ ($\|\cdot\|$ denotes the l_2 norm), where $\max_t \|\boldsymbol{x}_t\| \leq X$. We show that the perceptron in the same scenario actually guarantees a bound of $O(\|\boldsymbol{U}\|^2)$. We further exploit the stability of classification algorithms in the remainder of the paper, which leads to the following contributions:

Parameter free algorithm and improved guarantees for the multiclass hinge loss. In the delayed full information setting we provide a parameter-free algorithm with state-of-the-art surrogate regret guarantees. By *parameter-free* we mean that the algorithm does not require any input. In contrast, the state-of-the-art results of van der Hoeven [2020] require tuning an exploration parameter with knowledge of $\max_t \|\boldsymbol{x}_t\|$, and lack a delayed-feedback extension. Strictly speaking, one could use the `BOLD` algorithm of Joulani et al. [2013] and use the algorithm of van der Hoeven [2020] as a base learner to obtain a bound of order $K\|U\|^2 \max_t \|\boldsymbol{x}_t\|^2 D$. However, this requires maintaining D copies of the base algorithm.

For binary classification ($K = 2$) without delay, our framework yields a parameter-free version of van der Hoeven [2020, Theorem 2]: simultaneously for all powers $q \geq 1$ of hinge loss and all $\boldsymbol{U} \in \mathbb{R}^d$ (Corollary 2),

$$\mathbb{E}[M_T] \leq \sum_{t=1}^{T} \ell_t^q(\boldsymbol{U}) + \tfrac{3q}{2} \max_t \|\boldsymbol{x}_t\|^2 \|\boldsymbol{U}\|^2 ,$$

whereas van der Hoeven [2020] only provides results for $q = 1$.

For $K > 2$ (Corollary 4) we provide a parameter free algorithm that guarantees

$$\mathbb{E}[M_T] \leq \sum_{t=1}^{T} \ell_t(\boldsymbol{U}) + \min\Big\{ 2\|\boldsymbol{U}\|\sqrt{B_T},\ \tfrac{B_T}{A_T}\|\boldsymbol{U}\|^2 \Big\},$$

where $B_T \leq \sum_{t=1}^{T}(\|\boldsymbol{x}_t\|^2 + \sum_{s\in n_t} |\boldsymbol{x}_t^\top \boldsymbol{x}_s|)$ and $\frac{B_T}{A_T} \approx \frac{K}{T}\sum_{t=1}^{T}(\|\boldsymbol{x}_t\|^2 + \sum_{s\in n_t} |\boldsymbol{x}_t^\top \boldsymbol{x}_s|)$. Now, note that $\frac{B_T}{A_T}$ is approximately the average delay, which means that only a few large delays do not impact the bound. In contrast, by using the `BOLD` algorithm of Joulani et al. [2013], the bound scales with the largest delay.

In the delayed bandit setting (Corollary 5) our algorithm guarantees a surrogate regret bound of order $RXK\sqrt{T} + R^2X^2KD$ for $\boldsymbol{U}$ such that $\|\boldsymbol{U}\| \leq R$. This is a remarkable result because the cost of delayed feedback is additive in D rather than $O(\sqrt{DT})$, which would be expected based on results for bandits with delayed feedback (see for example Zierahn et al. [2025]) and is obtained in related work on online classification with surrogate losses (see below). We achieve this result by carefully exploiting the stability of online classification algorithms.

Finally, the fact that our algorithm is parameter-free yields a dynamic surrogate regret bound (Corollary 6) complementary to known literature.

Improved guarantees for the multiclass logistic loss. The logistic loss is a Fenchel–Young loss Sakaue et al. [2024], a class of surrogates whose curvature Sakaue et al. [2024] and Shibukawa et al. [2025] exploit to obtain constant surrogate regret in the full-information setting and $\sqrt{KT}$-rate regret in the bandit setting. We provide algorithms for the multiclass logistic loss in the delayed full-information and delayed bandit settings, with two structural improvements over prior work.

In the delayed full-information setting, Shibukawa et al. [2025] provide a bound of order $\|\boldsymbol{U}\|^2X^2D^2$ and a bound of order $\|\boldsymbol{U}\|^2X^2D$ by using the `BOLD` algorithm of Joulani et al. [2013]. In contrast, our algorithm guarantees $\mathbb{E}[M_T] \leq \sum_{t=1}^{T} \ell_t(\boldsymbol{U}) + \frac{(1+2D)\,X^2\,\|\boldsymbol{U}\|^2}{2(\log 2)(1-\log 2)}$, which holds for any $\boldsymbol{U} \in \mathbb{R}^{K\times d}$ (Theorem 7).

In the delayed bandit setting Shibukawa et al. [2025] guarantees a bound of order $O(RX\sqrt{(K+D)T})$. We considerably improve upon this guarantee and provide a surrogate regret bound of order $RX\sqrt{KT} + DX^2R^2$ (Theorem 8). The bandit cost $O(\sqrt{KT})$ and delay cost $O(D)$ are independent, so for any fixed D the delay contribution remains bounded regardless of T.

Beyond these theoretical improvements we also demonstrate empirical improvements. Our algorithm randomizes less than prior work: van der Hoeven [2020], van der Hoeven et al. [2021], Sakaue et al. [2024], Shibukawa et al. [2025] randomize on every round, while our stability-based algorithm predicts with what it considers the most likely outcome deterministically on stable rounds and randomizes only when a pending update could change the most likely outcome. Our experiments (Section 6) show this reduced dependence on randomness improves empirical performance.

Batch Setting Our techniques readily extend to regularized empirical risk minimization (ERM). Given an i.i.d. sample $S = \{(\boldsymbol{x}_t, y_t)\}_{t=1}^T$ and a regularized ERM output $\widehat{\boldsymbol{W}}$, stability and selective randomization yield $\mathbb{E}_{S,y'}[\mathbb{1}[y \neq y'(\boldsymbol{x}, \widehat{\boldsymbol{W}})]] \leq \mathbb{E}[\ell(\langle \boldsymbol{W}_\star, \boldsymbol{x}\rangle, y)] + c_\ell X^2 \|\boldsymbol{W}_\star\|^2/(T+1)$ for both the binary hinge loss (Theorem 9, $c_\ell = 2$) and the binary logistic loss (Theorem 10, $c_\ell = \frac{1}{\log 2 \log(\sqrt{2}+1)}$), for any $\boldsymbol{W}_\star \in \mathbb{R}^d$.

Organization. The rest of the paper is organized as follows. Section 2 develops the stability principle. Section 3 derives consequences for the multiclass hinge loss; Section 4 for the logistic loss. In Section 5 we show how stability extends to the batch setting. Section 6 provides empirical evaluation. However, we first discuss related work.

1.1 Further Related Work

The first work on online classification with surrogate losses was on the Perceptron [Rosenblatt, 1958, Novikoff, 1962]. It's multiclass variants for the full information setting were developed by [Crammer and Singer, 2003, Fink et al., 2006] and guarantee that $\mathcal{R}_T = O(\|\boldsymbol{U}\| X \sqrt{T})$. Kakade et al. [2008] developed a bandit version of the multiclass perceptron which guarantees $\mathcal{R}_T = O((KRX)^{1/3} T^{2/3})$. Cesa-Bianchi et al. [2005], Crammer et al. [2009] provide second-order algorithms for full information binary classification with improved guarantees at the cost of increased computational complexity. Specifically, these algorithms guarantee $\mathcal{R}_T = O(d \log(1 + R^2 X^2 T))$ with a quadratic version of the hinge loss at the cost of a $O(d^2)$ per round computational complexity. Additional results for these algorithms are obtained by Orabona et al. [2012]. A bandit version of these second-order algorithms was developed by Beygelzimer et al. [2017].

For the logistic loss [Foster et al., 2018] develop a computationally inefficient algorithm which guarantees $\mathcal{R} = O(dK \log(1 + RT))$ for the full information setting and $\mathcal{R} = O(K\sqrt{d \log(1 + RT)})$ for the bandit setting, which considerably improved upon the guarantees of `Newtron` [Hazan and Kale, 2011] at the cost of polynomial in T and d computational complexity.

In the literature different notions of stability have been exploited, see [Shalev-Shwartz et al., 2010, Bousquet and Elisseeff, 2002, Koren and Levy, 2015, Bousquet et al., 2020, Klochkov and Zhivotovskiy, 2021]. The idea in these works is to exploit the fact that most learning algorithms do not change their predictions very much if the one (x, y) pair in their training data is replaced by a different pair. Our notion is somewhat unique in the sense that we define stability for only a single feature vector $\boldsymbol{x}_t$, rather than for all possible (x, y) pairs.

Notation Lowercase bold letters ($\boldsymbol{x}, \boldsymbol{g}$) denote vectors and uppercase bold letters ($\boldsymbol{W}, \boldsymbol{U}$) denote weight vectors or matrices; $\|\cdot\|$ denotes the Euclidean norm for vectors and the Frobenius norm for matrices, and $\langle \cdot, \cdot \rangle$ the corresponding inner product. We write $\mathbb{1}[A]$ for the indicator of an event A, $\mathbf{1}_K \in \mathbb{R}^K$ for the all-ones vector, $\boldsymbol{e}_k$ for the k-th standard basis vector, and $\mathrm{sign}(\cdot)$ for the sign function with the convention $\mathrm{sign}(0) = -1$. We write $(\cdot)_+ = \max\{0, \cdot\}$. The rows of a matrix $\boldsymbol{W} \in \mathbb{R}^{K \times d}$ are denoted $\boldsymbol{W}^1, \ldots, \boldsymbol{W}^K$. For any round t, $\mathbb{E}_t[\,\cdot\,]$ denotes the conditional expectation given the predictions $y'_1, \ldots, y'_{t-1}$.

2 Stability in Regret Analysis and Algorithm Design

This section develops the notion of *stability* of online predictions as a tool for obtaining refined, data-dependent surrogate regret bounds and for designing improved learning algorithms.

The central idea is the following. In online classification, the zero-one loss is typically bounded by a surrogate loss: $\mathbb{1}[y'_t \neq y_t] \leq \ell_t(\boldsymbol{W}_t)$, where $\boldsymbol{W}_t$ are the parameters of the algorithm in round t. On rounds where the (deterministic) prediction is *stable*, meaning the algorithm's update does not

Table 1: Worst case upper bounds on the surrogate regret for online multiclass classification with delayed and bandit feedback. All bounds in this table suppress dependence on $\|\boldsymbol{U}\|$ and $\max_t \|\boldsymbol{x}_t\|$.

Reference	Feedback	Loss	Surrogate regret bound
[Joulani et al., 2013, Shibukawa et al., 2025]	Full-info & Delayed	logistic	$O(D)$
[Shibukawa et al., 2025]	Full-info & Delayed	logistic	$O(D^2)$
[Shibukawa et al., 2025]	Bandit & Delayed	logistic	$O(\min\{\sqrt{(K+D)T}, D^{1/3}T^{2/3}\})$
This work (Corollary 4)	Full-info & Delayed	hinge	$O(KD+K)$
This work (Corollary 5)	Bandit & Delayed	hinge	$O(K\sqrt{T}+KD)$
This work (Theorem 7)	Full-info & Delayed	logistic	$O(D)$
This work (Theorem 8)	Bandit & Delayed	logistic	$O(\sqrt{KT}+D)$

change its prediction on the current example, we can replace the current iterate by the updated one:

$$\mathbb{1}[y'_t \neq y_t] \leq \ell_t(\boldsymbol{W}_{t+1}). \tag{1}$$

This amounts to a free one-step lookahead: stable rounds allow us to charge mistakes to the post-update model, which has already incorporated feedback. Since the surrogate loss at $\boldsymbol{W}_{t+1}$ benefits from the update, stable rounds contribute less — or nothing — to the overall surrogate regret.

Before we formalize our notion of stability, we first explore its use in the binary classification setting. We begin by illustrating the value of stability in the Perceptron algorithm, where it admits an explicit characterization in terms of the margin. This characterization motivates algorithmic modifications that lead to improvements via the `GAPTRON` algorithm of van der Hoeven [2020]. Throughout this section we use the convention that $\mathcal{W} = \mathbb{R}^d$.

2.1 Application to Perceptron

We illustrate the main idea in a warm-up setting: exploiting stability leads to sharper surrogate regret bounds for the binary Perceptron. For completeness, the algorithm is included below (Algorithm 1). At round t, the Perceptron updates its weights whenever the margin condition $y_t\langle \boldsymbol{W}_t, \boldsymbol{x}_t\rangle < 1$ in line 5 of Algorithm 1 is satisfied. We write this compactly as $\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + \tau_t\, y_t \boldsymbol{x}_t$, where $\tau_t = \mathbb{1}[y_t\langle \boldsymbol{W}_t, \boldsymbol{x}_t\rangle < 1]$ indicates whether an update occurs. Let $\ell_t^q(\boldsymbol{W}) = (1 - y_t\langle \boldsymbol{W}, \boldsymbol{x}_t\rangle)_+^q$ denote the q-th power of the hinge loss. The classical analysis gives, for all $\boldsymbol{U} \in \mathbb{R}^d$,

$$M_T - \sum_{t=1}^{T} \ell_t^1(\boldsymbol{U}) \leq \|\boldsymbol{U}\| X \sqrt{\sum_{t=1}^{T} \tau_t}\,. \tag{2}$$

The proof is identical to that of Orabona [2025, Theorem 8.2], which uses the threshold $y_t\langle \boldsymbol{W}_t, \boldsymbol{x}_t\rangle < 0$. We show how to improve this bound by exploiting stability to identify rounds that do not contribute to surrogate regret.

Algorithm 1 Binary Perceptron

1: $\boldsymbol{W}_1 = \boldsymbol{0}$
2: **for** $t = 1, \ldots, T$ **do**
3: receive $\boldsymbol{x}_t$
4: predict $y'_t = \operatorname{sign}(\langle \boldsymbol{W}_t, \boldsymbol{x}_t\rangle)$
5: **if** $y_t\langle \boldsymbol{W}_t, \boldsymbol{x}_t\rangle < 1$ **then**
6: $\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + y_t \boldsymbol{x}_t$
7: **end if**
8: **end for**

The first observation to make is that the Perceptron is equivalent to OGD on the linearized losses $\widetilde{\ell}_t(\boldsymbol{W}) = \tau_t(1 - \langle \boldsymbol{W}, y_t \boldsymbol{x}_t\rangle)$. The second observation is that by definition of τ_t, we have that for all $\eta \in (0, 1]$, $\operatorname{sign}(\eta\langle \boldsymbol{W}_t, \boldsymbol{x}_t\rangle) = \operatorname{sign}(\langle \boldsymbol{W}_t, \boldsymbol{x}_t\rangle)$ and thus $\mathbb{1}[y'_t \neq y_t] \leq \widetilde{\ell}_t(\eta \boldsymbol{W}_t)$. This second observations allows us to analyze the Perceptron using OGD with any learning rate $\eta \in (0, \frac{1}{2}]$. A crucial observation that we make in the proof of our main result in this section is the following. If $\operatorname{sign}(\langle \boldsymbol{W}_t, \boldsymbol{x}_t\rangle) = \operatorname{sign}(\langle \boldsymbol{W}_{t+1}, \boldsymbol{x}_t\rangle)$ then for all $\eta \in (0, \frac{1}{2\max_t \|x_t\|_2^2}] : \mathbb{1}[y'_t \neq y_t] \leq \widetilde{\ell}_t(\eta \boldsymbol{W}_{t+1})$. Denote by $\mathcal{E}_t = \mathbb{1}[\operatorname{sign}(\langle \boldsymbol{W}_t, \boldsymbol{x}_t\rangle) \neq \operatorname{sign}(\langle \boldsymbol{W}_{t+1}, \boldsymbol{x}_t\rangle)]$. By the above and standard results for OGD,

we show that simultaneously for all $\eta \in \left(0, \frac{1}{2\max_t \|x_t\|^2}\right]$:

$$M_T \le \sum_{t=1}^{T} \left((1-\mathcal{E}_t)\widetilde{\ell}_t(\eta \boldsymbol{W}_{t+1}) + \mathcal{E}_t \widetilde{\ell}_t(\eta \boldsymbol{W}_t)\right) \le \sum_{t=1}^{T} \ell_t^1(\boldsymbol{U}) + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \frac{\eta}{2}\sum_{t=1}^{T} \tau_t \mathcal{E}_t \|\boldsymbol{x}_t\|^2 .$$

Taking the infimum over $\eta \in \left(0, \frac{1}{2\max_t \|x_t\|_2^2}\right]$ leads to the bound for $q = 1$. For $q > 1$ we do a few more manipulations (see Appendix A.1), which leads to the following result.

Theorem 1 (Improved bound for binary Perceptron)**.** *For any $\boldsymbol{U} \in \mathbb{R}^d$ and $q \ge 1$, the Perceptron (Algorithm 1) satisfies: $M_T - \sum_{t=1}^{T} \ell_t^q(\boldsymbol{U}) \le q\|\boldsymbol{U}\| \sqrt{\sum_{t=1}^{T} \tau_t \mathcal{E}_t \|\boldsymbol{x}_t\|^2} + q\|\boldsymbol{U}\|^2(1+\max_t \|\boldsymbol{x}_t\|^2)$.*

The proof is deferred to Appendix A.1. This bound reveals a hidden robustness of the perceptron algorithm: the data can be non-separable, yet as long as many predictions are stable the surrogate regret bound remains small. In Appendix A.1 we also provide a simple example where Theorem 1 shows that the surrogate regret of the perceptron is $O(\|\boldsymbol{U}\|^2)$ rather than the $O(\|\boldsymbol{U}\|\sqrt{T})$ surrogate regret bound provided by classical results. To illustrate the improvement, consider the nonseparable sequence $\boldsymbol{x}_t = \frac{1}{2}$, with $y_t = -1$ every fifth round and $y_t = +1$ otherwise. The Perceptron continues to update on $\Theta(T)$ rounds, but only its first update changes the prediction: $\mathcal{E}_1 = 1$ and $\mathcal{E}_t = 0$ for all $t > 2$. Consequently, $\sum_{t=1}^{T} \tau_t \mathcal{E}_t \|\boldsymbol{x}_t\|^2 = \frac{1}{4}$, and Theorem 1 with $q = 1$ gives

$$M_T - \sum_{t=1}^{T} \ell_t^1(\boldsymbol{U}) \le \frac{1}{2}\|\boldsymbol{U}\| + \frac{5}{4}\|\boldsymbol{U}\|^2.$$

This upper bound is independent of T, whereas the classical update-count bound grows as $O(\|\boldsymbol{U}\|\sqrt{T})$. Thus, even on nonseparable data, repeated updates need not accumulate surrogate regret when predictions remain stable. In what follows, we exploit this observation to design improved algorithms.

In what follows, we exploit this observation to design improved algorithms.

2.2 Stability as a Design Principle

The Perceptron analysis in the previous section shows that stable rounds do not contribute to surrogate regret, but the bound still accumulates surrogate regret over unstable rounds. In this section we eliminate this residual cost through algorithm design.

The key idea is to exploit the gap between the surrogate loss and the zero-one loss. On unstable rounds, we randomize our predictions to develop a gap between the surrogate loss and the expected 0-1 loss, which is later exploited in the analysis. This idea was first explored by van der Hoeven [2020] in the GAPTRON algorithm.Combining this idea with the stability principle leads to Algorithm 2. Algorithm 2 can be seen as a *parameter-free* version of GAPTRON for the binary hinge loss. Algorithm 2 requires no input and maintains parameters $\boldsymbol{W}_t \in \mathbb{R}^d$ that are very similar to the parameter of the perceptron. Algorithm 2 predicts with $\operatorname{sign}(\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle)$ if $|\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle| > \|\boldsymbol{x}_t\|^2$ and samples a prediction uniformly at random otherwise. Ideally we would like to check if $\operatorname{sign}(\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle) = \operatorname{sign}(\langle \boldsymbol{W}_{t+1}, \boldsymbol{x}_t \rangle)$. However, since $\boldsymbol{W}_{t+1}$ uses y_t we can not compute it to choose our prediction in round t. Instead, we use the proxy $|\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle| > \|\boldsymbol{x}_t\|^2$, which implies that $\operatorname{sign}(\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle) = \operatorname{sign}(\langle \boldsymbol{W}_{t+1}, \boldsymbol{x}_t \rangle)$ by definition of $\boldsymbol{W}_{t+1}$. To understand the reason that we sample $\widehat{y}_t$ uniformly at random if $|\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle| \le \|\boldsymbol{x}_t\|^2$ let us restrict to $q = 1$. We essentially show that for sufficiently small η, $\frac{1}{2} = \mathbb{E}[\mathbb{1}[y_t \ne y_t']] \le \ell_t^1(\eta \boldsymbol{W}_t) - \eta\|\boldsymbol{x}_t\|^2$ if $|\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle| \le \|\boldsymbol{x}_t\|^2$. Combined with the fact that on stable rounds (i.e. $|\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle| > \|\boldsymbol{x}_t\|^2$) we have that $\mathbb{1}[\widehat{y}_t \ne y_t] \le \ell_t^1(\boldsymbol{W}_{t+1}) \approx \ell_t^1(\boldsymbol{W}_t) - \|\boldsymbol{x}_t\|^2$, we have shown that our predictions satisfy $\mathbb{1}[\widehat{y}_t \ne y_t] \lesssim \ell_t^1(\eta \boldsymbol{W}_t) - \eta\|\boldsymbol{x}_t\|^2$. After this it is relatively straightforward to prove constant (in T) surrogate regret by exploiting standard surrogate regret bounds for online gradient descent, which we demonstrate in Theorem A.1.

A simple Corollary then shows that Algorithm 2 has constant (in T) surrogate regret and is simultaneously robust to excessively large $\|\boldsymbol{U}\|$, without knowledge of $\|\boldsymbol{U}\|$. Proofs are deferred to Appendix A.2.

Corollary 2. *With $A_T = \sum_{t=1}^{T} a_t \|\boldsymbol{x}_t\|^2$, and $N_T = \sum_{t=1}^{T} a_t$, and a_t as in line 5, Algorithm 2 guarantees, for any $\boldsymbol{U} \in \mathbb{R}^d$ and $q \ge 1$: $\mathbb{E}[M_T] \le \sum_{t=1}^{T} \ell_t^q(\boldsymbol{U}) + q \min\left\{\|\boldsymbol{U}\|\sqrt{3A_T}\,, \frac{3}{2}\|\boldsymbol{U}\|^2 \frac{A_T}{N_T}\right\}$*

Algorithm 2 Parameter-Free Gaptron for the Binary Hinge Loss

1: Initialize $\boldsymbol{W}_1 = 0$
2: **for** $t = 1$ **to** T **do**
3: Receive $\boldsymbol{x}_t$
4: Let $y_t^\star = \text{sign}(\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle)$
5: Set $a_t = \mathbb{1}[|\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle| \leq \|\boldsymbol{x}_t\|_2^2]$
6: Set $\boldsymbol{y}_t^\star$ if $a_t = 0$ and sample y_t' uniformly at random otherwise.
7: Receive $\mathbb{1}[y_t' \neq y_t]$
8: Update $\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + y_t \boldsymbol{x}_t \mathbb{1}[y_t \langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle \leq \|\boldsymbol{x}_t\|_2^2]$
9: **end for**

Compared to the Perceptron surrogate regret bound (Theorem 1), we have now obtained a constant surrogate regret bound yet we still maintain the robustness to large $\|\boldsymbol{U}\|$. This latter feature is not present in prior work. For example, [van der Hoeven, 2020] only guarantees a bound of order $\|\boldsymbol{U}\|^2 X^2$, where X is an upper bound on $\max_t \|\boldsymbol{x}_t\|$ which needs to be known by the learner. Note that Algorithm 2 does not need this knowledge: it is completely parameter free. Note that $A_T \leq TX^2$, but in general it can be much smaller.

3 Multiclass Hinge Loss

We extend the stability-based analysis to multiclass classification with K classes under the multiclass hinge loss of Crammer and Singer [2002] (Equation 7), in the delayed full information and delayed bandit feedback settings. Throughout, $\mathcal{W} = \mathbb{R}^{K \times d}$. Our algorithm (Algorithm 3) extends Algorithm 2 to multiclass and improves over [van der Hoeven, 2020, Theorems 2 and 5] in three ways: it handles delayed feedback and dynamic regret, randomizes only over a competitive subset of labels rather than uniformly over $[K]$, and is parameter-free. The non-delayed setting is recovered as the special case $D = 0$ (Appendix B.1, B.2), with algorithmic details in Appendix B.

We formalize stability in an online linear prediction setting: at round t, the learner maintains $\boldsymbol{W}_t \in \mathbb{R}^{K \times d}$, receives $\boldsymbol{x}_t$, predicts $y_t^\star = \arg\max_k \langle \boldsymbol{W}_t^k, \boldsymbol{x}_t \rangle$, and after observing y_t updates to $\boldsymbol{W}_{t+1}$.

Definition 3 (Stability)**.** A prediction at round t is *stable* if the deterministic prediction is invariant under the update for every possible label realization:

$$\arg\max_{k \in [K]} \langle \boldsymbol{W}_t^k, \boldsymbol{x}_t \rangle = \arg\max_{k \in [K]} \langle \boldsymbol{W}_{t+1}^k, \boldsymbol{x}_t \rangle \qquad \text{for every } y_t \in [K].$$

On stable rounds the deterministic prediction at $\boldsymbol{W}_{t+1}$ matches that at $\boldsymbol{W}_t$, so (1) applies. As mentioned in Section 2, we cannot observe stability in the sense of Definition 3 at prediction time. Therefore, we make use of proxy β_t, which results in Algorithm 3.

The remainder of this section analyzes Algorithm 3 across two settings: Section 3.1 analyzes the delayed full-information setting, and Section 3.2 the delayed bandit setting. Section 3.3 shows that we can use parameter-freeness to obtain dynamic regret bounds.

3.1 Delayed Full Information Feedback for the Multiclass Hinge Loss

Under delayed full-information feedback, the update at round t uses the gradients of all rounds in $\mathcal{F}_t$ — those whose feedback y_s arrives simultaneously. Stability (Definition 3) therefore depends on whether the total pending update, not a single round's gradient, suffices to alter the prediction on $\boldsymbol{x}_t$. This motivates the delay-adjusted instability threshold: $\beta_t = \|\boldsymbol{x}_t\|^2 + 2\sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle|$, which enlarges the non-delayed threshold $\|\boldsymbol{x}_t\|^2$ by the cross-terms between pending and current features. Unlike van der Hoeven [2020], our algorithm is parameter-free and uses selective rather than uniform exploration (line 6-7, Algorithm 3). Combined with Theorem B.1, parameter-freeness yields the following bound; proofs and details are in Appendix B.1.

Parameter-freeness in the full-information setting follows from a scaling argument. For the analysis, introduce auxiliary iterates $\boldsymbol{W}_t(\eta) = \eta \boldsymbol{W}_t$ for any $\eta > 0$. Since $m_t(\boldsymbol{W}_t(\eta), y) = \eta m_t(\boldsymbol{W}_t, y)$,

Algorithm 3 Stability Enhanced GAPTRON with Delayed Feedback For Multiclass Hinge Loss

Require: Exploration rate $\gamma \in [0,1]$, instability threshold

1: Initialize $\boldsymbol{W}_1 = \boldsymbol{0}$

2: **for** $t = 1$ **to** T **do**

3: Obtain $\boldsymbol{x}_t$ and set the instability threshold

$$\beta_t := \begin{cases} \|\boldsymbol{x}_t\|^2 + 2\sum_{s\in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| & \text{(full information)}, \\ \frac{K}{\gamma}\|\boldsymbol{x}_t\|^2 + 2\sum_{s\in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| & \text{(bandit)}, \\ 2\|\boldsymbol{x}_t\|^2 & \text{(dynamic regret)}. \end{cases}$$

4: Compute the margin $m_t(\boldsymbol{W}_t, y) = \langle \boldsymbol{W}_t^y, \boldsymbol{x}_t\rangle - \max_{k\neq y}\langle \boldsymbol{W}_t^k, \boldsymbol{x}_t\rangle$ for all $y \in [K]$

5: Let $y_t^\star = \arg\max_k m_t(\boldsymbol{W}_t, k)$, $m_t^\star = \max_k m_t(\boldsymbol{W}_t, k)$, and $a_t = \mathbb{1}[m_t^\star \le \beta_t]$

6: Compute $\mathcal{U}_t = \{y : m_t(\boldsymbol{W}_t, y) > -\beta_t\}$

7: Set $\boldsymbol{p}_t' = (1-\gamma)\big[(1-a_t)\,\boldsymbol{e}_{y_t^\star} + \frac{a_t}{|\mathcal{U}_t|}\,\boldsymbol{1}_{\mathcal{U}_t}\big] + \frac{\gamma}{K}\,\boldsymbol{1}$

8: Predict $y_t' \sim \boldsymbol{p}_t'$

9: Observe the set $\mathcal{F}_t$ of rounds whose feedback arrives during round t

10: For each $s \in \mathcal{F}_t$, compute $b_s := \mathbb{1}[a_s = 1 \text{ or } y_s^\star \neq y_s]$, $k_s^\star \in \arg\max_{k\neq y_s}\langle \boldsymbol{W}_s^k, \boldsymbol{x}_s\rangle$

- Full information / dynamic regret: $\boldsymbol{g}_s := b_s(\boldsymbol{e}_{k_s^\star} - \boldsymbol{e}_{y_s}) \otimes \boldsymbol{x}_s$
- Bandit: $\widehat{\boldsymbol{g}}_s := \frac{\mathbb{1}[y_s' = y_s]}{p_s'(y_s)}\,\boldsymbol{g}_s$

11: Update

- Full information / dynamic regret: $\boldsymbol{W}_{t+1} = \boldsymbol{W}_t - \sum_{s\in\mathcal{F}_t}\boldsymbol{g}_s$
- Bandit: $\boldsymbol{W}_{t+1} = \boldsymbol{W}_t - \sum_{s\in\mathcal{F}_t}\widehat{\boldsymbol{g}}_s$

12: **end for**

positive scaling preserves the greedy label. Moreover, scaling the instability threshold to $\eta\beta_t$ preserves both the instability indicator and the competitive set:

$$\mathbb{1}[\eta m_t^\star \le \eta\beta_t] = a_t, \qquad \{y : \eta m_t(\boldsymbol{W}_t, y) > -\eta\beta_t\} = \mathcal{U}_t.$$

Hence the prediction distribution and update directions remain unchanged. We can therefore apply the linearized OGD analysis to the same predictions for every $\eta > 0$, and optimize the resulting bound over η in hindsight. The learner itself requires no knowledge of the comparator norm, feature-norm bound, horizon, or maximum number of pending feedbacks. This scaling argument does not extend directly to the logistic algorithm, whose softmax probabilities depend on the magnitudes of the scores.

Corollary 4. *Under Theorem B.1's conditions, Algorithm 3 satisfies for any $\boldsymbol{U} \in \mathbb{R}^{K\times d}$:* $\mathbb{E}[M_T] \le \sum_{t=1}^T \ell_t(\boldsymbol{U}) + \min\Big\{2\|\boldsymbol{U}\|\sqrt{B_T},\ \frac{B_T}{A_T}\|\boldsymbol{U}\|^2\Big\}$, *where* $A_T = \sum_{t=1}^T \frac{a_t\,\mathbb{1}[y_t\in\mathcal{U}_t]}{|\mathcal{U}_t|}$; $B_T = \sum_{t=1}^T a_t\,\mathbb{1}[y_t\in\mathcal{U}_t]\big(\|\boldsymbol{x}_t\|^2 + 2\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|\big)$, *and ℓ_t denotes the multiclass hinge loss (7).*

Two features of the bound are notable. First, it is robust to excessive delays: a few high-delay rounds inflate the BOLD regret substantially but barely affect ours. Second, while van der Hoeven [2020]'s bound always scales with K, ours scales with $|\mathcal{U}_t|$ — and can be considerably smaller when only a few classes are competitive each round. See Appendix B.1 for details.

3.2 Delayed Bandit Feedback for the Multiclass Hinge Loss

We now turn to the bandit setting, where the learner observes only $\mathbb{1}[y_t \neq y_t']$, possibly with delay. OGD instead uses the importance-weighted estimator (line 11, Algorithm 3), which is unbiased ($\mathbb{E}_t[\widehat{\boldsymbol{g}}_t] = \boldsymbol{g}_t$) and computable from bandit feedback. Importance weighting requires positive sampling probability on every class, so we set $\gamma > 0$, ensuring $p_t'(y) \ge \gamma/K$. Bandit variance and delay cross-terms enter the instability threshold additively, $\beta_t = \frac{K}{\gamma}\|\boldsymbol{x}_t\|^2 + 2\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|$, so the delay cost in the bound is identical to the full-information case. The following result, with proofs in Appendix B.2, follows directly from Theorem B.2.

Corollary 5. *Under Theorem B.2's conditions, for any comparator $\boldsymbol{U} \in \mathbb{R}^{K\times d}$, $\gamma \in (0,1]$, Algorithm 3 satisfies* $\mathbb{E}[M_T] \le \mathbb{E}\Big[\sum_{t=1}^T \ell_t(\boldsymbol{U})\Big] + \min\Big\{\|\boldsymbol{U}\|\sqrt{2B_T},\ \frac{B_T}{2A_T}\|\boldsymbol{U}\|^2\Big\} + \gamma\,\frac{K-1}{K}\,T$, *where*

$A_T = \sum_{t=1}^{T} \frac{a_t \mathbb{1}[y_t \in \mathcal{U}_t]}{|\mathcal{U}_t|}$; $B_T = \sum_{t=1}^{T} a_t \mathbb{1}[y_t \in \mathcal{U}_t]\left[\left(\frac{K}{\gamma} + \frac{|\mathcal{U}_t|}{1-\gamma}\right)\|\boldsymbol{x}_t\|^2 + 4\sum_{s\in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|\right]$, *and* ℓ_t *denotes the multiclass hinge loss* (7)

Tuning γ for any $R \geq \|\boldsymbol{U}\|$ recovers the worst-case rate $XR\sqrt{2K(K-1)T} + 2KDX^2R^2$ (see Appendix B.2): bandit and delay costs enter additively.

3.3 Dynamic Surrogate Regret for the Multiclass Hinge Loss

The bounds so far compare against a fixed comparator. Parameter-freeness extends Corollary 4 to a comparator sequence $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_T$ by a minor algorithmic change: doubling the threshold to $\beta_t = 2\|\boldsymbol{x}_t\|^2$ absorbs an extra gradient cost, yielding a dynamic regret bound that interpolates between the static guarantee and the variability of the sequence.

Corollary 6 (Dynamic Surrogate Regret for the Multiclass Hinge Loss)**.** *Suppose* $D = 0$*. Let* $\gamma = 0$*,* ℓ_t *the multiclass hinge loss* (7)*, and* $\mathcal{A} = \{t : a_t = 1, y_t \in \mathcal{U}_t\}$*. Then for any comparator sequence* $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_T \in \mathbb{R}^{K\times d}$*, Algorithm 3 achieves*

$$\mathbb{E}[M_T] \leq \sum_{t=1}^{T} \ell_t(\boldsymbol{U}_t) + \min\Bigg\{ \inf_{\bar{\boldsymbol{U}}\in\mathbb{R}^{K\times d}} \frac{2\sum_{t\in\mathcal{A}}\|\boldsymbol{x}_t\|^2}{\sum_{t\in\mathcal{A}}|\mathcal{U}_t|^{-1}}\left(\|\bar{\boldsymbol{U}}\|^2 + \sum_{t=1}^{T}\|\boldsymbol{U}_t - \bar{\boldsymbol{U}}\|^2\right), \inf_{\bar{\boldsymbol{U}}\in\mathbb{R}^{K\times d}} 2\sqrt{2}\,X\sqrt{\left(\|\bar{\boldsymbol{U}}\|^2 + \sum_{t=1}^{T}\|\boldsymbol{U}_t - \bar{\boldsymbol{U}}\|^2\right)\sum_{t=1}^{T} a_t}\Bigg\}.$$

To our knowledge, this is the first dynamic surrogate regret bound of this form for the multiclass hinge loss. The bound is small when the comparator sequence is tightly clustered around its mean, with the proxy $\bar{\boldsymbol{U}}$ optimized in the analysis (no prior knowledge of the sequence required). Proof and discussion in Appendix B.2.

4 Logistic Loss

We now turn to the multiclass logistic loss, defined as

$$\ell_t(\boldsymbol{W}) = -\log_2\big(\sigma(\boldsymbol{W}, \boldsymbol{x}_t, y_t)\big), \tag{3}$$

where $\sigma(\boldsymbol{W}, \boldsymbol{x}, k) = \frac{\exp(\langle \boldsymbol{W}^k, \boldsymbol{x}\rangle)}{\sum_{j=1}^{K}\exp(\langle \boldsymbol{W}^j, \boldsymbol{x}\rangle)}$ is the softmax function. We write $\widetilde{\boldsymbol{p}}_t = (\widetilde{p}_t(1)), \ldots, \widetilde{p}_t(K))^\top$ with $\widetilde{p}_t(k) = \sigma(\boldsymbol{W}_t, \boldsymbol{x}_t, k)$. The full algorithm is given in Algorithm 4. Throughout we assume that $\mathcal{W} = \mathbb{R}^{K\times d}$. The algorithm is discussed in detail in Appendix C.

4.1 Delayed Full Information Feedback for the Logistic Loss

As in the hinge loss setting (Section 3.1), delayed OGD introduces cross-terms $\eta\sum_{s\in n_t}\langle \boldsymbol{g}_s, \boldsymbol{g}_t\rangle$. The gradient structure (11) bounds these by $\frac{2}{(\log 2)^2}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|$ (Appendix C.1), motivating the instability threshold $\beta_t = \frac{2\eta}{\log 2}\|\boldsymbol{x}_t\|^2 + \frac{2\eta}{(\log 2)^2}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|$.

Theorem 7 (Delayed Full-Information Mistake Bound, Logistic Loss)**.** *Let* ℓ_t *be the logistic loss* (3)*,* $\|\boldsymbol{x}_t\| \leq X$*,* $\gamma = 0$*,* $D = \max_t |n_t|$*. Then for any* $\boldsymbol{U} \in \mathbb{R}^{K\times d}$*, Algorithm 4 with* $\eta = \frac{(\log 2)(1-\log 2)}{(1+2D)\,X^2}$ *satisfies:* $\mathbb{E}[M_T] \leq \sum_{t=1}^{T}\ell_t(\boldsymbol{U}) + \frac{(1+2D)\,X^2\,\|\boldsymbol{U}\|^2}{2(\log 2)(1-\log 2)}$.

Unlike with the hinge loss (Section 3.1), the algorithm for the logistic loss is not parameter-free. We require knowledge of D and X. Knowledge of X can be alleviated by clever tuning of the learning rate such as is done with ODAFTRL [Flaspohler et al., 2021].

The resulting $O(\|\boldsymbol{U}\|^2X^2D)$ surrogate regret bound is independent of both T and K. A proof and detailed comparison with van der Hoeven [2020], Sakaue et al. [2024], and the ODAFTRL/BOLD bounds of Shibukawa et al. [2025] is given in Appendix C.1. A dynamic-regret variant in the non-delayed setting, is given in Appendix C.3.

Algorithm 4 Stability Enhanced GAPTRON with Delayed Feedback For Logistic Loss

Require: Learning rate $\eta > 0$, exploration rate $\gamma \in [0, 1]$.

Initialize $\boldsymbol{W}_1 = \boldsymbol{0}$

for $t = 1$ **to** T **do**

Obtain $\boldsymbol{x}_t$ and set the instability threshold

$$\beta_t = \begin{cases} \frac{2\eta}{\log 2}\|\boldsymbol{x}_t\|^2 + \frac{2\eta}{(\log 2)^2}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| & \text{(full information)}, \\ \frac{2\eta K}{\gamma \log 2}\|\boldsymbol{x}_t\|^2 + \frac{2\eta}{(\log 2)^2}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| & \text{(bandit)}, \\ \frac{4\eta}{\log 2}\|\boldsymbol{x}_t\|^2 & \text{(dynamic regret)}. \end{cases}$$

Compute the margins $m_t(\boldsymbol{W}_t, y) = \langle \boldsymbol{W}_t^y, \boldsymbol{x}_t\rangle - \max_{k\neq y}\langle \boldsymbol{W}_t^k, \boldsymbol{x}_t\rangle$ for all $y \in [K]$

Compute the softmax $\widetilde{\boldsymbol{p}}_t = \big(\sigma(\boldsymbol{W}_t, \boldsymbol{x}_t, k)\big)_{k\in[K]}$ and let $y_t^\star = \arg\max_k m_t(\boldsymbol{W}_t, k)$, $m_t^\star = \max_k m_t(\boldsymbol{W}_t, k)$, and $a_t = \mathbb{1}[m_t^\star \le \beta_t]$

Set $\boldsymbol{p}'_t = (1-\gamma)\big[(1-a_t)\,\boldsymbol{e}_{y_t^\star} + a_t\,\widetilde{\boldsymbol{p}}_t\big] + \frac{\gamma}{K}\,\boldsymbol{1}$

Predict $y'_t \sim \boldsymbol{p}'_t$

Observe the set $\mathcal{F}_t$ of rounds whose feedback arrives during round t

For each $s \in \mathcal{F}_t$, compute the gradient $\boldsymbol{g}_s$:

- Full information / dynamic regret: $\boldsymbol{g}_s = \nabla \ell_s(\boldsymbol{W}_s)$
- Bandit: $\boldsymbol{g}_s = \frac{\mathbb{1}[y'_s = y_s]}{p'_s(y_s)}\,\nabla \ell_s(\boldsymbol{W}_s)$

Update $\boldsymbol{W}_{t+1} = \boldsymbol{W}_t - \eta \sum_{s\in\mathcal{F}_t} \boldsymbol{g}_s$

end for

4.2 Delayed Bandit Feedback for the Logistic Loss

In the bandit setting, the importance-weighted gradient requires $\gamma > 0$ to ensure $p'_t(y) \ge \gamma/K$ (cf. Section 3.2). The cross-terms between rounds are not inflated by importance weighting (10), so the bandit and delay contributions enter β_t additively: $\beta_t = \frac{2\eta K}{\gamma \log 2}\|\boldsymbol{x}_t\|^2 + \frac{2\eta}{(\log 2)^2}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|$.

Theorem 8 (Delayed Bandit Mistake Bound, Logistic Loss)**.** *Let ℓ_t be the logistic loss (3), $\|\boldsymbol{x}_t\| \le X$, $\gamma \in (0,1)$, and $D = \max_t |n_t|$. Suppose $\boldsymbol{U} \in \mathbb{R}^{K\times d}$ with $\|\boldsymbol{U}\| \le R$ for some $R > 0$. Then Algorithm 4 with $\eta = \frac{(\log 2)(1-\log 2)}{(K/\gamma + 2D)\,X^2}$ and $\gamma = \min\Big\{1,\ R\,X\sqrt{\frac{K}{2(\log 2)(1-\log 2)\,T}}\Big\}$ satisfies:* $\mathbb{E}[M_T] \le \sum_{t=1}^T \ell_t(\boldsymbol{U}) + R\,X\sqrt{\frac{2K\,T}{(\log 2)(1-\log 2)}} + \frac{D\,X^2\,R^2}{(\log 2)(1-\log 2)} + O(KX^2R^2)$.

The bound decomposes as $\sum_t \ell_t(\boldsymbol{U}) + O(\|\boldsymbol{U}\|X\sqrt{KT}) + O(DX^2\|\boldsymbol{U}\|^2)$: the bandit cost $O(\sqrt{KT})$ and the delay cost $O(D)$ enter additively and independently, so for any fixed D the delay contribution remains bounded regardless of T. A proof and detailed comparison with the $O(\sqrt{(K+D)T})$ and $O(D^{1/3}T^{2/3})$ bounds of Shibukawa et al. [2025] is given in Appendix C.2.

5 Empirical Risk Minimization and Stability

The stability principle underlying our online mistake bounds has a direct counterpart for regularized empirical risk minimization (ERM). Let $S = ((\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_T, y_T))$ be drawn i.i.d. from $\mathcal{D}$ over $\mathbb{R}^d \times \{-1, +1\}$ with $\|\boldsymbol{x}\| \le X$, and let $\widehat{\boldsymbol{W}} = \arg\min_{\boldsymbol{W}} \sum_{t=1}^T \ell_t(\boldsymbol{W}) + \frac{1}{2\eta}\|\boldsymbol{W}\|^2$ denote the regularized ERM with respect to a example-specific loss $\ell_t(\boldsymbol{W})$. The randomized predictor $y'(\boldsymbol{x}; \boldsymbol{W})$ outputs $\mathrm{sign}(\langle \boldsymbol{W}, \boldsymbol{x}\rangle)$ when $|\langle \boldsymbol{W}, \boldsymbol{x}\rangle| > \beta(\boldsymbol{x})$ and a uniform random label otherwise, with a loss-dependent threshold $\beta(\boldsymbol{x})$. For both the hinge loss $\ell_t(\boldsymbol{W}) = (1 - y_t\langle \boldsymbol{W}, \boldsymbol{x}_t\rangle)_+$ and the (binary) logistic loss (3), leave-one-out stability of $\widehat{\boldsymbol{W}}$ yields an $O(1/T)$ classification error bound against any comparator. Proofs and discussion are deferred to Appendix D.

Theorem 9 (Hinge ERM)**.** *Let $\widehat{\boldsymbol{W}}$ be the regularized ERM with $\ell(\boldsymbol{W}, \boldsymbol{x}, y) = (1 - y\langle \boldsymbol{W}, \boldsymbol{x}\rangle)_+$, $y \in \{-1, +1\}$, and let y' use threshold $\beta(\boldsymbol{x}) = \eta\|\boldsymbol{x}\|^2$. With $\eta = 1/(4X^2)$, for every comparator $\boldsymbol{W}_\star \in \mathbb{R}^d$:* $\mathbb{E}_{S,y'}\Big[\mathbb{1}[y \neq y'(\boldsymbol{x}; \widehat{\boldsymbol{W}})]\Big] \le \mathbb{E}_{(\boldsymbol{x},y)}[\ell(\boldsymbol{W}_\star, \boldsymbol{x}, y)] + \frac{2X^2\|\boldsymbol{W}_\star\|^2}{T+1}$.

Theorem 10 (Logistic ERM). *Let* $\widehat{\boldsymbol{W}}$ *be the regularized ERM with* $\ell(\boldsymbol{W}, \boldsymbol{x}, y) = \log_2(1 + e^{-y\langle \boldsymbol{W}, \boldsymbol{x}\rangle})$, $\boldsymbol{W} \in \mathbb{R}^d$, $y \in \{-1, +1\}$. *Let* y' *use threshold* $\beta(\boldsymbol{x}) = \frac{\eta}{\log 2}\|\boldsymbol{x}\|^2$. *With* $\eta = \frac{(\log 2)\log(\sqrt{2}+1)}{2X^2}$, *for every comparator* $\boldsymbol{W}_\star \in \mathbb{R}^d$: $\mathbb{E}_{S,y'}\left[\mathbb{1}[y \neq y'(\boldsymbol{x}; \widehat{\boldsymbol{W}})]\right] \leq \mathbb{E}_{(\boldsymbol{x},y)}[\ell(\boldsymbol{W}_\star, \boldsymbol{x}, y)] + \frac{X^2\|\boldsymbol{W}_\star\|^2}{(\log 2)\log(\sqrt{2}+1)\,(T+1)}$.

Remark E.1 (A unified view). The proofs of Theorems 9 and 10 share a common backbone: the loss admits a representation $\ell(\langle \boldsymbol{W}, \boldsymbol{x}\rangle, y) = \psi(-y\langle \boldsymbol{W}, \boldsymbol{x}\rangle)$ for some non-decreasing function $\psi : \mathbb{R} \to \mathbb{R}_+$ — namely $\psi(z) = (1+z)_+$ for the hinge loss and $\psi(z) = \log_2(1+e^z)$ for the logistic loss. The pointwise lemmas (Lemma D.1 and D.3) convert the 0–1 indicator into ψ evaluated at $-y\langle \boldsymbol{W}, \boldsymbol{x}\rangle - G\eta\|\boldsymbol{x}\|^2$, where G is the Lipschitz constant of ℓ in its first argument; Cauchy–Schwarz and parameter stability supply a bound $-y\langle \widehat{\boldsymbol{W}}_i, \boldsymbol{x}\rangle \leq -y\langle \widehat{\boldsymbol{W}}_\star, \boldsymbol{x}\rangle + G\eta\|\boldsymbol{x}\|^2$; and these two $G\eta\|\boldsymbol{x}\|^2$ contributions cancel inside the argument of ψ. This is the only structural property of the loss the proof needs: any non-decreasing surrogate of the indicator with a matching pointwise-to-shifted-loss lemma and Lipschitz-constant-aware stability bound would yield the same $O(1/T)$ rate.

6 Simulations

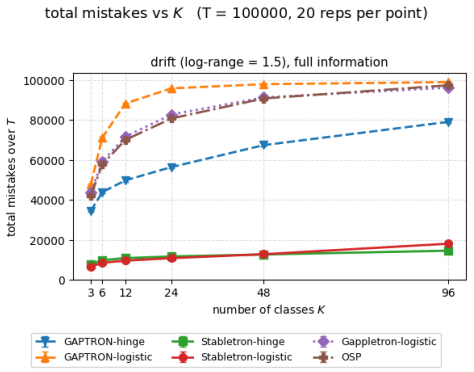


Figure 1: Number of mistakes for each algorithm under non-delayed feedback on SynSep with label noise $r = 0.1$, symmetric scale drift with $\rho = 1.5$ and full information feedback setting (20 repetitions). 1SD error bars are included.

We compare our algorithms (`Stabletron`) against `GAPTRON`, `Gappletron` [van der Hoeven et al., 2021], and `OSP` [Shibukawa et al., 2025] on SynSep [Kakade et al., 2008, van der Hoeven et al., 2021, Shibukawa et al., 2025]. To stress-test scale invariance, we incorporate per round feature scale drift into the SynSep data: each $\boldsymbol{x}_t$ is multiplied by an independent factor 10^{u_t} with $u_t \sim \text{Uniform}(-\rho, \rho)$. Without feature scale drift (Figure 4), `Stabletron` is competitive in both full-information and bandit settings. Under symmetric scale drift, `Stabletron`-Hinge and `Stabletron`-Logistic incur the fewest mistakes: their instability threshold $\beta_t \propto \|\boldsymbol{x}_t\|^2$ is invariant under rescaling of $\boldsymbol{x}_t$, whereas the comparators' absolute thresholds fail when feature norms shrink. The delayed-feedback experiments show the same pattern more sharply, with `Stabletron`-Logistic most robust to combined drift and delay. Details and further experiments are in Appendix E.

Acknowledgments and Disclosure of Funding

DvdH was supported by Netherlands Organization for Scientific Research (NWO), grant number VI.Veni.242.143 (https://doi.org/10.61686/SCGQQ55380).

References

Alina Beygelzimer, Francesco Orabona, and Chicheng Zhang. Efficient online bandit multiclass learning with $\widetilde{O}(\sqrt{T})$ regret. In *International Conference on Machine Learning*, volume 70, pages 488–497. PMLR, 2017.

Olivier Bousquet and André Elisseeff. Stability and generalization. *Journal of Machine Learning Research*, 2 (Mar):499–526, 2002.

Olivier Bousquet, Yegor Klochkov, and Nikita Zhivotovskiy. Sharper bounds for uniformly stable algorithms. In *Conference on Learning Theory*, pages 610–626. PMLR, 2020.

Nicolo Cesa-Bianchi, Alex Conconi, and Claudio Gentile. A second-order perceptron algorithm. *SIAM Journal on Computing*, 34(3):640–668, 2005.

Koby Crammer and Yoram Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, 2:265–292, 2002.

Koby Crammer and Yoram Singer. Ultraconservative online algorithms for multiclass problems. *Journal of Machine Learning Research*, 3(Jan):951–991, 2003.

Koby Crammer, Alex Kulesza, and Mark Dredze. Adaptive regularization of weight vectors. In *Advances in Neural Information Processing Systems*, volume 22, 2009.

Michael Fink, Shai Shalev-Shwartz, Yoram Singer, and Shimon Ullman. Online multiclass learning by interclass hypothesis sharing. In *International Conference on Machine Learning*, pages 313–320, 2006.

Genevieve E Flaspohler, Francesco Orabona, Judah Cohen, Soukayna Mouatadid, Miruna Oprescu, Paulo Orenstein, and Lester Mackey. Online learning with optimism and delay. In *International Conference on Machine Learning*, volume 139, pages 3363–3373. PMLR, 2021.

Dylan J Foster, Satyen Kale, Haipeng Luo, Mehryar Mohri, and Karthik Sridharan. Logistic regression: The importance of being improper. In *Conference on Learning Theory*, pages 167–208. PMLR, 2018.

Elad Hazan and Satyen Kale. Newtron: An efficient bandit algorithm for online multiclass prediction. In *Advances in Neural Information Processing Systems*, volume 24, 2011.

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

Pooria Joulani, Andras Gyorgy, and Csaba Szepesvari. Online learning under delayed feedback. In *International Conference on Machine Learning*, volume 28, pages 1453–1461. PMLR, 2013.

Sham M Kakade, Shai Shalev-Shwartz, and Ambuj Tewari. Efficient bandit algorithms for online multiclass prediction. In *International Conference on Machine Learning*, pages 440–447, 2008.

Yegor Klochkov and Nikita Zhivotovskiy. Stability and deviation optimal risk bounds with convergence rate $o(1/n)$. In *Advances in Neural Information Processing Systems*, volume 34, pages 5065–5076, 2021.

Tomer Koren and Kfir Levy. Fast rates for exp-concave empirical risk minimization. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

Albert BJ Novikoff. On convergence proofs on perceptrons. In *Proceedings of the Symposium on the Mathematical Theory of Automata*, volume 12, pages 615–622, 1962.

Francesco Orabona. A modern introduction to online learning, 2025.

Francesco Orabona, Nicolo Cesa-Bianchi, and Claudio Gentile. Beyond logarithmic bounds in online learning. In *International Conference on Artificial Intelligence and Statistics*, pages 823–831. PMLR, 2012.

Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386, 1958.

Shinsaku Sakaue, Han Bao, Taira Tsuchiya, and Taihei Oki. Online structured prediction with Fenchel–Young losses and improved surrogate regret for online multiclass classification with logistic loss. In *Conference on Learning Theory*, pages 4458–4486. PMLR, 2024.

Shinsaku Sakaue, Han Bao, and Yuzhou Cao. Non-stationary online structured prediction with surrogate losses, 2025.

Shai Shalev-Shwartz, Ohad Shamir, Nathan Srebro, and Karthik Sridharan. Learnability, stability and uniform convergence. *Journal of Machine Learning Research*, 11:2635–2670, 2010.

Yuki Shibukawa, Taira Tsuchiya, Shinsaku Sakaue, and Kenji Yamanishi. Bandit and delayed feedback in online structured prediction, 2025.

Dirk van der Hoeven. Exploiting the surrogate gap in online multiclass classification. In *Advances in Neural Information Processing Systems*. Curran Associates Inc., 2020.

Dirk van der Hoeven, Federico Fusco, and Nicolò Cesa-Bianchi. Beyond bandit feedback in online multiclass classification. In *Advances in Neural Information Processing Systems*, volume 34, pages 13280–13291, 2021.

Lukas Zierahn, Dirk Van Der Hoeven, Tal Lancewicki, Aviv Rosenberg, and Nicolo Cesa-Bianchi. A unified analysis of nonstochastic delayed feedback for combinatorial semi-bandits, linear bandits, and MDPs. *Journal of Machine Learning Research*, 26(104):1–60, 2025.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning*, pages 928–936, 2003.

A Details of Stability in Regret Analysis and Algorithm Design (Section 2)

A.1 Details of Section 2.1

A.1.1 Worked example for Theorem 1

Consider a one-dimensional example with $\boldsymbol{x}_t = \frac{1}{2}$ for all t, and

$$y_t = \begin{cases} +1, & \text{if } t \not\equiv 0 \pmod 5, \\ -1, & \text{if } t \equiv 0 \pmod 5. \end{cases}$$

This sequence is not linearly separable: every fifth label is flipped, so any fixed comparator $\boldsymbol{U}$ incurs nonzero hinge loss. The first ten rounds are shown in Table 2.

t	$\boldsymbol{x}_t$	y_t	$\boldsymbol{W}_t$	y'_t	τ_t	$\mathcal{E}_t$
1	$\frac{1}{2}$	+1	0	-1	1	1
2	$\frac{1}{2}$	+1	$\frac{1}{2}$	$+1$	1	1
3	$\frac{1}{2}$	+1	1	$+1$	1	0
4	$\frac{1}{2}$	+1	$\frac{3}{2}$	$+1$	1	0
5	$\frac{1}{2}$	-1	2	$+1$	1	0
6	$\frac{1}{2}$	+1	$\frac{3}{2}$	$+1$	1	0
7	$\frac{1}{2}$	+1	2	$+1$	0	0
8	$\frac{1}{2}$	+1	2	$+1$	0	0
9	$\frac{1}{2}$	+1	2	$+1$	0	0
10	$\frac{1}{2}$	-1	2	$+1$	1	0

Table 2: First ten rounds of the Perceptron on the example sequence. The instability indicator satisfies $\mathcal{E}_t = 0$ for all $t \geq 3$: after the first two rounds, $|\boldsymbol{W}_t| \geq 1 > \|\boldsymbol{x}_t\|^2 = \frac{1}{4}$, so no update can flip the prediction.

The algorithm continues to update ($\tau_t = 1$) on many rounds, but after $t = 2$ the margin always satisfies $|\boldsymbol{W}_t \cdot \boldsymbol{x}_t| > \|\boldsymbol{x}_t\|^2 = \frac{1}{4}$. This implies $\mathcal{E}_t = 0$ for $t > 2$, since $\operatorname{sign}(\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle) \neq \operatorname{sign}(\langle \boldsymbol{W}_{t+1}, \boldsymbol{x}_t \rangle)$ requires $|\boldsymbol{W}_t \cdot \boldsymbol{x}_t| \leq \|\boldsymbol{x}_t\|^2$. The classical bound gives $M_T - \sum_{t=1}^T \ell_t^1(\boldsymbol{U}) = O(\|\boldsymbol{U}\|\sqrt{T})$, since the number of updates $\sum_{t=1}^T \tau_t$ grows linearly with T, whereas our improved bound counts only the unstable update rounds and yields $M_T - \sum_{t=1}^T \ell_t^1(\boldsymbol{U}) = O(\|\boldsymbol{U}\|^2)$.

A.1.2 Proof of Theorem 1

Theorem 1 (RESTATED)**.** *For any $\boldsymbol{U} \in \mathbb{R}^d$ and $q \geq 1$, the Perceptron (Algorithm 1) satisfies:* $M_T - \sum_{t=1}^T \ell_t^q(\boldsymbol{U}) \leq q\|\boldsymbol{U}\| \sqrt{\sum_{t=1}^T \tau_t \mathcal{E}_t \|\boldsymbol{x}_t\|^2} + q\|\boldsymbol{U}\|^2(1 + \max_t \|\boldsymbol{x}_t\|^2)$.

Proof of Theorem 1. By definition of τ_t, for $\xi \geq 1$ we have that for all $\eta \in (0, \frac{1}{\xi}]$, $\operatorname{sign}(\eta\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle) = \operatorname{sign}(\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle)$ and thus

$$\begin{aligned} \mathbb{1}[y'_t \neq y_t] &\leq \widetilde{\ell}_t(\eta \boldsymbol{W}_t) - \mathbb{1}[y'_t = y_t]\tau_t(1 - y_t \eta \langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle) \\ &\leq \widetilde{\ell}_t(\eta \boldsymbol{W}_t) - \mathbb{1}[y'_t = y_t]\tau_t\Big(1 - \frac{1}{\xi}\Big) \end{aligned}$$

For $\xi \geq 1$, if $y_t\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle < 1$ then for $\eta \in (0, \min\{\frac{1}{2\xi}, \frac{1}{2\xi \max_t \|\boldsymbol{x}_t\|^2}\}]$ we have that $\eta y_t \langle \boldsymbol{W}_{t+1}, \boldsymbol{x}_t \rangle = \eta\big(y_t\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle + \tau_t \|\boldsymbol{x}_t\|^2\big) \leq \frac{1}{\xi}$. Therefore, we have that

$$(1 - \mathcal{E}_t)\mathbb{1}[y'_t \neq y_t] \leq (1 - \mathcal{E}_t)\widetilde{\ell}_t(\eta \boldsymbol{W}_{t+1}) - \mathbb{1}[y'_t = y_t]\tau_t\Big(1 - \frac{1}{\xi}\Big).$$

So far we have shown that

$$\begin{aligned} \sum_{t=1}^T \mathbb{1}[y'_t \neq y_t] &\leq \sum_{t=1}^T \Big((1 - \mathcal{E}_t)\widetilde{\ell}_t(\eta \boldsymbol{W}_{t+1}) + \mathcal{E}_t \widetilde{\ell}_t(\eta \boldsymbol{W}_t) - \mathbb{1}[y'_t = y_t]\tau_t\Big(1 - \frac{1}{\xi}\Big)\Big) \\ &= \sum_{t=1}^T \widetilde{\ell}_t(\eta \boldsymbol{W}_t) - \sum_{t=1}^T \Big((1 - \mathcal{E}_t)\eta\tau_t\|\boldsymbol{x}_t\|^2 + \mathbb{1}[y'_t = y_t]\tau_t\Big(1 - \frac{1}{\xi}\Big)\Big). \end{aligned}$$

where the last equality follows from $\eta y_t \langle \boldsymbol{W}_{t+1}, \boldsymbol{x}_t \rangle = \eta\big(y_t \langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle + \tau_t \|\boldsymbol{x}_t\|^2\big)$. Now, for any $\boldsymbol{U} \in \mathbb{R}^d$, the standard OGD regret bound (see e.g. [Orabona, 2025, Theorem 2.13]) then gives

$$\sum_{t=1}^{T} \widetilde{\ell}_t(\eta \boldsymbol{W}_t) - \widetilde{\ell}_t(\boldsymbol{U}) \le \frac{\|\boldsymbol{U}\|^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \tau_t \|\boldsymbol{x}_t\|^2 \,. \tag{4}$$

Therefore, we can conclude that for any $\eta \in (0, \min\{\frac{1}{2\xi}, \frac{1}{2\xi \max_t \|\boldsymbol{x}_t\|^2}\}]$

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{1}[y'_t \neq y_t] &\le \sum_{t=1}^{T} \widetilde{\ell}_t(\eta \boldsymbol{W}_t) - \sum_{t=1}^{T} \big((1-\mathcal{E}_t)\eta\tau_t \|\boldsymbol{x}_t\|^2 + \mathbb{1}[y'_t = y_t]\tau_t\Big(1 - \frac{1}{\xi}\Big)\big) \\
&\le \sum_{t=1}^{T} \big(\widetilde{\ell}_t(\boldsymbol{U}) - \mathbb{1}[y'_t = y_t]\tau_t\Big(1 - \frac{1}{\xi}\Big)\big) + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \tau_t \mathcal{E}_t \|\boldsymbol{x}_t\|^2 \\
&\le \sum_{t=1}^{T} \big(\tau_t \ell_t^1(\boldsymbol{U}) - \mathbb{1}[y'_t = y_t]\tau_t\Big(1 - \frac{1}{\xi}\Big)\big) + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \tau_t \mathcal{E}_t \|\boldsymbol{x}_t\|^2 \,. \qquad (5)
\end{aligned}$$

We now show the result for powers of $q > 1$. Denote by $\phi_T = \sum_{t: y'_t = y_t} \tau_t$ and note that $\sum_{t=1}^{T} \tau_t = M_T + \phi_T$. We apply Hölder's inequality with conjugate exponents p, q to $\sum_{t=1}^{T} \tau_t \ell_t(\boldsymbol{U})$, using $\tau_t^p = \tau_t$, and then Young's inequality ($ab \le a^p/p + b^q/q$):

$$\sum_{t=1}^{T} \tau_t \, \ell_t(\boldsymbol{U}) \;\le\; \Big(\sum_{t=1}^{T} \tau_t\Big)^{1/p} \Big(\sum_{t=1}^{T} \ell_t^q(\boldsymbol{U})\Big)^{1/q} \;\le\; \frac{M_T + \phi_T}{p} + \frac{1}{q} \sum_{t=1}^{T} \ell_t^q(\boldsymbol{U}) \,.$$

Substituting this into (5) gives us

$$M_T \le \frac{M_T + \phi_T}{p} - \Big(1 - \frac{1}{\xi}\Big)\phi_T + \frac{1}{q} \sum_{t=1}^{T} \ell_t^q(\boldsymbol{U}) + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \tau_t \mathcal{E}_t \|\boldsymbol{x}_t\|^2 \,,$$

Using $1 - 1/p = 1/q$ this implies that

$$\frac{M_T}{q} \le= \big(1 - 1/p\big) M_T \le \Big(\frac{1}{\xi} - \frac{1}{q}\Big)\phi_T + \frac{1}{q} \sum_{t=1}^{T} \ell_t^q(\boldsymbol{U}) + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \tau_t \mathcal{E}_t \|\boldsymbol{x}_t\|^2 \,.$$

Equivalently,

$$M_T \le q\Big(\frac{1}{\xi} - \frac{1}{q}\Big)\phi_T + \sum_{t=1}^{T} \ell_t^q(\boldsymbol{U}) + \frac{q\|\boldsymbol{U}\|^2}{2\eta} + \frac{q\eta}{2} \sum_{t=1}^{T} \tau_t \mathcal{E}_t \|\boldsymbol{x}_t\|^2 \,.$$

Setting $\eta = \min\{\sqrt{\frac{\|\boldsymbol{U}\|^2}{\sum_{t=1}^{T} \tau_t \mathcal{E}_t \|\boldsymbol{x}_t\|^2}}, \min\{\frac{1}{2q}, \frac{1}{2q \max_t \|\boldsymbol{x}_t\|^2}\}\}$ then completes the proof. □

A.2 Details of Section 2.2

Corollary 2 is a direct consequence of the following theorem.

Theorem A.1. *Algorithm 2 guarantees that for any $\boldsymbol{U} \in \mathbb{R}^d$*

$$\mathbb{E}[M_T] \le \sum_{t=1}^{T} \ell_t^q(\boldsymbol{U}) + \inf_{\eta > 0\,,\, q \ge 1} \left\{ q\frac{\|\boldsymbol{U}\|^2}{2\eta} + \sum_{t=1}^{T} \mathbb{1}[|\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle| \le \|\boldsymbol{x}_t\|^2]\big(-\tfrac{1}{2} + \tfrac{3q}{2}\eta\|\boldsymbol{x}_t\|^2\big) \right\}.$$

Proof of Theorem A.1. We use the notation $\boldsymbol{W}_t(\eta) = \eta \boldsymbol{W}_t$ and $a_t = \mathbb{1}[|\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle| \le \|\boldsymbol{x}_t\|^2]$. Additionally define

$$b_t := \mathbb{1}[y_t \langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle \le \|\boldsymbol{x}_t\|^2], \qquad \widetilde{\ell}_t^q(\boldsymbol{W}) := b_t \ell_t^q(\boldsymbol{W}), \qquad \widetilde{\ell}_t(\boldsymbol{W}) := \widetilde{\ell}_t^1(\boldsymbol{W}).$$

The update vector used by the algorithm is the gradient of the linear part of the hinge loss, i.e. $\boldsymbol{g}_t = -b_t y_t \boldsymbol{x}_t$. It can be seen that $\boldsymbol{W}_{t+1}(\eta) = \boldsymbol{W}_t(\eta) - \eta \boldsymbol{g}_t$ and thus iterates $\boldsymbol{W}_1(\eta), \dots \boldsymbol{W}_T(\eta)$ are iterates of online gradient descent [Zinkevich, 2003] on ℓ_t^q.

Since y_t^* depends only on the sign of $\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle$, which positive scaling preserves, $y_t^* = \operatorname{sign}(\langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle) = \operatorname{sign}(\eta \langle \boldsymbol{W}_t, \boldsymbol{x}_t \rangle)$, i.e. y_t' is independent of η whenever $a_t = 0$. The uniform sampling, and the uniform sampling rule are independent of η, since a_t is independent of η. Therefore the sequence of issued predictions y_t' would be the same if we replaced the parameters $\boldsymbol{W}_t$ in Algorithm 2 by $\boldsymbol{W}_t(\eta)$.

For any $\boldsymbol{U}$,

$$\widetilde{\ell}_t(\boldsymbol{U}) = b_t(1 - y_t \langle \boldsymbol{U}, \boldsymbol{x}_t \rangle)_+ \geq b_t(1 - y_t \langle \boldsymbol{U}, \boldsymbol{x}_t \rangle) = b_t + \langle \boldsymbol{g}_t, \boldsymbol{U} \rangle.$$

Since $\langle \boldsymbol{g}_t, \boldsymbol{W}_t(\eta) \rangle = -b_t y_t \langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle$, we obtain

$$b_t(1 - y_t \langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle) - \widetilde{\ell}_t(\boldsymbol{U}) \leq \langle \boldsymbol{g}_t, \boldsymbol{W}_t(\eta) - \boldsymbol{U} \rangle.$$

We therefore have

$$\begin{aligned}
\mathbb{E}[M_T] &= \sum_{t=1}^{T} \left((1 - a_t) \mathbb{1}[y_t^\star \neq y_t] + \tfrac{1}{2} a_t \right) \\
&= \sum_{t=1}^{T} \left((1 - a_t) \mathbb{1}[y_t^\star \neq y_t] + \tfrac{1}{2} a_t - b_t(1 - y_t \langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle) \right) \\
&\quad + \sum_{t=1}^{T} b_t(1 - y_t \langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle) \\
&\leq \sum_{t=1}^{T} \left((1 - a_t) \mathbb{1}[y_t^\star \neq y_t] + \tfrac{1}{2} a_t - b_t(1 - y_t \langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle) + \frac{\eta}{2} \|\boldsymbol{g}_t\|^2 \right) \\
&\quad + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \sum_{t=1}^{T} \widetilde{\ell}_t^1(\boldsymbol{U}),
\end{aligned}$$

where the inequality follows from for example [Orabona, 2025, Theorem 2.13]. Next we show that

$$\begin{aligned}
S_t &:= (1 - a_t) \mathbb{1}[y_t^\star \neq y_t] + \tfrac{1}{2} a_t - b_t(1 - y_t \langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle) + \frac{\eta}{2} \|\boldsymbol{g}_t\|^2 \\
&\leq a_t \left(-\tfrac{1}{2} + \tfrac{3}{2} \eta \|\boldsymbol{x}_t\|^2 \right).
\end{aligned}$$

We split the analysis into four cases.

$\boxed{y_t^\star = y_t,\ a_t = 0}$

Since $a_t = 0$ and $y_t^\star = y_t$, we have $b_t = 0$ and $\boldsymbol{g}_t = \boldsymbol{0}$. Consequently $S_t = 0$.

$\boxed{y_t^\star \neq y_t,\ a_t = 0}$

Since $y_t^\star \neq y_t$, we have $b_t = 1$ and $y_t \langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle = -|\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle| < 0$. Therefore $b_t(1 - y_t \langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle) = 1 + |\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle|$ and $\|\boldsymbol{g}_t\|^2 = \|\boldsymbol{x}_t\|^2$.

$$\begin{aligned}
S_t &= 1 - (1 + |\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle|) + \tfrac{\eta}{2} \|\boldsymbol{x}_t\|^2 \\
&= -|\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle| + \tfrac{\eta}{2} \|\boldsymbol{x}_t\|^2.
\end{aligned}$$

Since $a_t = 0$ implies $|\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t \rangle| > \eta \|\boldsymbol{x}_t\|^2$:

$$S_t < -\tfrac{\eta}{2} \|\boldsymbol{x}_t\|^2 \leq 0.$$

$\boxed{y_t^\star \neq y_t,\ a_t = 1}$

Here $b_t = 1$, $b_t(1 - y_t\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t\rangle) = 1 + |\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t\rangle|$, and $\|\boldsymbol{g}_t\|^2 = \|\boldsymbol{x}_t\|^2$.

$$\begin{aligned} S_t &= \tfrac{1}{2} - (1 + |\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t\rangle|) + \tfrac{\eta}{2}\|\boldsymbol{x}_t\|^2 \\ &= -\tfrac{1}{2} - |\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t\rangle| + \tfrac{\eta}{2}\|\boldsymbol{x}_t\|^2 \le -\tfrac{1}{2} + \tfrac{\eta}{2}\|\boldsymbol{x}_t\|^2. \end{aligned}$$

$\boxed{y_t^\star = y_t,\ a_t = 1}$

Here $b_t = 1$ and

$$b_t(1 - y_t\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t\rangle) = 1 - |\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t\rangle| \ge 1 - \eta\|\boldsymbol{x}_t\|^2,$$

while $\|\boldsymbol{g}_t\|^2 = \|\boldsymbol{x}_t\|^2$.

$$S_t \le \tfrac{1}{2} - (1 - \eta\|\boldsymbol{x}_t\|^2) + \tfrac{\eta}{2}\|\boldsymbol{x}_t\|^2 = -\tfrac{1}{2} + \tfrac{3}{2}\eta\|\boldsymbol{x}_t\|^2.$$

In all four cases, $S_t \le a_t(-\frac{1}{2} + \frac{3}{2}\eta\|\boldsymbol{x}_t\|^2)$. For $q = 1$, using $\widetilde{\ell}_t(\boldsymbol{U}) \le \ell_t(\boldsymbol{U})$ gives the result. Starting from

$$\begin{aligned} \mathbb{E}[M_T] &= \sum_{t=1}^{T}(1 - a_t)\mathbb{1}[y_t^* \ne y_t] + \frac{1}{2}a_t \\ &\le \sum_{t=1}^{T}\ell_t(\boldsymbol{U})\mathbb{1}[y_t\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t\rangle \le \eta\|\boldsymbol{x}_t\|^2] + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \sum_{t=1}^{T} S_t, \end{aligned} \tag{6}$$

we apply Hölder's inequality with $\frac{1}{p} + \frac{1}{q} = 1$ followed by Young's inequality (for $q > 1$) to (6):

$$\begin{aligned} \mathbb{E}[M_T] &\le \left(\sum_{t=1}^{T}\ell_t^q(\boldsymbol{U})\right)^{\frac{1}{q}} \cdot \left(\sum_{t=1}^{T}\mathbb{1}[y_t\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t\rangle \le \eta\|\boldsymbol{x}_t\|^2]\right)^{\frac{1}{p}} + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \sum_{t=1}^{T} S_t \\ &\le \frac{1}{q}\sum_{t=1}^{T}\ell_t^q(\boldsymbol{U}) + \frac{1}{p}\sum_{t=1}^{T}\mathbb{1}[y_t\langle \boldsymbol{W}_t(\eta), \boldsymbol{x}_t\rangle \le \eta\|\boldsymbol{x}_t\|^2] + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \sum_{t=1}^{T} S_t. \end{aligned}$$

Using the upper bound $\sum_{t=1}^{T} S_t \le \sum_{t=1}^{T} a_t\big(-\frac{1}{2} + \frac{3}{2}\eta X^2\big)$ together with (6) to substitute the indicator sum, we get

$$\mathbb{E}[M_T] \le \tfrac{1}{q}\sum_{t=1}^{T}\ell_t^q(\boldsymbol{U}) + \tfrac{1}{p}\mathbb{E}[M_T] + \frac{1}{2}\sum_{t=1}^{T} a_t\left(\tfrac{1}{p} - 1 + 3\eta X^2\right) + \tfrac{\|\boldsymbol{U}\|^2}{2\eta}.$$

Since $1 - \frac{1}{p} = \frac{1}{q}$, rearranging and multiplying both sides by q yields

$$\mathbb{E}[M_T] \le \sum_{t=1}^{T}\ell_t^q(\boldsymbol{U}) + \sum_{t=1}^{T} a_t\left(\tfrac{3q}{2}\eta X^2 - \tfrac{1}{2}\right) + q\tfrac{\|\boldsymbol{U}\|^2}{2\eta}.$$

□

Corollary 2 (RESTATED). *With $A_T = \sum_{t=1}^{T} a_t\|\boldsymbol{x}_t\|^2$, and $N_T = \sum_{t=1}^{T} a_t$, and a_t as in line 5, Algorithm 2 guarantees, for any $\boldsymbol{U} \in \mathbb{R}^d$ and $q \ge 1$: $\mathbb{E}[M_T] \le \sum_{t=1}^{T}\ell_t^q(\boldsymbol{U}) + q\min\left\{\|\boldsymbol{U}\|\sqrt{3A_T}\,,\ \frac{3}{2}\|\boldsymbol{U}\|^2\frac{A_T}{N_T}\right\}$*

Proof of Corollary 2. Since the analysis allows us to choose η freely in Theorem A.1, we can set $\eta = \frac{\|\boldsymbol{U}\|}{X\sqrt{3A_T}}$ to achieve the first bound, and $\eta = \frac{N_T}{3A_T}$ to achieve the second bound. □

B Details of Multiclass Hinge Loss (Section 3)

We use online gradient descent (OGD) with a constant learning rate as the base learner, following van der Hoeven [2020], modified to handle delayed feedback by aggregating the updates of all rounds whose feedback arrives during the current round: $\boldsymbol{W}_{t+1}(\eta) = \boldsymbol{W}_t(\eta) - \eta \sum_{s \in \mathcal{F}_t} \boldsymbol{g}_s(\eta)$, where $\boldsymbol{g}_s(\eta)$ is the gradient of the linear part of the multiclass hinge loss, defined below.

We define the margin of a label y as $m_t(\boldsymbol{W}_t(\eta), y) = \langle \boldsymbol{W}_t^y(\eta), \boldsymbol{x}_t \rangle - \max_{k \neq y} \langle \boldsymbol{W}_t^k(\eta), \boldsymbol{x}_t \rangle$, the greedy prediction as $y_t^\star = \arg\max_k \langle \boldsymbol{W}_t^k(\eta), \boldsymbol{x}_t \rangle$, and the top margin as $m_t^*(\eta) = \max_k m_t(\boldsymbol{W}_t(\eta), k)$. To ease notation, we write $\boldsymbol{W}_t = \boldsymbol{W}_t(1)$ and $m_t^* = m_t^*(1)$. We use a truncated multiclass hinge loss that zeroes out the loss on stable correct rounds:

$$\ell_t(\boldsymbol{W}) = \begin{cases} \max\{1 - m_t(\boldsymbol{W}, y_t),\, 0\}, & \text{if } m_t^* \leq \beta_t \\ \max\{1 - m_t(\boldsymbol{W}, y_t),\, 0\}, & \text{if } y_t^* \neq y_t \text{ and } m_t^* > \beta_t \\ 0, & \text{if } y_t^\star = y_t \text{ and } m_t^\star > \beta_t\,, \end{cases} \tag{7}$$

where β_t is the instability threshold specified per setting in Algorithm 3. We can now define our update as,

$$g_t(\eta) := \mathbb{1}[a_t = 1 \text{ or } y_t^\star \neq y_t](\boldsymbol{e}_{k_t^\star} - \boldsymbol{e}_{y_t}) \otimes \boldsymbol{x}_t.$$

To ease notation we will write $\boldsymbol{g}_t = \boldsymbol{g}_t(1)$. Since $\mathcal{W}$ is unconstrained, the iterates scale linearly: $\boldsymbol{W}_t(\eta) = \eta \boldsymbol{W}_t(1)$. In the bandit setting, the updates are multiplied by the importance weight $\mathbb{1}[y_t = y_t']/p_t'(y_t)$ to provide an unbiased estimate.

Two components encode the stability-based design: an instability indicator $a_t = \mathbb{1}[m_t^* \leq \beta_t]$ that flags rounds where the pending update could flip the prediction on $\boldsymbol{x}_t$, and a *competitive set*:

$$\mathcal{U}_t = \{y : m_t(\boldsymbol{W}_t, y) > -\beta_t\} \qquad \text{(line 5)}\,, \tag{8}$$

collecting the labels that could become the top prediction once the updates are applied. This contrasts with the original `GAPTRON` [van der Hoeven, 2020, Algorithm 1], where uniform exploration over all K classes incurs a mistake probability of $(K-1)/K$ per unstable round; restricting to $\mathcal{U}_t$ reduces this to $(|\mathcal{U}_t| - 1)/|\mathcal{U}_t|$, which can be substantially smaller when $|\mathcal{U}_t| \ll K$.

Another improvement over [van der Hoeven, 2020, Algorithm 1] is that Algorithm 3 is parameter-free in η: every prediction it produces is independent of the step size. First, $y_t^\star$ depends only on the ordering of $\{\langle \boldsymbol{W}_t^k, \boldsymbol{x}_t \rangle\}_k$, which positive scaling (with η) preserves. Second, a_t and $\mathcal{U}_t$ are defined in terms of $\boldsymbol{W}_t(1)$ and the data alone: β_t contains no η. Third, the prediction distribution at line 6, consisting of $(1-\gamma)\boldsymbol{e}_{y_t^\star}$ on stable rounds – uniform over $\mathcal{U}_t$ on unstable rounds, plus the γ-uniform exploration term used in the bandit setting – is built from $y_t^\star$, a_t, $\mathcal{U}_t$, and γ, none of which depend on η. The user may therefore run Algorithm 3 without specifying η, and η is left as a free parameter (the predictions would be the same if the OGD updates were ran with any $\eta > 0$) in the analysis to be optimized in hindsight.

B.1 Details of Section 3.1

Corollary 4 gives two bounds and uses the better one. The first is linear in $\|\boldsymbol{U}\|$ and tighter when $\|\boldsymbol{U}\|$ is large or there are few unstable rounds; the second is quadratic in $\|\boldsymbol{U}\|$, eliminates the surrogate gap entirely via a data-dependent choice of η in the analysis, and is preferable when $\|\boldsymbol{U}\|$ is small or there are many unstable rounds (and is constant in T). Both depend only on quantities averaged over unstable rounds: B_T aggregates feature norms and pending cross-terms, while A_T measures the average inverse size of the competitive set. Neither bound involves worst-case quantities such as K, $\max_t \|\boldsymbol{x}_t\|$, or D. When $D = 0$ (i.e. $n_t = \emptyset$ for all t), the cross-terms in B_T vanish and both bounds are constant in T.

Theorem B.1 gives the first surrogate regret bound for the multiclass hinge loss that is additive rather than multiplicative in the delay. The worst-case form of Corollary 4, obtained by upper-bounding $|\mathcal{U}_t| \leq K$ and $\|\boldsymbol{x}_t\|^2 \leq X^2 := \max_t \|\boldsymbol{x}_t\|^2$, gives $\mathbb{E}[M_T] \leq \sum_{t=1}^T \ell_t(\boldsymbol{U}) + K(1 + 2D)X^2\|\boldsymbol{U}\|^2$: linear in D, constant in T. This worst case is attained only when every round is unstable ($a_t = 1$), every feature has norm X, and every round has D pending labels. In typical sequences, instability is intermittent and the realized cross-terms are well below $D \cdot X^2$, so the data-dependent bound is substantially sharper.

Beyond the worst case, the bound is robust to outliers along several axes simultaneously. A round with very large $\|\boldsymbol{x}_t\|$, large $|\mathcal{U}_t|$, or large $|n_t|$ contributes to B_T only if it is also unstable; if $a_t = 0$ the round drops out entirely. When $a_t = 1$, the anomaly is averaged over active rounds rather than entering as a uniform maximum. Consequently, a single round with $\|\boldsymbol{x}_t\|$, $|\mathcal{U}_t|$, or $|n_t|$ much larger than the rest therefore does not inflate the bound through an X^2, K, or D factor, with its contribution being divided by the number of unstable rounds.

Both robustness and the additive-in-D form rest on parameter-freeness in η. van der Hoeven [2020, Theorem 4.2] achieves the non-delayed bound $O(K^2X^2\|\boldsymbol{U}\|^2)$ by setting η in advance using a priori knowledge of X. Our algorithm requires no such tuning: η does not enter the prediction distribution or the instability indicator, so the bound of Corollary 4 holds simultaneously for all $\eta > 0$ and is optimized in hindsight against the realized data, without prior knowledge of $\boldsymbol{U}$, X, T, or D. Specializing to $D = 0$ recovers a non-delayed worst-case bound of $KX^2\|\boldsymbol{U}\|^2$.

Theorem B.1. *Set $\gamma = 0$, and let ℓ_t denote the multiclass hinge loss (7). Then for any $\boldsymbol{U} \in \mathbb{R}^{K\times d}$, Algorithm 3 satisfies*

$$\mathbb{E}[M_T] \le \sum_{t=1}^{T} \ell_t(\boldsymbol{U}) + \inf_{\eta>0}\left\{ \frac{\|\boldsymbol{U}\|^2}{2\eta} + \sum_{t=1}^{T} a_t\, \mathbb{1}[y_t \in \mathcal{U}_t]\left(-\frac{1}{|\mathcal{U}_t|} + 2\eta\|\boldsymbol{x}_t\|^2 + 4\eta \sum_{s\in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| \right) \right\}.$$

New Proof of Theorem B.1. Write

$$b_t = \mathbb{1}[a_t = 1 \text{ or } y_t^\star \neq y_t], \qquad \boldsymbol{g}_t = b_t(\boldsymbol{e}_{k_t^\star} - \boldsymbol{e}_{y_t}) \otimes \boldsymbol{x}_t, \qquad k_t^\star \in \arg\max_{k\neq y_t} \langle \boldsymbol{W}_t^k, \boldsymbol{x}_t\rangle.$$

The vectors $\boldsymbol{g}_t$ are used in the update $\boldsymbol{W}_{t+1} = \boldsymbol{W}_t - \sum_{s\in\mathcal{F}_t} \boldsymbol{g}_s$. The surrogate loss remains as in Equation (7), i.e. $\ell_t(\boldsymbol{W}) = b_t(1 - m_t(\boldsymbol{W}, y_t))_+$.

For any $\eta > 0$, consider the auxiliary sequence

$$\boldsymbol{W}_{t+1}(\eta) = \boldsymbol{W}_t(\eta) - \eta \sum_{s\in\mathcal{F}_t} \boldsymbol{g}_s, \qquad \boldsymbol{W}_1(\eta) = \boldsymbol{0}$$

We first establish that Algorithm 3 produces the same predictions if we ran it with any $\eta > 0$, then prove the bound for the iterate sequence $\boldsymbol{W}_t(\eta)$, and we can conclude by taking the infimum over η, since the same surrogate regret bounds hold for any value of η.

Since $\mathcal{W} = \mathbb{R}^{K\times d}$ is unconstrained and $\boldsymbol{W}_1 = \boldsymbol{0}$, the recurrence gives $\boldsymbol{W}_t(\eta) = \eta\boldsymbol{W}_t$, and hence $m_t(\boldsymbol{W}_t(\eta), y) = \eta m_t(\boldsymbol{W}_t, y)$ and $m_t^\star(\eta) = \eta m_t^\star$. The greedy prediction $y_t^\star$ depends only on the ordering of $\{\langle \boldsymbol{W}_t^k(\eta), \boldsymbol{x}_t\rangle\}_k$, which positive scaling preserves. The instability indicator $a_t = \mathbb{1}[m_t^\star \le \beta_t]$ and competitive set $\mathcal{U}_t = \{y : m_t(\boldsymbol{W}_t, y) > -\beta_t\}$ are defined directly in terms of the actual iterates $\boldsymbol{W}_t = \boldsymbol{W}_t(1)$ and β_t, both quantities independent of η. Positive scaling obviously preserves the maximizer $k_t^\star$. Thus the same update vector $\boldsymbol{g}_t$ is used for every $\eta > 0$: when $b_t = 1$, it equals $(\boldsymbol{e}_{k_t^\star} - \boldsymbol{e}_{y_t}) \otimes \boldsymbol{x}_t$; when $b_t = 0$, it is zero. In particular, this update condition does not involve the unit-margin cutoff of the hinge loss.

The prediction distribution $\boldsymbol{p}_t'$, formed from $y_t^\star$, a_t, and $\mathcal{U}_t$, is therefore unchanged. This justifies treating η as a free parameter in the analysis: although η does not appear in Algorithm 3, we may apply the linearized OGD bound to the auxiliary sequence $\boldsymbol{W}_t(\eta)$ for any step size $\eta > 0$.

For any comparator $\boldsymbol{U}$,

$$\begin{aligned}\ell_t(\boldsymbol{U}) &= b_t\left(1 - \langle \boldsymbol{U}^{y_t}, \boldsymbol{x}_t\rangle + \max_{k\neq y_t}\langle \boldsymbol{U}^k, \boldsymbol{x}_t\rangle\right)_+ \\ &\ge b_t\left(1 - \langle \boldsymbol{U}^{y_t}, \boldsymbol{x}_t\rangle + \langle \boldsymbol{U}^{k_t^\star}, \boldsymbol{x}_t\rangle\right) = b_t + \langle \boldsymbol{g}_t, \boldsymbol{U}\rangle,\end{aligned}$$

by noting that if $b_t = 0$, the left hand side is 0 as well. Moreover, $\langle \boldsymbol{g}_t, \boldsymbol{W}_t(\eta)\rangle = -\eta b_t m_t(\boldsymbol{W}_t, y_t)$. Consequently,

$$b_t(1 - \eta m_t(\boldsymbol{W}_t, y_t)) - \ell_t(\boldsymbol{U}) \le \langle \boldsymbol{g}_t, \boldsymbol{W}_t(\eta) - \boldsymbol{U}\rangle.$$

Let $\boldsymbol{W}_t^\star(\eta) = -\eta\sum_{s=1}^{t-1} \boldsymbol{g}_s$ denote the reference iterate using all of the same update vectors up to round $t-1$. Here $n_t = [t-1] \setminus \bigcup_{r<t} \mathcal{F}_r$ denotes the rounds whose feedback is still pending before

the prediction at round t. Since $\boldsymbol{W}_t(\eta) - \boldsymbol{W}_t^\star(\eta) = \eta \sum_{s \in n_t} \boldsymbol{g}_s$, the linearized regret decomposes as

$$\sum_{t=1}^{T} \langle \boldsymbol{g}_t, \boldsymbol{W}_t(\eta) - \boldsymbol{U} \rangle = \sum_{t=1}^{T} \langle \boldsymbol{g}_t, \boldsymbol{W}_t^\star(\eta) - \boldsymbol{U} \rangle + \eta \sum_{t=1}^{T} \sum_{s \in n_t} \langle \boldsymbol{g}_s, \boldsymbol{g}_t \rangle.$$

The first term on the right-hand side is bounded by $\frac{\|\boldsymbol{U}\|^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \|\boldsymbol{g}_t\|^2$ via [Orabona, 2025, Theorem 2.13].

Expanding the conditional mistake probability under $\boldsymbol{p}'_t$ (with $\gamma = 0$) gives

$$\begin{aligned} \mathbb{E}_t[\mathbb{1}[y_t \neq y'_t]] &= (1 - a_t)\mathbb{1}[y_t \neq y_t^\star] \\ &\quad + a_t \left(\mathbb{1}[y_t \in \mathcal{U}_t] \frac{|\mathcal{U}_t| - 1}{|\mathcal{U}_t|} + \mathbb{1}[y_t \notin \mathcal{U}_t] \right) \end{aligned}$$

Combining the preceding inequalities and taking expectations,

$$\mathbb{E}[M_T] \leq \sum_{t=1}^{T} \ell_t(\boldsymbol{U}) + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \sum_{t=1}^{T} S_t,$$

where

$$\begin{aligned} S_t = (1 - a_t)\mathbb{1}[y_t \neq y_t^\star] + a_t \left(\mathbb{1}[y_t \in \mathcal{U}_t] \frac{|\mathcal{U}_t| - 1}{|\mathcal{U}_t|} + \mathbb{1}[y_t \notin \mathcal{U}_t] \right) \\ - b_t(1 - \eta m_t(\boldsymbol{W}_t, y_t)) + \frac{\eta}{2} \|\boldsymbol{g}_t\|^2 + \eta \sum_{s \in n_t} \langle \boldsymbol{g}_s, \boldsymbol{g}_t \rangle. \end{aligned} \tag{9}$$

We bound the cross-terms by $\eta \sum_{s \in n_t} \langle \boldsymbol{g}_s, \boldsymbol{g}_t \rangle \leq 2\eta \sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle|$ and use $\|\boldsymbol{g}_t\|^2 \leq 2\|\boldsymbol{x}_t\|^2$ throughout. We show

$$S_t \leq a_t \mathbb{1}[y_t \in \mathcal{U}_t] \left(-\frac{1}{|\mathcal{U}_t|} + 2\eta \|\boldsymbol{x}_t\|^2 + 4\eta \sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle| \right)$$

by case analysis on a_t, $y_t^\star$, and whether $y_t \in \mathcal{U}_t$.

$\boxed{a_t = 0,\ y_t^\star = y_t}$

Here $b_t = 0$ and $\boldsymbol{g}_t = \boldsymbol{0}$, so $S_t = 0$.

$\boxed{a_t = 0,\ y_t \neq y_t^\star}$

Here $b_t = 1$. Since $y_t \neq y_t^\star$, $m_t(\boldsymbol{W}_t, y_t) \leq -m_t^\star$. Hence

$$\begin{aligned} S_t &\leq \eta m_t(\boldsymbol{W}_t, y_t) + \eta \|\boldsymbol{x}_t\|^2 + 2\eta \sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle| \\ &\leq \eta(-m_t^\star + \beta_t) < 0, \end{aligned}$$

where the last inequality uses $a_t = 0$, i.e., $m_t^\star > \beta_t$.

$\boxed{a_t = 1,\ y_t \neq y_t^\star,\ y_t \notin \mathcal{U}_t}$

Here $b_t = 1$. Since $y_t \notin \mathcal{U}_t$, $m_t(\boldsymbol{W}_t, y_t) \leq -\beta_t$. Hence

$$S_t \leq \eta m_t(\boldsymbol{W}_t, y_t) + \eta \beta_t \leq 0.$$

$\boxed{a_t = 1,\ y_t \neq y_t^\star,\ y_t \in \mathcal{U}_t}$

Here $b_t = 1$ and $m_t(\boldsymbol{W}_t, y_t) \le 0$, so $b_t(1 - \eta m_t(\boldsymbol{W}_t, y_t)) \ge 1$. Hence

$$S_t \le \frac{|\mathcal{U}_t| - 1}{|\mathcal{U}_t|} - 1 + \eta\|\boldsymbol{x}_t\|^2 + 2\eta \sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle|$$
$$= -\frac{1}{|\mathcal{U}_t|} + \eta\|\boldsymbol{x}_t\|^2 + 2\eta \sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle|.$$

$\boxed{a_t = 1,\ y_t^\star = y_t,\ y_t \in \mathcal{U}_t}$

Here $b_t = 1$, $m_t(\boldsymbol{W}_t, y_t) = m_t^\star$, and $m_t^\star \le \beta_t$. Hence

$$S_t \le \frac{|\mathcal{U}_t| - 1}{|\mathcal{U}_t|} - (1 - \eta m_t^\star) + \eta\|\boldsymbol{x}_t\|^2 + 2\eta \sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle|$$
$$= -\frac{1}{|\mathcal{U}_t|} + \eta m_t^\star + \eta\|\boldsymbol{x}_t\|^2 + 2\eta \sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle|$$
$$\le -\frac{1}{|\mathcal{U}_t|} + 2\eta\|\boldsymbol{x}_t\|^2 + 4\eta \sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle|.$$

This is the binding case: one copy of $\eta\beta_t$ comes from $m_t^\star \le \beta_t$ and one from the update-norm and cross-term bounds, giving the factor of 2 on $\eta\|\boldsymbol{x}_t\|^2$ and 4 on the cross-terms.

The case $y_t^\star = y_t$, $y_t \notin \mathcal{U}_t$ is impossible since $y_t^\star \in \mathcal{U}_t$. Combining the bounds (case 4 is dominated by case 5), we obtain the uniform upper bound stated above. Substituting into $\mathbb{E}[M_T] \le \sum_{t=1}^T \ell_t(\boldsymbol{U}) + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \sum_{t=1}^T S_t$ and taking the infimum over $\eta > 0$ yields the theorem. □

Corollary 4 (RESTATED). *Under Theorem B.1's conditions, Algorithm 3 satisfies for any $\boldsymbol{U} \in \mathbb{R}^{K \times d}$: $\mathbb{E}[M_T] \le \sum_{t=1}^T \ell_t(\boldsymbol{U}) + \min\Big\{2\|\boldsymbol{U}\|\sqrt{B_T},\ \frac{B_T}{A_T}\|\boldsymbol{U}\|^2\Big\}$, where $A_T = \sum_{t=1}^T \frac{a_t\,\mathbb{1}[y_t \in \mathcal{U}_t]}{|\mathcal{U}_t|}$; $B_T = \sum_{t=1}^T a_t\,\mathbb{1}[y_t \in \mathcal{U}_t]\big(\|\boldsymbol{x}_t\|^2 + 2\sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle|\big)$, and ℓ_t denotes the multiclass hinge loss (7).*

Proof of Corollary 4. Theorem B.1 holds for all $\eta > 0$ simultaneously, since the algorithm's predictions do not depend on η. Writing the surrogate-gap sum as $-A_T + 2\eta B_T$, the bound becomes

$$\mathbb{E}[M_T] \le \sum_{t=1}^T \ell_t(\boldsymbol{U}) + \inf_{\eta > 0} \left\{ \frac{\|\boldsymbol{U}\|^2}{2\eta} + 2\eta B_T \right\} - A_T.$$

Setting $\eta = \|\boldsymbol{U}\|/(2\sqrt{B_T})$ yields the first bound $2\|\boldsymbol{U}\|\sqrt{B_T} - A_T$. Alternatively, choosing $\eta = A_T/(2B_T)$ makes the surrogate-gap sum $\frac{\|\boldsymbol{U}\|^2}{2\eta} - A_T + 2\eta B_T$ equal to $\frac{B_T}{A_T}\|\boldsymbol{U}\|^2$, yielding the second bound. □

B.2 Details of Section 3.2

The estimator inflates the gradient variance by the inverse sampling probability. On stable rounds ($a_t = 0$), the only mass on labels $y \ne y_t^\star$ comes from the uniform γ/K component, so $p_t'(y_t) \ge \gamma/K$ and $\frac{\eta}{2}\,\mathbb{E}_t[\|\widehat{\boldsymbol{g}}_t\|^2] \le \frac{\eta K}{\gamma}\|\boldsymbol{x}_t\|^2$. On unstable rounds ($a_t = 1$) with $y_t \in \mathcal{U}_t$, the $(1 - \gamma)$ component additionally puts mass $1/|\mathcal{U}_t|$ on each competitive label, giving $p_t'(y_t) \ge 1/K$ and the tighter bound $\frac{\eta}{2}\,\mathbb{E}_t[\|\widehat{\boldsymbol{g}}_t\|^2] \le \eta K\|\boldsymbol{x}_t\|^2$. On unstable rounds with $y_t \notin \mathcal{U}_t$, the large margin deficit $m_t(\boldsymbol{W}_t, y_t) \le -\beta_t$ ensures the loss term $-\,\mathbb{E}_t[\ell_t(\boldsymbol{W}_t(\eta))]$ absorbs the gradient cost in the surrogate-gap analysis.

Let $\mathcal{G}_t$ denote the filtration containing everything observed up to round t. For $s < t$, $\widehat{\boldsymbol{g}}_s$ is $\mathcal{G}_{t-1}$-measurable, while $\widehat{\boldsymbol{g}}_t$ depends only on the randomness at round t and satisfies $\mathbb{E}[\widehat{\boldsymbol{g}}_t \mid \mathcal{G}_{t-1}] = \boldsymbol{g}_t$. By the tower rule:

$$\mathbb{E}[\langle \widehat{\boldsymbol{g}}_s, \widehat{\boldsymbol{g}}_t \rangle] = \mathbb{E}[\langle \widehat{\boldsymbol{g}}_s\,, \mathbb{E}[\widehat{\boldsymbol{g}}_t \mid \mathcal{G}_{t-1}] \rangle] = \mathbb{E}[\langle \widehat{\boldsymbol{g}}_s, \boldsymbol{g}_t \rangle] = \mathbb{E}[\langle \boldsymbol{g}_s, \boldsymbol{g}_t \rangle]. \tag{10}$$

This separation underlies the additive form of β_t in line 3 of Algorithm 3.

Theorem B.2. *Let $\gamma \in (0,1)$, and ℓ_t multiclass hinge loss. Then for any $\boldsymbol{U} \in \mathcal{W}$, Algorithm 3 satisfies*

$$\mathbb{E}[M_T] \le \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(\boldsymbol{U})\right] + \gamma \frac{K-1}{K} T + \inf_{\eta>0}\left\{\frac{\|\boldsymbol{U}\|^2}{2\eta} + \sum_{t=1}^{T} a_t \mathbb{1}[y_t \in \mathcal{U}_t]\left(-\frac{1}{|\mathcal{U}_t|} + \eta\left(\frac{K}{\gamma} + \frac{|\mathcal{U}_t|}{1-\gamma}\right)\|\boldsymbol{x}_t\|^2 + 4\eta \sum_{s\in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|\right)\right\}.$$

Proof of Theorem B.2. Recall that,

$$b_t = \mathbb{1}[a_t = 1 \text{ or } y_t^\star \neq y_t], \qquad k_t^\star \in \arg\max_{k \neq y_t} \langle \boldsymbol{W}_t^k, \boldsymbol{x}_t\rangle, \qquad \boldsymbol{g}_t = b_t(\boldsymbol{e}_{k_t^\star} - \boldsymbol{e}_{y_t}) \otimes \boldsymbol{x}_t,$$

and define the importance weighted update,

$$\widehat{\boldsymbol{g}}_t := \frac{\mathbb{1}[y_t' = y_t]}{p_t'(y_t)} \boldsymbol{g}_t.$$

The comparator loss remains as in Equation (7), i.e. $\ell_t(\boldsymbol{W}) = b_t(1 - m_t(\boldsymbol{W}, y_t))_+$.

The parameter-freeness argument of Theorem B.1 carries over: the iterates satisfy $\boldsymbol{W}_t(\eta) = \eta\, \boldsymbol{W}_t$, the prediction distribution $\boldsymbol{p}_t'$ is η-invariant, and the importance-weighted update $\widehat{\boldsymbol{g}}_t = \frac{\mathbb{1}[y_t' = y_t]}{p_t'(y_t)} \boldsymbol{g}_t$ is η-invariant since p_t' and $\boldsymbol{g}_t$ are. Consequently the surrogate regret would be the same if we ran OGD with any step size η.

Fix any $\eta > 0$ and consider the iterate sequence $\boldsymbol{W}_t(\eta)$. Let $\boldsymbol{W}_t^\star(\eta) = -\eta \sum_{s=1}^{t-1} \widehat{\boldsymbol{g}}_s$ denote the iterate that would have received all feedback up to round t.

Let $\mathbb{E}_t$ denote expectation over the prediction at round t, conditional on all previous prediction draws and the fixed data and feedback schedule. We have

$$\mathbb{E}_t[\widehat{\boldsymbol{g}}_t] = \boldsymbol{g}_t, \qquad \mathbb{E}_t[\|\widehat{\boldsymbol{g}}_t\|^2] = \frac{\|\boldsymbol{g}_t\|^2}{p_t'(y_t)} = \frac{2b_t\|\boldsymbol{x}_t\|^2}{p_t'(y_t)}.$$

For every comparator $\boldsymbol{U}$,

$$\ell_t(\boldsymbol{U}) \ge b_t\big(1 + \langle \boldsymbol{U}^{k_t^\star} - \boldsymbol{U}^{y_t}, \boldsymbol{x}_t\rangle\big) = b_t + \langle \boldsymbol{g}_t, \boldsymbol{U}\rangle.$$

Since $\langle \boldsymbol{g}_t, \boldsymbol{W}_t(\eta)\rangle = -\eta b_t m_t(\boldsymbol{W}_t, y_t)$, it follows that

$$b_t(1 - \eta m_t(\boldsymbol{W}_t, y_t)) - \ell_t(\boldsymbol{U}) \le \langle \boldsymbol{g}_t, \boldsymbol{W}_t(\eta) - \boldsymbol{U}\rangle = \mathbb{E}_t[\langle \widehat{\boldsymbol{g}}_t, \boldsymbol{W}_t(\eta) - \boldsymbol{U}\rangle].$$

Let $\boldsymbol{W}_t^\star(\eta) = -\eta \sum_{s=1}^{t-1} \widehat{\boldsymbol{g}}_s$ denote the reference iterate formed from all previous update vectors. The linearized regret decomposes as

$$\sum_{t=1}^{T} \langle \widehat{\boldsymbol{g}}_t, \boldsymbol{W}_t(\eta) - \boldsymbol{U}\rangle = \sum_{t=1}^{T} \langle \widehat{\boldsymbol{g}}_t, \boldsymbol{W}_t^\star(\eta) - \boldsymbol{U}\rangle + \eta \sum_{t=1}^{T} \sum_{s\in n_t} \langle \widehat{\boldsymbol{g}}_s, \widehat{\boldsymbol{g}}_t\rangle.$$

The first term is bounded by $\frac{\|\boldsymbol{U}\|^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \|\widehat{\boldsymbol{g}}_t\|^2$ via, for example, [Orabona, 2025, Theorem 2.13].

For the delay cross-terms, we use Equation (10). Adding and subtracting $\mathbb{E}_t[\ell_t(\boldsymbol{W}_t(\eta))]$ and expanding the conditional mistake probability under $\boldsymbol{p}_t'$:

$$\mathbb{E}_t[\mathbb{1}[y_t' \neq y_t]] = (1-\gamma)\left[(1-a_t)\mathbb{1}[y_t \neq y_t^\star] + a_t\left(\mathbb{1}[y_t \in \mathcal{U}_t]\frac{|\mathcal{U}_t| - 1}{|\mathcal{U}_t|} + \mathbb{1}[y_t \notin \mathcal{U}_t]\right)\right] + \gamma \frac{K-1}{K}.$$

Combining and separating the $\gamma \frac{K-1}{K} T$ exploration cost,

$$\mathbb{E}[M_T] \le \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(\boldsymbol{U})\right] + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \gamma \frac{K-1}{K} T + (1-\gamma) \sum_{t=1}^{T} \mathbb{E}[S_t],$$

where

$$S_t = (1-a_t)\mathbb{1}[y_t \neq y_t^\star] + a_t\left(\mathbb{1}[y_t \in \mathcal{U}_t]\frac{|\mathcal{U}_t|-1}{|\mathcal{U}_t|} + \mathbb{1}[y_t \notin \mathcal{U}_t]\right)$$
$$- b_t(1-\eta m_t(\boldsymbol{W}_t, y_t)) + \frac{\eta}{2}\,\mathbb{E}_t[\|\widehat{\boldsymbol{g}}_t\|^2] + \eta\sum_{s\in n_t}\langle \boldsymbol{g}_s, \boldsymbol{g}_t\rangle.$$

We bound the cross-terms by $|\langle \boldsymbol{g}_s, \boldsymbol{g}_t\rangle| \le 2|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|$. For the second-moment term, on rounds with $b_t = 1$,

$$\frac{\eta}{2}\,\mathbb{E}_t[\|\widehat{\boldsymbol{g}}_t\|^2] = \frac{\eta\|\boldsymbol{x}_t\|^2}{p_t'(y_t)}.$$

We have $p_t'(y_t) \ge \gamma/K$ on every round, and $p_t'(y_t) \ge (1-\gamma)/|\mathcal{U}_t|$ when $a_t = 1$ and $y_t \in \mathcal{U}_t$. Recall

$$\beta_t = \frac{K}{\gamma}\|\boldsymbol{x}_t\|^2 + 2\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|.$$

Thus $a_t = 0$ means $m_t^\star > \beta_t$, and $y_t \notin \mathcal{U}_t$ implies $m_t(\boldsymbol{W}_t, y_t) \le -\beta_t$.

We show

$$\mathbb{E}_t[S_t] \le a_t\,\mathbb{1}[y_t \in \mathcal{U}_t]\left(-\frac{1}{|\mathcal{U}_t|} + \eta\left(\frac{K}{\gamma} + \frac{|\mathcal{U}_t|}{1-\gamma}\right)\|\boldsymbol{x}_t\|^2 + 4\eta\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|\right)$$

by case analysis.

$\boxed{a_t = 0,\ y_t^\star = y_t}$

Here $b_t = 0$, so the affine term and update vector are zero. Hence $S_t = 0$.

$\boxed{a_t = 0,\ y_t \neq y_t^\star}$

Here $b_t = 1$, and $m_t(\boldsymbol{W}_t, y_t) \le -m_t^\star$. Hence

$$S_t \le -\eta m_t^\star + \frac{\eta K}{\gamma}\|\boldsymbol{x}_t\|^2 + 2\eta\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|$$
$$= \eta(-m_t^\star + \beta_t) < 0,$$

where the last inequality uses $m_t^\star > \beta_t$.

$\boxed{a_t = 1,\ y_t \neq y_t^\star,\ y_t \notin \mathcal{U}_t}$

Here $b_t = 1$, and the non-exploration component makes a mistake with probability 1. Since $m_t(\boldsymbol{W}_t, y_t) \le -\beta_t$ and $p_t'(y_t) = \gamma/K$,

$$S_t \le 1 - (1+\eta\beta_t) + \frac{\eta K}{\gamma}\|\boldsymbol{x}_t\|^2 + 2\eta\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|$$
$$= -\eta\beta_t + \eta\beta_t = 0.$$

$\boxed{a_t = 1,\ y_t \neq y_t^\star,\ y_t \in \mathcal{U}_t}$

Here $b_t = 1$ and $m_t(\boldsymbol{W}_t, y_t) \le 0$, so $1 - \eta m_t(\boldsymbol{W}_t, y_t) \ge 1$. Moreover, $p_t'(y_t) \ge (1-\gamma)/|\mathcal{U}_t|$. Hence

$$S_t \le -\frac{1}{|\mathcal{U}_t|} + \frac{\eta|\mathcal{U}_t|}{1-\gamma}\|\boldsymbol{x}_t\|^2 + 2\eta\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|.$$

This is bounded above by the claimed expression.

$\boxed{a_t = 1,\ y_t^\star = y_t,\ y_t \in \mathcal{U}_t}$

Here $b_t = 1$, $m_t(\boldsymbol{W}_t, y_t) = m_t^\star \le \beta_t$, and $p_t'(y_t) \ge (1-\gamma)/|\mathcal{U}_t|$. Hence

$$\begin{aligned} S_t &\le \frac{|\mathcal{U}_t|-1}{|\mathcal{U}_t|} - (1-\eta m_t^\star) + \frac{\eta|\mathcal{U}_t|}{1-\gamma}\|\boldsymbol{x}_t\|^2 + 2\eta\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| \\ &= -\frac{1}{|\mathcal{U}_t|} + \eta m_t^\star + \frac{\eta|\mathcal{U}_t|}{1-\gamma}\|\boldsymbol{x}_t\|^2 + 2\eta\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| \\ &\le -\frac{1}{|\mathcal{U}_t|} + \eta\beta_t + \frac{\eta|\mathcal{U}_t|}{1-\gamma}\|\boldsymbol{x}_t\|^2 + 2\eta\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| \\ &= -\frac{1}{|\mathcal{U}_t|} + \eta\left(\frac{K}{\gamma} + \frac{|\mathcal{U}_t|}{1-\gamma}\right)\|\boldsymbol{x}_t\|^2 + 4\eta\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|. \end{aligned}$$

This is the binding case. The threshold contributes $\eta(K/\gamma)\|\boldsymbol{x}_t\|^2$ and one copy of the delay term; the second-moment and cross-term bounds contribute $\eta|\mathcal{U}_t|\|\boldsymbol{x}_t\|^2/(1-\gamma)$ and the other copy.

The case $y_t^\star = y_t$, $y_t \notin \mathcal{U}_t$ is impossible since $y_t^\star \in \mathcal{U}_t$. Combining the bounds (case 4 is dominated by case 5) gives

$$\begin{aligned} \mathbb{E}[M_T] \le \mathbb{E}\left[\sum_{t=1}^T \ell_t(\boldsymbol{U})\right] + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \gamma\frac{K-1}{K}T \\ + \mathbb{E}\left[\sum_{t=1}^T a_t \mathbb{1}[y_t \in \mathcal{U}_t]\left(-\frac{1}{|\mathcal{U}_t|} + \eta\left(\frac{K}{\gamma} + \frac{|\mathcal{U}_t|}{1-\gamma}\right)\|\boldsymbol{x}_t\|^2 + 4\eta\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|\right)\right]. \end{aligned}$$

All quantities inside the expectation, other than the displayed factors of η, are determined by the actual run. The bound therefore holds for every deterministic $\eta > 0$. Taking the infimum over $\eta > 0$ outside the expectation yields the theorem. □

Corollary 5 (RESTATED)**.** *Under Theorem B.2's conditions, for any comparator $\boldsymbol{U} \in \mathbb{R}^{K\times d}$, $\gamma \in (0,1]$, Algorithm 3 satisfies $\mathbb{E}[M_T] \le \mathbb{E}\left[\sum_{t=1}^T \ell_t(\boldsymbol{U})\right] + \min\left\{\|\boldsymbol{U}\|\sqrt{2B_T},\ \frac{B_T}{2A_T}\|\boldsymbol{U}\|^2\right\} + \gamma\,\frac{K-1}{K}\,T$, where $A_T = \sum_{t=1}^T \frac{a_t\,\mathbb{1}[y_t\in\mathcal{U}_t]}{|\mathcal{U}_t|}$; $B_T = \sum_{t=1}^T a_t\,\mathbb{1}[y_t \in \mathcal{U}_t]\left[\left(\frac{K}{\gamma} + \frac{|\mathcal{U}_t|}{1-\gamma}\right)\|\boldsymbol{x}_t\|^2 + 4\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|\right]$, and ℓ_t denotes the multiclass hinge loss (7)*

Proof of Corollary 5. Theorem B.2 holds for all $\eta > 0$ simultaneously. Writing the surrogate-gap sum as $-A_T + \eta B_T$, the bound becomes

$$\mathbb{E}[M_T] \le \mathbb{E}\left[\sum_{t=1}^T \ell_t(\boldsymbol{U})\right] + \inf_{\eta>0}\left\{\frac{\|\boldsymbol{U}\|^2}{2\eta} + \eta B_T\right\} - A_T + \gamma\,\frac{K-1}{K}\,T.$$

Setting $\eta = \|\boldsymbol{U}\|/\sqrt{2B_T}$ yields the first bound $\|\boldsymbol{U}\|\sqrt{2B_T} - A_T$. Alternatively, choosing $\eta = A_T/B_T$ makes the surrogate-gap sum $\frac{\|\boldsymbol{U}\|^2}{2\eta} - A_T + \eta B_T$ equal to $\frac{B_T}{2A_T}\|\boldsymbol{U}\|^2$, yielding the second bound. □

The structure of the bound reflects the additive separation in the instability threshold (line 3, Algorithm 3): the bandit variance contributes via the coefficient $\frac{K}{\gamma} + \frac{|\mathcal{U}_t|}{1-\gamma}$ on $\|\boldsymbol{x}_t\|^2$, while the delay cross-terms enter additively. On rounds where $|\mathcal{U}_t| \ll K$ the variance coefficient is much sharper than the worst-case $\frac{K}{\gamma(1-\gamma)}$, and when $D = 0$ the cross-terms vanish, recovering the non-delayed bandit setting.

In the non-delayed bandit setting ($D = 0$), van der Hoeven [2020, Theorem 5] achieves a worst-case mistake bound of $O(K\sqrt{T})$ for the multiclass hinge loss, but requires both η and γ to be set using a priori knowledge of X, K, $\|\boldsymbol{U}\|$, and T. Our algorithm is parameter-free in η: neither the prediction

distribution nor the instability indicator depend on η, so the bound holds simultaneously for all $\eta > 0$ and can be optimized in hindsight. While γ remains a genuine hyperparameter, the data-dependent bound of Corollary 5 adapts to the realized sequence: rounds with small $|\mathcal{U}_t|$ contribute via $|\mathcal{U}_t|/(1-\gamma)$ rather than $K/(1-\gamma)$, and rounds with large $\|\boldsymbol{x}_t\|$ or large $|n_t|$ are absorbed into averages over active rounds rather than dominating through uniform X^2 and K. van der Hoeven [2020] randomizes uniformly over all K classes on unstable rounds, incurring a per-round mistake probability of $(K-1)/K$; our focused exploration over $\mathcal{U}_t$ reduces this to $(|\mathcal{U}_t|-1)/|\mathcal{U}_t|$, reflected in the self-correction term $-1/|\mathcal{U}_t|$ rather than $-1/K$.

Remark B.2 (Worst-case delayed bandit bound). Suppose $\|\boldsymbol{U}\| \le R$ for some $R > 0$ and $\|\boldsymbol{x}_t\| \le X$ for all t. Then under the conditions of Corollary 5, using $|\mathcal{U}_t| \le K$ gives the variance bound $\frac{K}{\gamma} + \frac{|\mathcal{U}_t|}{1-\gamma} \le \frac{K}{\gamma(1-\gamma)}$, and the corollary's second bound implies

$$\mathbb{E}[M_T] \le \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(\boldsymbol{U})\right] + \frac{K^2X^2\|\boldsymbol{U}\|^2}{2\gamma(1-\gamma)} + 2KDX^2\,\|\boldsymbol{U}\|^2 + \gamma\,\frac{K-1}{K}\,T.$$

Setting $\gamma = \min\left\{\frac{1}{2},\ KXR\sqrt{\frac{K}{2(K-1)T}}\right\}$ yields the worst-case rate

$$\mathbb{E}[M_T] \le \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(\boldsymbol{U})\right] + XR\sqrt{2K(K-1)T} + 2KDX^2\,\|\boldsymbol{U}\|^2 + O(K^2X^2R^2).$$

Since $R \ge \|\boldsymbol{U}\|$ we obtain the bound. The bandit cost $O(K\sqrt{T})$ and the delay cost $O(KD)$ enter additively and independently: for any fixed D, the delay contribution remains bounded regardless of T.

B.3 Details of Section 3.3

Corollary 6 (RESTATED)**.** *Suppose $D = 0$. Let $\gamma = 0$, ℓ_t the multiclass hinge loss (7), and $\mathcal{A} = \{t : a_t = 1, y_t \in \mathcal{U}_t\}$. Then for any comparator sequence $\boldsymbol{U}_1, \dots, \boldsymbol{U}_T \in \mathbb{R}^{K\times d}$, Algorithm 3 achieves*

$$\mathbb{E}[M_T] \le \sum_{t=1}^{T} \ell_t(\boldsymbol{U}_t) + \min\left\{ \inf_{\bar{\boldsymbol{U}}\in\mathbb{R}^{K\times d}} \frac{2\sum_{t\in\mathcal{A}}\|\boldsymbol{x}_t\|^2}{\sum_{t\in\mathcal{A}}|\mathcal{U}_t|^{-1}}\left(\|\bar{\boldsymbol{U}}\|^2 + \sum_{t=1}^{T}\|\boldsymbol{U}_t - \bar{\boldsymbol{U}}\|^2\right),\right.$$
$$\left. \inf_{\bar{\boldsymbol{U}}\in\mathbb{R}^{K\times d}} 2\sqrt{2}\,X\sqrt{\left(\|\bar{\boldsymbol{U}}\|^2 + \sum_{t=1}^{T}\|\boldsymbol{U}_t - \bar{\boldsymbol{U}}\|^2\right)\sum_{t=1}^{T} a_t}\right\}.$$

Proof of Corollary 6. The parameter-freeness of Theorem B.1 carries over: we may analyze the iterates as if produced by OGD with any step size $\eta > 0$.

For every comparator $\boldsymbol{U}_t$,

$$\ell_t(\boldsymbol{U}_t) \ge b_t\big(1 + \langle \boldsymbol{U}_t^{k_t^\star} - \boldsymbol{U}_t^{y_t}, \boldsymbol{x}_t\rangle\big) = b_t + \langle \boldsymbol{g}_t, \boldsymbol{U}_t\rangle.$$

Since $\langle \boldsymbol{g}_t, \boldsymbol{W}_t(\eta)\rangle = -\eta b_t m_t(\boldsymbol{W}_t, y_t)$, for any fixed $\bar{\boldsymbol{U}} \in \mathcal{W}$,

$$b_t(1 - \eta m_t(\boldsymbol{W}_t, y_t)) - \ell_t(\boldsymbol{U}_t) \le \langle \boldsymbol{g}_t, \boldsymbol{W}_t(\eta) - \bar{\boldsymbol{U}}\rangle + \langle \boldsymbol{g}_t, \bar{\boldsymbol{U}} - \boldsymbol{U}_t\rangle.$$

The first inner product telescopes via the one-step OGD inequality [Orabona, 2025, Lemma 2.12]; for the second, Young's inequality gives

$$\langle \boldsymbol{g}_t, \bar{\boldsymbol{U}} - \boldsymbol{U}_t\rangle \le \frac{\eta}{2}\|\boldsymbol{g}_t\|^2 + \frac{1}{2\eta}\|\bar{\boldsymbol{U}} - \boldsymbol{U}_t\|^2.$$

Summing over $t = 1, \dots, T$ and using $\boldsymbol{W}_1 = \boldsymbol{0}$,

$$\sum_{t=1}^{T}\big(b_t(1 - \eta m_t(\boldsymbol{W}_t, y_t)) - \ell_t(\boldsymbol{U}_t)\big) \le \frac{\|\bar{\boldsymbol{U}}\|^2}{2\eta} + \frac{1}{2\eta}\sum_{t=1}^{T}\|\bar{\boldsymbol{U}} - \boldsymbol{U}_t\|^2 + \eta\sum_{t=1}^{T}\|\boldsymbol{g}_t\|^2.$$

Adding and subtracting $b_t(1 - \eta m_t(\boldsymbol{W}_t, y_t))$ and expanding the conditional mistake probability under $\boldsymbol{p}'_t$ (with $D = 0$ and $\gamma = 0$) gives

$$\mathbb{E}[M_T] \leq \sum_{t=1}^{T} \ell_t(\boldsymbol{U}_t) + \frac{\|\bar{\boldsymbol{U}}\|^2}{2\eta} + \frac{1}{2\eta}\sum_{t=1}^{T} \|\bar{\boldsymbol{U}} - \boldsymbol{U}_t\|^2 + \sum_{t=1}^{T} S_t,$$

where

$$\begin{aligned} S_t = (1 - a_t)\mathbb{1}[y_t \neq y_t^\star] + a_t \left(\mathbb{1}[y_t \in \mathcal{U}_t] \frac{|\mathcal{U}_t| - 1}{|\mathcal{U}_t|} + \mathbb{1}[y_t \notin \mathcal{U}_t] \right) \\ - b_t(1 - \eta m_t(\boldsymbol{W}_t, y_t)) + \eta \|\boldsymbol{g}_t\|^2. \end{aligned}$$

Note the factor $\eta\|\boldsymbol{g}_t\|^2$ (not $\frac{\eta}{2}\|\boldsymbol{g}_t\|^2$), arising from the OGD bound ($\frac{\eta}{2}$) plus the Young's inequality application ($\frac{\eta}{2}$). We show

$$S_t \leq a_t \mathbb{1}[y_t \in \mathcal{U}_t] \left(-\frac{1}{|\mathcal{U}_t|} + 4\eta \|\boldsymbol{x}_t\|^2 \right)$$

by case analysis, using $\|\boldsymbol{g}_t\|^2 = 2b_t\|\boldsymbol{x}_t\|^2 \leq 2\|\boldsymbol{x}_t\|^2$ throughout.

$\boxed{a_t = 0,\ y_t^\star = y_t}$

Here $b_t = 0$, hence $S_t = 0$.

$\boxed{a_t = 0,\ y_t \neq y_t^\star}$

Here $b_t = 1$. Since $y_t \neq y_t^\star$, $m_t(\boldsymbol{W}_t, y_t) \leq -m_t^\star$. Thus

$$\begin{aligned} S_t &\leq \eta m_t(\boldsymbol{W}_t, y_t) + 2\eta\|\boldsymbol{x}_t\|^2 \\ &\leq \eta(-m_t^\star + 2\|\boldsymbol{x}_t\|^2) < 0, \end{aligned}$$

where the last inequality uses $a_t = 0$, that is, $m_t^\star > \beta_t = 2\|\boldsymbol{x}_t\|^2$.

$\boxed{a_t = 1,\ y_t \neq y_t^\star,\ y_t \notin \mathcal{U}_t}$

Here $b_t = 1$. Since $y_t \notin \mathcal{U}_t$, $m_t(\boldsymbol{W}_t, y_t) \leq -\beta_t = -2\|\boldsymbol{x}_t\|^2$. Consequently,

$$S_t \leq \eta m_t(\boldsymbol{W}_t, y_t) + 2\eta\|\boldsymbol{x}_t\|^2 \leq 0.$$

$\boxed{a_t = 1,\ y_t \neq y_t^\star,\ y_t \in \mathcal{U}_t}$

Here $b_t = 1$. Since $m_t(\boldsymbol{W}_t, y_t) \leq 0$, $1 - \eta m_t(\boldsymbol{W}_t, y_t) \geq 1$. Hence

$$\begin{aligned} S_t &\leq \frac{|\mathcal{U}_t| - 1}{|\mathcal{U}_t|} - 1 + 2\eta\|\boldsymbol{x}_t\|^2 \\ &= -\frac{1}{|\mathcal{U}_t|} + 2\eta\|\boldsymbol{x}_t\|^2 \leq -\frac{1}{|\mathcal{U}_t|} + 4\eta\|\boldsymbol{x}_t\|^2. \end{aligned}$$

$\boxed{a_t = 1,\ y_t^\star = y_t,\ y_t \in \mathcal{U}_t}$

Here $b_t = 1$ and $m_t(\boldsymbol{W}_t, y_t) = m_t^\star \leq \beta_t = 2\|\boldsymbol{x}_t\|^2$. Hence

$$\begin{aligned} S_t &\leq \frac{|\mathcal{U}_t| - 1}{|\mathcal{U}_t|} - (1 - \eta m_t^\star) + 2\eta\|\boldsymbol{x}_t\|^2 \\ &= -\frac{1}{|\mathcal{U}_t|} + \eta m_t^\star + 2\eta\|\boldsymbol{x}_t\|^2 \\ &\leq -\frac{1}{|\mathcal{U}_t|} + 4\eta\|\boldsymbol{x}_t\|^2. \end{aligned}$$

This is the binding case: $\eta m_t^\star \le 2\eta\|\boldsymbol{x}_t\|^2$ from the widened threshold combines with $2\eta\|\boldsymbol{x}_t\|^2$ from the update cost.

The case $y_t^\star = y_t,\ y_t \notin \mathcal{U}_t$ is impossible since $y_t^\star \in \mathcal{U}_t$ by definition.

Let $C(\bar{\boldsymbol{U}}) := \|\bar{\boldsymbol{U}}\|^2 + \sum_{t=1}^T \|\boldsymbol{U}_t - \bar{\boldsymbol{U}}\|^2$. We have:

$$\mathbb{E}[M_T] \le \sum_{t=1}^T \ell_t(\boldsymbol{U}_t) + \frac{C(\bar{\boldsymbol{U}})}{2\eta} - \sum_{t\in\mathcal{A}} \frac{1}{|\mathcal{U}_t|} + 4\eta \sum_{t\in\mathcal{A}} \|\boldsymbol{x}_t\|^2 .$$

For the first bound, choose $\eta = \frac{\sum_{t\in\mathcal{A}} |\mathcal{U}_t|^{-1}}{4\sum_{t\in\mathcal{A}} \|\boldsymbol{x}_t\|^2}$ so that the last two terms cancel; then $\frac{1}{2\eta} = \frac{2\sum_{t\in\mathcal{A}} \|\boldsymbol{x}_t\|^2}{\sum_{t\in\mathcal{A}} |\mathcal{U}_t|^{-1}}$. For the second bound, drop the negative term and use $\sum_{t\in\mathcal{A}} \|\boldsymbol{x}_t\|^2 \le X^2 \sum_{t=1}^T a_t$ to obtain $\frac{C(\bar{\boldsymbol{U}})}{2\eta} + 4\eta X^2 \sum_{t=1}^T a_t$. This is minimised at $\eta = \frac{1}{2X}\sqrt{C(\bar{\boldsymbol{U}})/(2\sum_{t=1}^T a_t)}$, giving $2\sqrt{2}\, X \sqrt{C(\bar{\boldsymbol{U}}) \sum_{t=1}^T a_t}$, and taking $\inf_{\bar{\boldsymbol{U}}\in\mathcal{W}}$ yields the result for both bounds. □

The bound is small when the comparator sequence is tightly clustered. Choosing $\bar{\boldsymbol{U}} = \frac{1}{T}\sum_{t=1}^T \boldsymbol{U}_t$ and using the identity $\sum_{t=1}^T \|\boldsymbol{U}_t - \bar{\boldsymbol{U}}\|^2 = \sum_{t=1}^T \|\boldsymbol{U}_t\|^2 - T\|\bar{\boldsymbol{U}}\|^2$ gives an infimum over $\bar{\boldsymbol{U}}$ at most $\sum_{t=1}^T \|\boldsymbol{U}_t\|^2$. In the worst case, $\sum_{t\in\mathcal{A}} \|\boldsymbol{x}_t\|^2 / \sum_{t\in\mathcal{A}} |\mathcal{U}_t|^{-1} \le KX^2$, recovering a bound of $2KX^2\big(\|\bar{\boldsymbol{U}}\|^2 + \sum_{t=1}^T \|\boldsymbol{U}_t - \bar{\boldsymbol{U}}\|^2\big)$.

The parameter-freeness in η enables us to compare with $\bar{\boldsymbol{U}}^* = \frac{1}{T+1}\sum_{t=1}^T \boldsymbol{U}_t$, which yields

$$\inf_{\bar{\boldsymbol{U}}} \left(\frac{\|\bar{\boldsymbol{U}}\|^2}{2\eta} + \frac{1}{2\eta}\sum_{t=1}^T \|\boldsymbol{U}_t - \bar{\boldsymbol{U}}\|^2 \right) \le \frac{1}{2\eta}\left(\sum_{t=1}^T \|\boldsymbol{U}_t\|^2 - \frac{1}{T+1}\left\| \sum_{t=1}^T \boldsymbol{U}_t \right\|^2 \right),$$

which is small when the comparator sequence is tightly clustered.

The dominant notion of comparator-sequence variability in the dynamic regret literature is the path length $P_T = \sum_{t=2}^T \|\boldsymbol{U}_t - \boldsymbol{U}_{t-1}\|$, which yields bounds of the form $\sqrt{T(1+P_T)}$ for convex losses [Zinkevich, 2003, Orabona, 2025]. Our bound instead measures variability via the squared deviation $\sum_t \|\boldsymbol{U}_t - \bar{\boldsymbol{U}}\|^2$ around an analytic proxy $\bar{\boldsymbol{U}}$ minimized over $\mathcal{W}$ in the analysis. The two notions capture qualitatively different phenomena and are incomparable in general: for rapidly oscillating comparators with bounded magnitude, the path length grows much faster than the squared deviation. The squared-deviation form arises naturally from the Young's inequality step in the proof, and the parameter-freeness of η allows us to optimize over $\bar{\boldsymbol{U}}$ without prior knowledge of the comparator sequence's structure.

C Details of Logistic Loss (Section 4)

We first discuss Algorithm 4. The gradient of the logistic loss is

$$\boldsymbol{g}_t = \nabla \ell_t(\boldsymbol{W}_t) = \frac{1}{\log 2}(\widetilde{\boldsymbol{p}}_t - \boldsymbol{e}_{y_t}) \otimes \boldsymbol{x}_t, \tag{11}$$

so the k-th row is $\boldsymbol{g}_t^k = c_t^k\, \boldsymbol{x}_t$ where $c_t^k = \frac{1}{\log 2}(\widetilde{p}_t(k)) - \mathbb{1}[k = y_t])$. On unstable rounds, we sample from the softmax distribution $\widetilde{\boldsymbol{p}}_t$. This choice is natural for the logistic loss because the softmax is the loss-induced prediction distribution: $\widetilde{p}_t(k)$ is the model's predicted probability of $y_t = k$, and the loss $\log_2(1/\widetilde{p}_t(y_t))$ is small precisely when the algorithm places large probability on the correct label. Another benefit is that softmax sampling implicitly performs the role of the competitive set $\mathcal{U}_t$ used in the hinge case: classes with small $\widetilde{p}_t(k)$ (which are the non-competitive classes) receive exponentially small mass anyway, so an explicit hard restriction adds little. Uniform sampling over $[K]$ (the choice in the original `GAPTRON` [van der Hoeven, 2020]) would forfeit both properties: it decouples the mistake probability from the model's confidence, incurring the $(K-1)/K$ mistake rate per unstable round and the resulting K-dependence in the bound. Softmax sampling avoids this:

the bound is K-independent in the full-information setting and has $\sqrt{K}$-dependence in the bandit setting. The full prediction distribution at round t is

$$\boldsymbol{p}'_t = (1-\gamma)\big[(1-a_t)\,\boldsymbol{e}_{y_t^\star} + a_t\,\widetilde{\boldsymbol{p}}_t\big] + \frac{\gamma}{K}\mathbf{1}_K\,, \tag{12}$$

which places mass on $y_t^\star$ on stable rounds, samples from the softmax on unstable rounds, and mixes in a uniform-$[K]$ component of weight γ. In the full-information setting we set $\gamma = 0$; in the bandit setting we set $\gamma > 0$, since importance weighting requires $p'_t(y) > 0$ for every y (which otherwise would not be the case on stable rounds).

Recall that we define the margin of a label y as $m_t(\boldsymbol{W}_t, y) = \langle \boldsymbol{W}_t^y, \boldsymbol{x}_t\rangle - \max_{k\neq y}\langle \boldsymbol{W}_t^k, \boldsymbol{x}_t\rangle$. The prediction at round t under $\boldsymbol{W}_{t+1} = \boldsymbol{W}_t - \eta\,\boldsymbol{g}_t$ can flip from $y_t^\star$ to class k only if

$$m_t(\boldsymbol{W}_t, k) < \eta\,|c_t^{y_t^\star} - c_t^k|\,\|\boldsymbol{x}_t\|^2. \tag{13}$$

Since $\widetilde{p}_t(k) \in [0,1]$ and $\mathbb{1}[k = y_t] \in \{0,1\}$, we have $|c_t^{y_t^\star} - c_t^k| \leq \frac{2}{\log 2}$ uniformly. We define stability for the logistic loss as:

$$\max_k m_t(\boldsymbol{W}_t, k) = m_t^\star > \frac{2\eta}{\log 2}\|\boldsymbol{x}_t\|^2\,, \tag{14}$$

giving the non-delayed instability threshold $\beta_t = \frac{2\eta}{\log 2}\|\boldsymbol{x}_t\|^2$. This is enlarged with delay cross-terms in Section 4.1 and with a multiplicative bandit-variance factor in Section 4.2, analogous to Sections 3.1 and 3.2.

For the hinge loss, predictions and gradients depend on $\boldsymbol{W}_t$ only through orderings of inner products $\langle \boldsymbol{W}_t^k, \boldsymbol{x}_t\rangle$, which positive scaling preserves; this is what made a_t, $\mathcal{U}_t$, and $\boldsymbol{g}_t$ invariant under the choice of η. The softmax does not have this invariance: the inner products enter $\widetilde{\boldsymbol{p}}_t$ through the exponential rather than just through their ordering, so

$$\widetilde{p}_t(k) = \frac{\exp(\eta\langle \boldsymbol{W}_t^k, \boldsymbol{x}_t\rangle)}{\sum_{j=1}^K \exp(\eta\langle \boldsymbol{W}_t^j, \boldsymbol{x}_t\rangle)}$$

depends on η nonlinearly. As η grows, $\widetilde{\boldsymbol{p}}_t$ concentrates increasingly on $y_t^\star$; as η shrinks, it flattens toward the uniform distribution. The prediction distribution (12) and the importance weight $1/p'_t(y_t)$ inherit this η-dependence. This is structural to the logistic loss rather than a limitation of our approach: the same η-dependence appears in van der Hoeven [2020], Sakaue et al. [2024], and Shibukawa et al. [2025], whose iterates likewise depend on the learning rate. Compared to Algorithm 3, we therefore require an explicit choice of η, and the bound is no longer optimizable in hindsight.

The results in Section 4 rely on two elementary properties of the logistic loss. We have the following bound on the gradient (van der Hoeven [2020, Lemma 2]):

$$\|\nabla \ell_t(\boldsymbol{W}_t)\|^2 \leq \frac{2}{\log 2}\,\|\boldsymbol{x}_t\|^2\,\ell_t(\boldsymbol{W}_t)\,. \tag{15}$$

Lemma C.1. $\|\widetilde{\boldsymbol{p}}_t - \boldsymbol{e}_{y_t}\|^2 \leq 2(1-\widetilde{p}_t(y_t))$.

Proof. Write $p = \widetilde{p}_t(y_t)$. Since $\widetilde{p}_t(k) \in [0,1]$, we have $\widetilde{p}_t(k)^2 \leq \widetilde{p}_t(k)$ for each k, hence

$$\|\widetilde{\boldsymbol{p}}_t - \boldsymbol{e}_{y_t}\|^2 = (1-p)^2 + \sum_{k\neq y_t}\widetilde{p}_t(k)^2 \leq (1-p) + \sum_{k\neq y_t}\widetilde{p}_t(k) = 2(1-p)\,. \qquad \square$$

Lemma C.2. *For all* $p \in (0,1)$*:* $\dfrac{\log_2(1/p)}{1-p} > \dfrac{1}{\log 2}$.

Proof. The function $\log$ is strictly concave on $(0,\infty)$, so it lies strictly below its tangent line at any point. The tangent line at $x_0 = 1$ is $x - 1$, giving $\log p < p - 1$ for $p \in (0,1)$. Hence $-\log p > 1 - p > 0$. Dividing by $(1-p)\log 2 > 0$:

$$\frac{\log_2(1/p)}{1-p} = \frac{-\log p}{(1-p)\log 2} > \frac{1}{\log 2}\,. \qquad \square$$

We also use the following lower bound on the logistic loss. If $y_t^\star \neq y_t$, then

$$\ell_t(\boldsymbol{W}_t) = \log_2\Big(1 + \textstyle\sum_{k\neq y_t}\exp\big(\langle \boldsymbol{W}_t^k - \boldsymbol{W}_t^{y_t}, \boldsymbol{x}_t\rangle\big)\Big) \geq \log_2\big(1+\exp(m_t^\star)\big) \geq 1 + \frac{m_t^\star}{2\log 2}\,. \tag{16}$$

This inequality follows from convexity of $\log_2(1+e^x)$: the function lies above its tangent at $x_0 = 0$, which equals $1 + x/(2\log 2)$ since $\log_2(1+e^0) = 1$ and $\frac{d}{dx}\log_2(1+e^x)\big|_{x=0} = \frac{1}{2\log 2}$.

C.1 Details of Section 4.1

Bounding the cross-terms

Using the gradient structure (11) and Cauchy–Schwarz:

$$\begin{aligned} |\langle \boldsymbol{g}_s, \boldsymbol{g}_t \rangle| &= |\langle \boldsymbol{p}_s - \boldsymbol{e}_{y_s}, \boldsymbol{p}_t - \boldsymbol{e}_{y_t} \rangle \cdot \langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle| \\ &\le \frac{1}{(\log 2)^2} \|\boldsymbol{p}_s - \boldsymbol{e}_{y_s}\|_2 \, \|\boldsymbol{p}_t - \boldsymbol{e}_{y_t}\|_2 \, |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle| \\ &\le \frac{2}{(\log 2)^2} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle|, \end{aligned} \tag{17}$$

where the final step uses $\|\boldsymbol{p}_t - \boldsymbol{e}_{y_t}\|_2 \le \sqrt{2}$. Combined with (14), this yields the threshold in line 3 of Algorithm 4.

C.1.1 Proof of Theorem 7

Theorem 7 (RESTATED). *Let ℓ_t be the logistic loss (3), $\|\boldsymbol{x}_t\| \le X$, $\gamma = 0$, $D = \max_t |n_t|$. Then for any $\boldsymbol{U} \in \mathbb{R}^{K \times d}$, Algorithm 4 with $\eta = \frac{(\log 2)(1-\log 2)}{(1+2D)\,X^2}$ satisfies: $\mathbb{E}[M_T] \le \sum_{t=1}^{T} \ell_t(\boldsymbol{U}) + \frac{(1+2D)\,X^2\,\|\boldsymbol{U}\|^2}{2(\log 2)(1-\log 2)}$.*

Proof of Theorem 7. Let $\boldsymbol{W}_t^\star = -\eta \sum_{s=1}^{t-1} \boldsymbol{g}_s$ denote the iterate that would have received all feedback up to round t. By convexity of ℓ_t, the surrogate regret is upper bounded by the linearized regret $\sum_t \langle \boldsymbol{g}_t, \boldsymbol{W}_t - \boldsymbol{U} \rangle$, which decomposes via $\boldsymbol{W}_t - \boldsymbol{W}_t^\star = \eta \sum_{s \in n_t} \boldsymbol{g}_s$ as

$$\sum_{t=1}^{T} \langle \boldsymbol{g}_t, \boldsymbol{W}_t - \boldsymbol{U} \rangle = \sum_{t=1}^{T} \langle \boldsymbol{g}_t, \boldsymbol{W}_t^\star - \boldsymbol{U} \rangle + \eta \sum_{t=1}^{T} \sum_{s \in n_t} \langle \boldsymbol{g}_s, \boldsymbol{g}_t \rangle.$$

The first term is bounded by $\frac{\|\boldsymbol{U}\|^2}{2\eta} + \frac{\eta}{2} \sum_t \|\boldsymbol{g}_t\|^2$ via [Orabona, 2025, Theorem 2.13]. Adding and subtracting $\ell_t(\boldsymbol{W}_t)$ and expanding the conditional mistake probability under $\boldsymbol{p}'_t$ (with $\gamma = 0$) yields

$$\mathbb{E}[M_T] \le \sum_{t=1}^{T} \ell_t(\boldsymbol{U}) + \frac{\|\boldsymbol{U}\|^2}{2\eta} + \sum_{t=1}^{T} S_t,$$

where

$$S_t = (1 - a_t)\, \mathbb{1}[y_t^\star \ne y_t] + a_t\,(1 - \widetilde{p}_t(y_t)) - \ell_t(\boldsymbol{W}_t) + \frac{\eta}{2}\|\boldsymbol{g}_t\|^2 + \eta \sum_{s \in n_t} \langle \boldsymbol{g}_s, \boldsymbol{g}_t \rangle.$$

It suffices to show $S_t \le 0$ on each round.

By Lemma C.1, $\|\widetilde{\boldsymbol{p}}_t - \boldsymbol{e}_{y_t}\|^2 \le 2(1 - \widetilde{p}_t(y_t))$, so

$$\frac{\eta}{2}\|\boldsymbol{g}_t\|^2 = \frac{\eta \|\boldsymbol{x}_t\|^2}{2(\log 2)^2} \|\widetilde{\boldsymbol{p}}_t - \boldsymbol{e}_{y_t}\|^2 \le \frac{\eta \|\boldsymbol{x}_t\|^2}{(\log 2)^2}(1 - \widetilde{p}_t(y_t)), \tag{18}$$

and bounding $\|\widetilde{\boldsymbol{p}}_s - \boldsymbol{e}_{y_s}\| \le \sqrt{2}$ for $s \in n_t$,

$$\eta \sum_{s \in n_t} |\langle \boldsymbol{g}_s, \boldsymbol{g}_t \rangle| \le \frac{2\eta}{(\log 2)^2}(1 - \widetilde{p}_t(y_t))^{1/2} \sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle| \le \frac{2\eta}{(\log 2)^2} \sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle|. \tag{19}$$

$\boxed{a_t = 0,\ y_t^\star = y_t}$

The loss is non-negative and the first term is zero. Using (18) and (19),

$$S_t \le -\ell_t(\boldsymbol{W}_t) + \frac{\eta \|\boldsymbol{x}_t\|^2}{(\log 2)^2}(1 - \widetilde{p}_t(y_t)) + \frac{2\eta}{(\log 2)^2} \sum_{s \in n_t} |\langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle|.$$

By Lemma C.2, $\ell_t(\boldsymbol{W}_t) = \log_2(1/\widetilde{p}_t(y_t)) > (1-\widetilde{p}_t(y_t))/\log 2$, so dividing through by $1-\widetilde{p}_t(y_t) > 0$:

$$\frac{S_t}{1-\widetilde{p}_t(y_t)} < -\frac{1}{\log 2} + \frac{\eta\|\boldsymbol{x}_t\|^2}{(\log 2)^2} + \frac{2\eta}{(\log 2)^2}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|,$$

which is non-positive whenever $\eta(\|\boldsymbol{x}_t\|^2 + 2\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|) \le \log 2$.

$\boxed{a_t = 0,\ y_t^\star \neq y_t}$

The bound (16) combined with $a_t = 0$ (i.e., $m_t^\star > \beta_t$) gives

$$\ell_t(\boldsymbol{W}_t) > 1 + \frac{\eta\|\boldsymbol{x}_t\|^2}{(\log 2)^2} + \frac{2\eta}{(\log 2)^3}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|.$$

Combining with (18), (19), and $(1-\widetilde{p}_t(y_t)) \le 1$, the gradient and cross-term contributions cancel against the loss bound, leaving

$$S_t < \frac{2\eta}{(\log 2)^2}\left(1 - \frac{1}{\log 2}\right)\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| \le 0,$$

since $\frac{1}{\log 2} > 1$.

$\boxed{a_t = 1}$

Write $p = \widetilde{p}_t(y_t)$. For $p = 1$, S_t is trivially non-positive. Using (18), (19), and $(1-p) \le 1$,

$$S_t \le (1-p) - \log_2(1/p) + \frac{\eta\|\boldsymbol{x}_t\|^2}{(\log 2)^2}(1-p) + \frac{2\eta}{(\log 2)^2}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|.$$

For $p \in (0,1)$, dividing by $1-p > 0$ and applying Lemma C.2:

$$\frac{S_t}{1-p} < 1 - \frac{1}{\log 2} + \frac{\eta}{(\log 2)^2}\left(\|\boldsymbol{x}_t\|^2 + 2\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|\right),$$

which is non-positive whenever $\eta(\|\boldsymbol{x}_t\|^2 + 2\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|) \le (\log 2)(1-\log 2)$.

Using $\|\boldsymbol{x}_t\|^2 + 2\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| \le (1+2D)X^2$, the condition $\eta \le \frac{(\log 2)(1-\log 2)}{(1+2D)X^2}$ suffices to ensure $S_t \le 0$ in all cases (since $(\log 2)(1-\log 2) < \log 2$). The result follows by substituting into the surrogate-gap bound. □

Two features distinguish this result from the hinge loss analogue (Theorem B.1). First, the condition on η is independent of K: the ratio $\frac{\log_2(1/p)}{1-p} > \frac{1}{\log 2}$ (Lemma C.2) provides enough slack to absorb the gradient cost without a K-dependent correction. Second, the bound involves no self-correction term $-1/|\mathcal{U}_t|$; the surrogate gap vanishes entirely rather than being merely non-positive on average. The trade-off is that the algorithm does not exploit the competitive set structure, so the per-round cost on unstable rounds is $(1-\widetilde{p}_t(y_t))$ rather than $(|\mathcal{U}_t|-1)/|\mathcal{U}_t|$.

Remark C.3 (Non-delayed setting). When $D = 0$, the threshold reduces to $\beta_t = \frac{2\eta}{\log 2}\|\boldsymbol{x}_t\|^2$ and the condition becomes $\eta \le \frac{(\log 2)(1-\log 2)}{X^2}$. This improves the K-dependent condition $\eta \le \frac{\log 2}{KX^2}$ of van der Hoeven [2020, Theorem 1] and matches the surrogate regret bound of Sakaue et al. [2024, Theorem 12], who achieve $\frac{X^2\|\boldsymbol{U}\|^2}{2\log 2(1-\log 2)}$ using randomized decoding with Fenchel–Young losses. However, their algorithm randomizes on every round, whereas our algorithm randomizes only on unstable rounds.

Comparison with Shibukawa et al. [2025]. Shibukawa et al. [2025, Theorems 4.3 and E.3] obtain a surrogate regret bound of $O(\min\{D^2, D^{2/3}T^{1/3}\})$ for delayed full-information feedback using the ODAFTRL algorithm, and Shibukawa et al. [2025, Section 4.2] achieve $O(D+1)$ using BOLD, which maintains $D+1$ independent copies of the base learner. Our bound $(1+2D)X^2\|\boldsymbol{U}\|^2/(2\log 2(1-\log 2))$ is also $O(D)$ and independent of both T and K. The key difference is algorithmic: BOLD runs $D+1$ independent learners, each observing only $T/(D+1)$ examples, while our algorithm maintains a single model that processes every round. When T is not much larger than D, each BOLD copy has very few examples, whereas our approach benefits from the full sequence. Furthermore, our instability indicator adapts to the data geometry: when the pending features are nearly orthogonal to $\boldsymbol{x}_t$, the threshold is small and the algorithm predicts deterministically, whereas BOLD and ODAFTRL randomize on every round regardless of the margin.

C.2 Details of Section 4.2

Theorem 8 (RESTATED). *Let ℓ_t be the logistic loss (3), $\|\boldsymbol{x}_t\| \le X$, $\gamma \in (0,1)$, and $D = \max_t |n_t|$. Suppose $\boldsymbol{U} \in \mathbb{R}^{K\times d}$ with $\|\boldsymbol{U}\| \le R$ for some $R > 0$. Then Algorithm 4 with $\eta = \frac{(\log 2)(1-\log 2)}{(K/\gamma+2D)\,X^2}$ and $\gamma = \min\Big\{1,\ R\,X\sqrt{\frac{K}{2(\log 2)(1-\log 2)\,T}}\Big\}$ satisfies: $\mathbb{E}[M_T] \le \sum_{t=1}^T \ell_t(\boldsymbol{U}) + R\,X\sqrt{\frac{2K\,T}{(\log 2)(1-\log 2)}} + \frac{D\,X^2\,R^2}{(\log 2)(1-\log 2)} + O(KX^2R^2)$.*

Proof of Theorem 8. Since $p'_t(y) \ge \gamma/K > 0$ for all y, the importance-weighted gradient $\widehat{\boldsymbol{g}}_t = \frac{\mathbb{1}[y'_t=y_t]}{p'_t(y_t)}\boldsymbol{g}_t$ satisfies $\mathbb{E}_t[\widehat{\boldsymbol{g}}_t] = \boldsymbol{g}_t$. Following the proof of Theorem 7 with $\widehat{\boldsymbol{g}}_t$ in place of $\boldsymbol{g}_t$, and using the cross-term reduction $\mathbb{E}[\langle \widehat{\boldsymbol{g}}_s, \widehat{\boldsymbol{g}}_t\rangle \mid \boldsymbol{g}_s, \boldsymbol{g}_t] = \langle \boldsymbol{g}_s, \boldsymbol{g}_t\rangle$ (Equation (10)), it suffices to show $S_t \le 0$ where

$$S_t = (1-\gamma)\big[(1-a_t)\,\mathbb{1}[y_t^\star \ne y_t] + a_t\,(1-\widetilde{p}_t(y_t))\big] - \ell_t(\boldsymbol{W}_t) + \frac{\eta}{2}\frac{\|\boldsymbol{g}_t\|^2}{p'_t(y_t)} + \eta\sum_{s\in n_t}\langle \boldsymbol{g}_s, \boldsymbol{g}_t\rangle.$$

By Lemma C.1,

$$\frac{\|\boldsymbol{g}_t\|^2}{p'_t(y_t)} \le \frac{2\,\|\boldsymbol{x}_t\|^2}{(\log 2)^2}\,\frac{(1-\widetilde{p}_t(y_t))}{p'_t(y_t)}, \tag{20}$$

and bounding $\|\widetilde{\boldsymbol{p}}_s - \boldsymbol{e}_{y_s}\| \le \sqrt{2}$ for $s \in n_t$,

$$\eta\sum_{s\in n_t}|\langle \boldsymbol{g}_s, \boldsymbol{g}_t\rangle| \le \frac{2\eta}{(\log 2)^2}\,(1-\widetilde{p}_t(y_t))^{1/2}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| \le \frac{2\eta}{(\log 2)^2}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|. \tag{21}$$

$\boxed{a_t = 0,\ y_t^\star \ne y_t}$

Here $p'_t(y_t) = \gamma/K$ and the first term is $(1-\gamma)$. The stability condition $a_t = 0$ gives $m_t^\star > \frac{2\eta K}{\gamma\,\log 2}\|\boldsymbol{x}_t\|^2 + \frac{4\eta}{(\log 2)^2}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|$, so by (16):

$$\ell_t(\boldsymbol{W}_t) > 1 + \frac{\eta K\,\|\boldsymbol{x}_t\|^2}{(\log 2)^2\,\gamma} + \frac{2\eta}{(\log 2)^3}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|.$$

Using (20) with $p'_t(y_t) = \gamma/K$ and $(1-\widetilde{p}_t(y_t)) \le 1$, and (21):

$$\begin{aligned} S_t &< (1-\gamma) - 1 - \frac{\eta K\,\|\boldsymbol{x}_t\|^2}{(\log 2)^2\,\gamma} - \frac{2\eta}{(\log 2)^3}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| \\ &\quad + \frac{\eta K\,\|\boldsymbol{x}_t\|^2}{(\log 2)^2\,\gamma}(1-\widetilde{p}_t(y_t)) + \frac{2\eta}{(\log 2)^2}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| \\ &= -\gamma - \frac{\eta K\,\|\boldsymbol{x}_t\|^2}{(\log 2)^2\,\gamma}\,\widetilde{p}_t(y_t) + \frac{2\eta}{(\log 2)^2}\left(1 - \frac{1}{\log 2}\right)\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle| \le 0, \end{aligned}$$

since $\frac{1}{\log 2} > 1$.

$\boxed{a_t = 0,\ y_t^\star = y_t}$

Here $p'_t(y_t) = (1-\gamma) + \gamma/K \geq 1-\gamma$ and the first term is 0. Using (20) with $(1-\widetilde{p}_t(y_t)) < 1$ and (21):

$$S_t \leq -\ell_t(\boldsymbol{W}_t) + \frac{\eta\,\|\boldsymbol{x}_t\|^2}{(\log 2)^2\,(1-\gamma)}(1-\widetilde{p}_t(y_t)) + \frac{2\eta}{(\log 2)^2}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|.$$

Dividing by $1-\widetilde{p}_t(y_t) > 0$ and applying Lemma C.2:

$$\frac{S_t}{1-\widetilde{p}_t(y_t)} < -\frac{1}{\log 2} + \frac{\eta\,X^2}{(\log 2)^2\,(1-\gamma)} + \frac{2\eta\,DX^2}{(\log 2)^2} \leq 0,$$

whenever $\eta \leq \frac{(1-\gamma)\,\log 2}{(1+2D(1-\gamma))\,X^2}$, which is implied by $\eta = \frac{(\log 2)(1-\log 2)}{(K/\gamma+2D)\,X^2}$ since $K/\gamma \geq K \geq 2 > 1/(1-\gamma)$.

$\boxed{a_t = 1}$

Here $p'_t(y_t) = (1-\gamma)\widetilde{p}_t(y_t) + \gamma/K \geq \gamma/K$ and the first term is $(1-\gamma)(1-\widetilde{p}_t(y_t))$. Using (20) with $p'_t(y_t) \geq \gamma/K$, $(1-\widetilde{p}_t(y_t)) \leq 1$, and (21):

$$S_t \leq (1-\gamma)(1-\widetilde{p}_t(y_t)) - \ell_t(\boldsymbol{W}_t) + \frac{\eta K\,\|\boldsymbol{x}_t\|^2}{(\log 2)^2\,\gamma}(1-\widetilde{p}_t(y_t)) + \frac{2\eta}{(\log 2)^2}\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|.$$

Dividing by $1-\widetilde{p}_t(y_t) > 0$ and applying Lemma C.2:

$$\frac{S_t}{1-\widetilde{p}_t(y_t)} < (1-\gamma) - \frac{1}{\log 2} + \frac{\eta\,X^2}{(\log 2)^2}\left(\frac{K}{\gamma} + 2D\right) \leq 0,$$

whenever $\eta \leq \frac{(\log 2)(1-\log 2+\gamma\,\log 2)}{(K/\gamma+2D)\,X^2}$, which is satisfied by $\eta = \frac{(\log 2)(1-\log 2)}{(K/\gamma+2D)\,X^2}$.

Hence $S_t \leq 0$ in all cases. The comparator term decomposes as $\frac{K\,X^2\,\|\boldsymbol{U}\|^2}{2(\log 2)(1-\log 2)\,\gamma} + \frac{D\,X^2\,\|\boldsymbol{U}\|^2}{(\log 2)(1-\log 2)}$, where only the first depends on γ. Assuming $\|\boldsymbol{U}\| \leq R$ and minimizing $\frac{K\,X^2\,\|\boldsymbol{U}\|^2}{2(\log 2)(1-\log 2)\,\gamma} + \gamma\,\frac{K-1}{K}\,T$ over $\gamma > 0$ gives the stated bound. □

The delay contribution $\frac{DX^2\|\boldsymbol{U}\|^2}{(\log 2)(1-\log 2)}$ is independent of both γ and T, confirming the decoupling: optimizing γ affects only the bandit term. The delay cost is also independent of K, since the cross-terms use the full-information gradients by (10).

Comparison with Shibukawa et al. [2025]. When $D = 0$, the bound reduces to $\sum_t \ell_t(\boldsymbol{U}) + \|\boldsymbol{U}\|X\sqrt{\frac{2KT}{(\log 2)(1-\log 2)}}$, which matches Shibukawa et al. [2025, Theorem 3.5]. In the delayed bandit setting, Shibukawa et al. [2025, Theorem 5.1] achieve a surrogate regret of $O(\sqrt{(K+D)T})$, where the delay inflates the effective K inside the square root. For any fixed D, their delay cost grows as $O(\sqrt{DT})$, which diverges with T. Our bound separates the two costs entirely: the bandit cost $O(\sqrt{KT})$ and the delay cost $O(D)$ are independent, so the delay contribution remains bounded regardless of T. This separation is a consequence of (10): the importance-weighting variance inflates only the diagonal gradient term, and the instability threshold (line 3, Algorithm 4) assigns different coefficients to the two components. Shibukawa et al. [2025, Theorem 5.2] also achieve an $O(D^{1/3}T^{2/3})$ bound that is independent of K, using a pseudo-inverse matrix estimator; our bound retains $\sqrt{K}$ dependence through the bandit term but has the superior $\sqrt{T}$ rate. An advantage of their framework is the more general setting of online structured prediction with Fenchel–Young losses.

C.3 Dynamic Surrogate Regret Bound for the Logistic Loss

The dynamic regret bound follows by the same deviation argument as Corollary 6.

Corollary C.3. *Suppose $D = 0$. Let $\gamma = 0$, and ℓ_t the logistic loss (3). Then for any comparator sequence $\boldsymbol{U}_1, \dots, \boldsymbol{U}_T \in \mathbb{R}^{K\times d}$ setting $\eta = \frac{(\log 2)(1-\log 2)}{2X^2}$ yields:* $\mathbb{E}[M_T] \le \sum_{t=1}^T \ell_t(\boldsymbol{U}_t) + \inf_{\bar{\boldsymbol{U}}\in\mathcal{W}} \frac{X^2}{(\log 2)(1-\log 2)} \Big(\|\bar{\boldsymbol{U}}\|^2 + \sum_{t=1}^T \|\boldsymbol{U}_t - \bar{\boldsymbol{U}}\|^2 \Big)$.

The dynamic surrogate regret is controlled entirely by the deviation of the comparator sequence around the analytic proxy $\bar{\boldsymbol{U}}$, with no K-dependence.

Proof of Corollary C.3. By convexity, for any fixed $\bar{\boldsymbol{U}} \in \mathcal{W}$:

$$\ell_t(\boldsymbol{W}_t) - \ell_t(\boldsymbol{U}_t) \le \langle \boldsymbol{g}_t, \boldsymbol{W}_t - \bar{\boldsymbol{U}} \rangle + \langle \boldsymbol{g}_t, \bar{\boldsymbol{U}} - \boldsymbol{U}_t \rangle .$$

Telescoping the first term via OGD and applying Young's inequality to the second:

$$\sum_{t=1}^T \big(\ell_t(\boldsymbol{W}_t) - \ell_t(\boldsymbol{U}_t)\big) \le \frac{\|\bar{\boldsymbol{U}}\|^2}{2\eta} + \frac{1}{2\eta}\sum_{t=1}^T \|\bar{\boldsymbol{U}} - \boldsymbol{U}_t\|^2 + \eta \sum_{t=1}^T \|\boldsymbol{g}_t\|^2 .$$

Adding and subtracting $\ell_t(\boldsymbol{W}_t)$ and expanding the conditional mistake probability gives

$$\mathbb{E}[M_T] \le \sum_{t=1}^T \ell_t(\boldsymbol{U}_t) + \frac{\|\bar{\boldsymbol{U}}\|^2}{2\eta} + \frac{1}{2\eta}\sum_{t=1}^T \|\bar{\boldsymbol{U}} - \boldsymbol{U}_t\|^2 + \sum_{t=1}^T S_t ,$$

where, with $D = 0$ and $\gamma = 0$, the surrogate gap is

$$S_t \;=\; (1 - a_t)\, \mathbb{1}[y_t^\star \neq y_t] + a_t\,(1 - \widetilde{p}_t(y_t)) - \ell_t(\boldsymbol{W}_t) + \eta\, \|\boldsymbol{g}_t\|^2 .$$

Compared to the surrogate gap in Theorem 7, the gradient term is doubled ($\eta\|\boldsymbol{g}_t\|^2$ rather than $\frac{\eta}{2}\|\boldsymbol{g}_t\|^2$). The case analysis proceeds as in Theorem 7 with $D = 0$, but with the doubled gradient cost.

$\boxed{a_t = 0,\ y_t^\star = y_t}$

Dividing by $(1 - \widetilde{p}_t(y_t)) > 0$ and applying Lemma C.2:

$$\frac{S_t}{1 - \widetilde{p}_t(y_t)} < -\frac{1}{\log 2} + \frac{2\eta\, \|\boldsymbol{x}_t\|^2}{(\log 2)^2} \le 0$$

whenever $\eta \le \frac{\log 2}{2X^2}$.

$\boxed{a_t = 0,\ y_t^\star \neq y_t}$

The stability condition $m_t^\star > \frac{4\eta}{\log 2}\|\boldsymbol{x}_t\|^2$ gives, by (16): $\ell_t(\boldsymbol{W}_t) > 1 + \frac{2\eta\,\|\boldsymbol{x}_t\|^2}{(\log 2)^2}$. Hence $S_t < 1 - 1 - \frac{2\eta\,\|\boldsymbol{x}_t\|^2}{(\log 2)^2} + \frac{2\eta\,\|\boldsymbol{x}_t\|^2}{(\log 2)^2} = 0$.

$\boxed{a_t = 1}$

Dividing by $(1 - \widetilde{p}_t(y_t)) > 0$ and applying Lemma C.2:

$$\frac{S_t}{1 - \widetilde{p}_t(y_t)} < 1 - \frac{1}{\log 2} + \frac{2\eta\, X^2}{(\log 2)^2} \le 0$$

whenever $\eta \le \frac{(\log 2)(1-\log 2)}{2X^2}$.

Since $S_t \le 0$ in all cases, the surrogate gap sum vanishes, and taking the infimum over $\bar{\boldsymbol{U}}$ yields the result. □

Choosing $\bar{\boldsymbol{U}} = \frac{1}{T}\sum_{t=1}^T \boldsymbol{U}_t$ and using $\sum_{t=1}^T \|\boldsymbol{U}_t - \bar{\boldsymbol{U}}\|^2 = \sum_{t=1}^T \|\boldsymbol{U}_t\|^2 - T\|\bar{\boldsymbol{U}}\|^2$ shows that the infimum is at most $\frac{X^2}{(\log 2)(1-\log 2)} \sum_{t=1}^T \|\boldsymbol{U}_t\|^2$. The bound is independent of K, inheriting the K-independence of the static logistic loss result.

Comparison with Sakaue et al. [2025]. Concurrent work by Sakaue et al. [2025] extends the surrogate-gap framework to non-stationary online structured prediction with full-information feedback, obtaining a bound of the form $\sum_t \ell_t(\boldsymbol{U}_t) + C(1 + P_T)$, where $P_T = \sum_{t=1}^{T-1} \|\boldsymbol{U}_{t+1} - \boldsymbol{U}_t\|$ is the path length of the comparator sequence. Their analysis covers general structured prediction via convolutional Fenchel–Young losses; ours is specific to the multiclass logistic loss (and multiclass hinge loss). The two bounds measure comparator variability differently: their P_T tracks consecutive-comparator distances, whereas ours tracks the squared deviation $\sum_t \|\boldsymbol{U}_t - \bar{\boldsymbol{U}}\|^2$ around an analytic proxy $\bar{\boldsymbol{U}}$, which we minimize over $\mathcal{W}$. The two bounds are incomparable in general: $\sum_t \|\boldsymbol{U}_t\|^2$ can be much smaller than P_T for rapidly oscillating sequences with bounded magnitude, while P_T can be much smaller than $\sum_t \|\boldsymbol{U}_t\|^2$ for stationary or slowly drifting sequences with large magnitude.

D Details of Stability and Empirical Risk Minimization (Section 5)

Both results follow the same proof strategy: a leave-one-out symmetrisation argument due to Koren and Levy [2015] reduces the classification error to a per-sample surrogate loss difference, parameter stability of the regularised ERM controls this difference, and a pointwise indicator-to-surrogate inequality converts the resulting bound into a classification error guarantee. The two theorems differ only in the choice of instability threshold $\beta(\boldsymbol{x})$ and the per-loss Lipschitz constant entering the stability lemma.

Suppose we observe an independent datapoint $(\boldsymbol{x}_{T+1}, y_{T+1})$. Let $\phi^{-i}(\boldsymbol{W}) = \sum_{t\in[T+1]\setminus\{i\}} \ell_t(\boldsymbol{W}) + \frac{1}{2\eta}\|\boldsymbol{W}\|^2$ denote the regularised objective with the i-th sample removed, and write $\widehat{\boldsymbol{W}}_i = \arg\min_{\boldsymbol{W}\in\mathbb{R}^d} \phi^{-i}(\boldsymbol{W})$ for its minimiser. Likewise, let $\widehat{\boldsymbol{W}}_\star = \arg\min_{\boldsymbol{W}\in\mathbb{R}^d} \phi(\boldsymbol{W})$ denote the minimiser of the full-sample objective $\phi(\boldsymbol{W}) = \sum_{t=1}^{T+1} \ell_t(\boldsymbol{W}) + \frac{1}{2\eta}\|\boldsymbol{W}\|^2$.

D.1 Hinge Loss (Proof of Theorem 9)

Recall that $\beta(\boldsymbol{x}) = \eta\|\boldsymbol{x}\|^2$ and in this section we work with the loss $\ell(\boldsymbol{W}, \boldsymbol{x}, y) = (1 - y\langle \boldsymbol{W}, \boldsymbol{x}\rangle)_+$, and $\ell_t(\boldsymbol{W}) = \ell(\boldsymbol{W}, \boldsymbol{x}_t, y_t)$. We require the following two lemmas.

Lemma D.1 (Pointwise indicator-to-hinge)**.** *For any $\boldsymbol{W} \in \mathbb{R}^d$ and any $(\boldsymbol{x}, y) \in \mathbb{R}^d \times \{-1, +1\}$ the prediction y' defined relative to $\boldsymbol{W}$ satisfies*

$$\mathbb{E}_{y'}[\mathbb{1}[y \neq y'(\boldsymbol{x}; \boldsymbol{W})]] \leq (1 - y\langle \boldsymbol{W} + \eta y \boldsymbol{x}, \boldsymbol{x}\rangle)_+ , \tag{22}$$

provided that $\eta \leq \frac{1}{4X^2}$.

Proof of Lemma D.1. We write $\ell_+ = (1 - y\langle \boldsymbol{W}, \boldsymbol{x}\rangle - \eta\|\boldsymbol{x}\|^2)_+$ for the right-hand side of (22) and $p = \mathbb{E}[\mathbb{1}[y \neq y'(\boldsymbol{x}; \boldsymbol{W})]]$ for the left-hand side. We show $p \leq \ell_+$ by splitting into four cases based on the magnitude of $\langle \boldsymbol{W}, \boldsymbol{x}\rangle$ relative to the threshold $\eta\|\boldsymbol{x}\|^2$ and on whether $\operatorname{sign}(\langle \boldsymbol{W}, \boldsymbol{x}\rangle)$ agrees with y.

$$\boxed{|\langle \boldsymbol{W}, \boldsymbol{x}\rangle| > \eta\|\boldsymbol{x}\|^2 \text{ and } \operatorname{sign}(\langle \boldsymbol{W}, \boldsymbol{x}\rangle) = y}$$

The predictor outputs y deterministically, so $p = 0$. The bound $p \leq \ell_+$ holds trivially since $\ell_+ \geq 0$.

$$\boxed{|\langle \boldsymbol{W}, \boldsymbol{x}\rangle| > \eta\|\boldsymbol{x}\|^2 \text{ and } \operatorname{sign}(\langle \boldsymbol{W}, \boldsymbol{x}\rangle) \neq y}$$

The predictor outputs $-y$ deterministically, so $p = 1$. We have $y\langle \boldsymbol{W}, \boldsymbol{x}\rangle = -|\langle \boldsymbol{W}, \boldsymbol{x}\rangle| < -\eta\|\boldsymbol{x}\|^2$, so

$$\ell_+ = 1 + |\langle \boldsymbol{W}, \boldsymbol{x}\rangle| - \eta\|\boldsymbol{x}\|^2 \geq 1 + \eta\|\boldsymbol{x}\|^2 - \eta\|\boldsymbol{x}\|^2 = 1.$$

Hence $p \leq \ell_+$.

$$\boxed{|\langle \boldsymbol{W}, \boldsymbol{x}\rangle| \leq \eta\|\boldsymbol{x}\|^2 \text{ and } \operatorname{sign}(\langle \boldsymbol{W}, \boldsymbol{x}\rangle) \neq y}$$

The predictor is uniform on $\{-1, +1\}$, so $p = 1/2$. We have $y\langle \boldsymbol{W}, \boldsymbol{x}\rangle \leq 0$, so

$$\ell_+ \geq 1 - \eta\|\boldsymbol{x}\|^2 \geq \tfrac{1}{2}$$

provided $\eta\|\boldsymbol{x}\|^2 \le 1/2$, which follows from $\eta \le 1/(4X^2)$.

$\boxed{|\langle \boldsymbol{W}, \boldsymbol{x}\rangle| \le \eta\|\boldsymbol{x}\|^2 \text{ and } \operatorname{sign}(\langle \boldsymbol{W}, \boldsymbol{x}\rangle) = y}$

The predictor is uniform on $\{-1, +1\}$, so $p = 1/2$. We have $y\langle \boldsymbol{W}, \boldsymbol{x}\rangle = |\langle \boldsymbol{W}, \boldsymbol{x}\rangle| \le \eta\|\boldsymbol{x}\|^2$, so

$$\ell_+ \ge 1 - \eta\|\boldsymbol{x}\|^2 - \eta\|\boldsymbol{x}\|^2 \ge 1 - 2\eta\|\boldsymbol{x}\|^2 \ge \tfrac{1}{2}$$

provided $\eta\|\boldsymbol{x}\|^2 \le 1/4$, which follows from $\eta \le 1/(4X^2)$.

In all four cases $p \le \ell_+$. □

The difference between $\widehat{\boldsymbol{W}}_i$ and $\widehat{\boldsymbol{W}}_\star$ measures the stability of ERM to the removal of a single sample. We bound this difference using the strong convexity of the regularized objective in the following lemma.

Lemma D.2 (Parameter stability)**.** *For each $i \in [T+1]$,*

$$\|\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star\| \le \eta\|\boldsymbol{x}_i\|. \tag{23}$$

Proof of Lemma D.2. Since ϕ is $\frac{1}{\eta}$-strongly convex and $\widehat{\boldsymbol{W}}_\star$ minimises ϕ, strong convexity at the minimiser gives

$$\phi(\widehat{\boldsymbol{W}}_i) - \phi(\widehat{\boldsymbol{W}}_\star) \ge \tfrac{1}{2\eta}\|\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star\|^2. \tag{24}$$

For the upper bound, expand ϕ to second order around $\widehat{\boldsymbol{W}}_i$, with $\boldsymbol{z}$ on the segment between $\widehat{\boldsymbol{W}}_\star$ and $\widehat{\boldsymbol{W}}_i$:

$$\phi(\widehat{\boldsymbol{W}}_\star) = \phi(\widehat{\boldsymbol{W}}_i) - \nabla\phi(\widehat{\boldsymbol{W}}_i)^\top(\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star) + \tfrac{1}{2}(\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star)^\top\nabla^2\phi(\boldsymbol{z})(\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star).$$

Optimality of $\widehat{\boldsymbol{W}}_i$ for ϕ^{-i} yields $\nabla\phi(\widehat{\boldsymbol{W}}_i) = \nabla\phi^{-i}(\widehat{\boldsymbol{W}}_i) + \nabla\ell_i(\widehat{\boldsymbol{W}}_i) = \nabla\ell_i(\widehat{\boldsymbol{W}}_i)$, and $\nabla^2\phi(\boldsymbol{z}) \succeq \frac{1}{\eta}I_d$ where I_d denotes the d-dimensional identity matrix, so

$$\phi(\widehat{\boldsymbol{W}}_i) - \phi(\widehat{\boldsymbol{W}}_\star) \le \nabla\ell_i(\widehat{\boldsymbol{W}}_i)^\top(\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star) - \tfrac{1}{2\eta}\|\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star\|^2.$$

Young's inequality, $a^\top b \le \frac{\eta}{2}\|a\|^2 + \frac{1}{2\eta}\|b\|^2$, applied to the linear term, together with $\|\nabla\ell_i(\widehat{\boldsymbol{W}}_i)\| \le \|\boldsymbol{x}_i\|$ for the hinge loss, gives

$$\phi(\widehat{\boldsymbol{W}}_i) - \phi(\widehat{\boldsymbol{W}}_\star) \le \tfrac{\eta}{2}\|\boldsymbol{x}_i\|^2. \tag{25}$$

Combining (24) and (25), $\frac{1}{2\eta}\|\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star\|^2 \le \frac{\eta}{2}\|\boldsymbol{x}_i\|^2$, and taking square roots yields $\|\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star\| \le \eta\|\boldsymbol{x}_i\|$. □

We can now state our main result.

Theorem 9 (RESTATED)**.** *Let $\widehat{\boldsymbol{W}}$ be the regularized ERM with $\ell(\boldsymbol{W}, \boldsymbol{x}, y) = (1 - y\langle \boldsymbol{W}, \boldsymbol{x}\rangle)_+$, $y \in \{-1, +1\}$, and let y' use threshold $\beta(\boldsymbol{x}) = \eta\|\boldsymbol{x}\|^2$. With $\eta = 1/(4X^2)$, for every comparator $\boldsymbol{W}_\star \in \mathbb{R}^d$: $\mathbb{E}_{S,y'}\left[\mathbb{1}[y \ne y'(\boldsymbol{x}; \widehat{\boldsymbol{W}})]\right] \le \mathbb{E}_{(\boldsymbol{x},y)}[\ell(\boldsymbol{W}_\star, \boldsymbol{x}, y)] + \frac{2X^2\|\boldsymbol{W}_\star\|^2}{T+1}$.*

Proof of Theorem 9. Lemma D.1 gives the pointwise upper bound, provided $\eta \le 1/(4X^2)$:

$$\begin{aligned}\mathbb{E}_{y'}[\mathbb{1}[y'(\boldsymbol{x}_i; \widehat{\boldsymbol{W}}_i) \ne y_i]] &\le (1 - y\langle \widehat{\boldsymbol{W}}_i, \boldsymbol{x}\rangle - \eta\|\boldsymbol{x}_i\|^2)_+ \\ &\le (1 - y_i\langle \widehat{\boldsymbol{W}}_\star, \boldsymbol{x}_i\rangle)_+ \\ &= \ell_i(\widehat{\boldsymbol{W}}_\star).\end{aligned}$$

The second inequality applies Cauchy-Schwarz and Lemma D.2 to bound $y_i\langle \widehat{\boldsymbol{W}}_i, \boldsymbol{x}_i\rangle \ge y_i\langle \widehat{\boldsymbol{W}}_\star, \boldsymbol{x}_i\rangle - \eta\|\boldsymbol{x}_i\|^2$. The crucial observation made by Koren and Levy [2015] is that $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{W}}_{T+1}$ has the same distribution as $\widehat{\boldsymbol{W}}_i$ for all $i \in [T+1]$. Using this, and taking expectations yields:

$$\mathbb{E}_{S,y'}\big[\mathbb{1}[y'(\boldsymbol{x}; \widehat{\boldsymbol{W}}) \ne y]\big] = \frac{1}{T+1}\sum_{i=1}^{T+1}\mathbb{E}_{S,y}\big[\mathbb{1}[y'(\boldsymbol{x}_i; \widehat{\boldsymbol{W}}_i) \ne y_i]\big] \le \mathbb{E}_S\left[\frac{1}{T+1}\sum_{i=1}^{T+1}\ell(\widehat{\boldsymbol{W}}_\star, \boldsymbol{x}_i, y_i)\right].$$

By optimality of $\widehat{\boldsymbol{W}}_\star$ for ϕ,

$$\sum_{i=1}^{T+1} \ell_i(\widehat{\boldsymbol{W}}_\star) + \frac{1}{2\eta}\|\widehat{\boldsymbol{W}}_\star\|^2 \;\le\; \sum_{i=1}^{T+1} \ell_i(\boldsymbol{W}_\star) + \frac{1}{2\eta}\|\boldsymbol{W}_\star\|^2.$$

Dropping $\|\widehat{\boldsymbol{W}}_\star\|^2 \ge 0$, dividing by $T+1$, taking expectations, and using that $(\boldsymbol{x}_i, y_i)$ are i.i.d. yields:

$$\mathbb{E}_{S,y'}\big[\mathbb{1}[y'(\boldsymbol{x}; \widehat{\boldsymbol{W}}) \neq y]\big] \;\le\; \mathbb{E}_S[\ell(\boldsymbol{W}_\star, \boldsymbol{x}, y)] + \frac{\|\boldsymbol{W}_\star\|^2}{2\eta(T+1)}.$$

With $\eta = 1/(4X^2)$ this gives the rate $\frac{2X^2\|\boldsymbol{W}_\star\|^2}{T+1}$. □

D.2 Logistic Loss with Binary Specialization (Proof of Theorem 10)

The argument carries over to the logistic loss with only the two helper lemmas modified. The rest of the proof is structurally identical.

Remark D.4 (Binary specialization). For binary classification ($K = 2$), the loss (3) admits a different parametrization. Writing $\boldsymbol{W} = \boldsymbol{W}^{+1} - \boldsymbol{W}^{-1} \in \mathbb{R}^d$ and labels $y \in \{-1, +1\}$, one has

$$-\log_2 \sigma(\boldsymbol{W}, \boldsymbol{x}, y) \;=\; \log_2\big(1 + e^{-y\langle \boldsymbol{W}, \boldsymbol{x}\rangle}\big),$$

so the loss depends on $\boldsymbol{W}$ only through the scalar $y\langle \boldsymbol{W}, \boldsymbol{x}\rangle$. Throughout this subsection we work in this lower-dimensional parametrization, with $\boldsymbol{W} \in \mathbb{R}^d$, $\ell_t(\boldsymbol{W}) = \log_2(1 + e^{-y_t\langle \boldsymbol{W}, \boldsymbol{x}_t\rangle})$, and $\|\boldsymbol{W}\|$ the Euclidean norm. The matrix parametrization $\boldsymbol{W} \in \mathbb{R}^{K\times d}$ from (3) is recovered by setting $\boldsymbol{W}^{+1} = \frac{1}{2}\boldsymbol{W}$ and $\boldsymbol{W}^{-1} = -\frac{1}{2}\boldsymbol{W}$.

Throughout this section, set $\beta(\boldsymbol{x}) = \frac{\eta\|\boldsymbol{x}\|^2}{\log 2}$ and $\ell_t(\boldsymbol{W}) = \log_2(1 + e^{-y_t\langle \boldsymbol{W}, \boldsymbol{x}_t\rangle})$. We require the following two lemmas.

Lemma D.3 (Pointwise indicator-to-logistic)**.** *For any $\boldsymbol{W} \in \mathbb{R}^d$ and any $(\boldsymbol{x}, y) \in \mathbb{R}^d \times \{-1, +1\}$ the prediction y' defined relative to $\boldsymbol{W}$ satisfies*

$$\mathbb{E}_{y'}\big[\mathbb{1}[y \neq y'(\boldsymbol{x}; \boldsymbol{W})]\big] \;\le\; \log_2\Big(1 + e^{-y\langle \boldsymbol{W}, \boldsymbol{x}\rangle - \frac{\eta}{\log 2}\|\boldsymbol{x}\|^2}\Big), \tag{26}$$

provided that $\eta \le \frac{(\log 2)\log(\sqrt{2}+1)}{2X^2}$.

Proof of Lemma D.3. We write $\ell_+ \;=\; \log_2\Big(1 + e^{-y\langle \boldsymbol{W}, \boldsymbol{x}\rangle - \frac{\eta}{\log 2}\|\boldsymbol{x}\|^2}\Big)$ for the right-hand side of (26) and $p = \mathbb{E}_{y'}[\mathbb{1}[y \neq y'(\boldsymbol{x}; \boldsymbol{W})]]$ for the left-hand side. We show $p \le \ell_+$ by splitting into four cases based on the magnitude of $\langle \boldsymbol{W}, \boldsymbol{x}\rangle$ relative to the threshold $\frac{\eta}{\log 2}\|\boldsymbol{x}\|^2$ and on whether $\operatorname{sign}(\langle \boldsymbol{W}, \boldsymbol{x}\rangle)$ agrees with y.

$$\boxed{|\langle \boldsymbol{W}, \boldsymbol{x}\rangle| > \tfrac{\eta}{\log 2}\|\boldsymbol{x}\|^2 \text{ and } \operatorname{sign}(\langle \boldsymbol{W}, \boldsymbol{x}\rangle) = y}$$

The predictor outputs y deterministically, so $p = 0$. The bound $p \le \ell_+$ holds trivially since $\ell_+ \ge 0$.

$$\boxed{|\langle \boldsymbol{W}, \boldsymbol{x}\rangle| > \tfrac{\eta}{\log 2}\|\boldsymbol{x}\|^2 \text{ and } \operatorname{sign}(\langle \boldsymbol{W}, \boldsymbol{x}\rangle) \neq y}$$

The predictor outputs $-y$ deterministically, so $p = 1$. We have $y\langle \boldsymbol{W}, \boldsymbol{x}\rangle = -|\langle \boldsymbol{W}, \boldsymbol{x}\rangle| < -\frac{\eta}{\log 2}\|\boldsymbol{x}\|^2$, so $-y\langle \boldsymbol{W}, \boldsymbol{x}\rangle - \frac{\eta}{\log 2}\|\boldsymbol{x}\|^2 > 0$ and

$$\ell_+ \;\ge\; \log_2(1 + e^0) \;=\; 1.$$

$$\boxed{|\langle \boldsymbol{W}, \boldsymbol{x}\rangle| \le \tfrac{\eta}{\log 2}\|\boldsymbol{x}\|^2 \text{ and } \operatorname{sign}(\langle \boldsymbol{W}, \boldsymbol{x}\rangle) \neq y}$$

The predictor is uniform on $\{-1, +1\}$, so $p = 1/2$. We have $y\langle \boldsymbol{W}, \boldsymbol{x}\rangle \le 0$, so

$$\ell_+ \;\ge\; \log_2\Big(1 + e^{-\frac{\eta}{\log 2}\|\boldsymbol{x}\|^2}\Big) \;\ge\; \tfrac{1}{2}$$

provided $\frac{\eta}{\log 2}\|\boldsymbol{x}\|^2 \le \log(\sqrt{2}+1)$, which follows from $\eta \le \frac{(\log 2)\log(\sqrt{2}+1)}{2X^2}$.

$\boxed{|\langle \boldsymbol{W}, \boldsymbol{x}\rangle| \le \frac{\eta}{\log 2}\|\boldsymbol{x}\|^2 \text{ and } \operatorname{sign}(\langle \boldsymbol{W}, \boldsymbol{x}\rangle) = y}$

The predictor is uniform on $\{-1,+1\}$, so $p = 1/2$. We have $y\langle \boldsymbol{W}, \boldsymbol{x}\rangle = |\langle \boldsymbol{W}, \boldsymbol{x}\rangle| \le \frac{\eta}{\log 2}\|\boldsymbol{x}\|^2$, so

$$\ell_+ \;\ge\; \log_2\!\Big(1 + e^{-\frac{2\eta}{\log 2}\|\boldsymbol{x}\|^2}\Big) \;\ge\; \tfrac{1}{2}$$

provided $\frac{2\eta}{\log 2}\|\boldsymbol{x}\|^2 \le \log(\sqrt{2}+1)$, which follows from $\eta \le \frac{(\log 2)\log(\sqrt{2}+1)}{2X^2}$.

In all four cases $p \le \ell_+$. □

Lemma D.4 (Parameter stability)**.** *For each $i \in [T+1]$,*

$$\|\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star\| \;\le\; \frac{\eta\,\|\boldsymbol{x}_i\|}{\log 2}\,. \tag{27}$$

Proof of Lemma D.4. Since ϕ is $\frac{1}{\eta}$-strongly convex and $\widehat{\boldsymbol{W}}_\star$ minimises ϕ, strong convexity at the minimiser gives

$$\phi(\widehat{\boldsymbol{W}}_i) - \phi(\widehat{\boldsymbol{W}}_\star) \;\ge\; \tfrac{1}{2\eta}\|\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star\|^2. \tag{28}$$

For the upper bound, expand ϕ to second order around $\widehat{\boldsymbol{W}}_i$, with $\boldsymbol{z}$ on the segment between $\widehat{\boldsymbol{W}}_\star$ and $\widehat{\boldsymbol{W}}_i$:

$$\phi(\widehat{\boldsymbol{W}}_\star) = \phi(\widehat{\boldsymbol{W}}_i) - \nabla\phi(\widehat{\boldsymbol{W}}_i)^\top(\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star) + \tfrac{1}{2}(\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star)^\top\nabla^2\phi(\boldsymbol{z})(\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star).$$

Optimality of $\widehat{\boldsymbol{W}}_i$ for ϕ^{-i} yields $\nabla\phi(\widehat{\boldsymbol{W}}_i) = \nabla\ell_i(\widehat{\boldsymbol{W}}_i)$, and $\nabla^2\phi(\boldsymbol{z}) \succeq \frac{1}{\eta} I_d$ where I_d denotes the d-dimensional identity matrix, so

$$\phi(\widehat{\boldsymbol{W}}_i) - \phi(\widehat{\boldsymbol{W}}_\star) \;\le\; \nabla\ell_i(\widehat{\boldsymbol{W}}_i)^\top(\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star) - \tfrac{1}{2\eta}\|\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star\|^2.$$

Young's inequality, $a^\top b \le \frac{\eta}{2}\|a\|^2 + \frac{1}{2\eta}\|b\|^2$, applied to the linear term, together with $\|\nabla\ell_i(\widehat{\boldsymbol{W}}_i)\| \le \frac{\|\boldsymbol{x}_i\|}{\log 2}$ for the logistic loss, gives

$$\phi(\widehat{\boldsymbol{W}}_i) - \phi(\widehat{\boldsymbol{W}}_\star) \;\le\; \tfrac{\eta}{2(\log 2)^2}\|\boldsymbol{x}_i\|^2. \tag{29}$$

Combining (28) and (29), $\frac{1}{2\eta}\|\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star\|^2 \le \frac{\eta}{2(\log 2)^2}\|\boldsymbol{x}_i\|^2$, and taking square roots yields $\|\widehat{\boldsymbol{W}}_i - \widehat{\boldsymbol{W}}_\star\| \le \frac{\eta}{\log 2}\|\boldsymbol{x}_i\|$. □

We can now state our main theorem.

Theorem 10 (RESTATED)**.** *Let $\widehat{\boldsymbol{W}}$ be the regularized ERM with $\ell(\boldsymbol{W}, \boldsymbol{x}, y) = \log_2(1 + e^{-y\langle \boldsymbol{W}, \boldsymbol{x}\rangle})$, $\boldsymbol{W} \in \mathbb{R}^d$, $y \in \{-1,+1\}$. Let y' use threshold $\beta(\boldsymbol{x}) = \frac{\eta}{\log 2}\|\boldsymbol{x}\|^2$. With $\eta = \frac{(\log 2)\log(\sqrt{2}+1)}{2X^2}$, for every comparator $\boldsymbol{W}_\star \in \mathbb{R}^d$: $\mathbb{E}_{S,y'}\Big[\mathbb{1}[y \ne y'(\boldsymbol{x}; \widehat{\boldsymbol{W}})]\Big] \le \mathbb{E}_{(\boldsymbol{x},y)}[\ell(\boldsymbol{W}_\star, \boldsymbol{x}, y)] + \frac{X^2\|\boldsymbol{W}_\star\|^2}{(\log 2)\log(\sqrt{2}+1)\,(T+1)}$.*

Proof of Theorem 10. Lemma D.3 gives the pointwise upper bound, provided $\eta \le \frac{(\log 2)\log(\sqrt{2}+1)}{2X^2}$:

$$\begin{aligned}
\mathbb{E}_{y'}\big[\mathbb{1}[y'(\boldsymbol{x}_i; \widehat{\boldsymbol{W}}_i) \ne y_i]\big] &\le \log_2\!\Big(1 + e^{-y_i\langle \widehat{\boldsymbol{W}}_i, \boldsymbol{x}_i\rangle - \frac{\eta}{\log 2}\|\boldsymbol{x}_i\|^2}\Big) \\
&\le \log_2\!\Big(1 + e^{-y_i\langle \widehat{\boldsymbol{W}}_\star, \boldsymbol{x}_i\rangle}\Big) \\
&= \ell_i(\widehat{\boldsymbol{W}}_\star).
\end{aligned}$$

The second inequality applies Cauchy–Schwarz and Lemma D.4 to bound $-y_i\langle \widehat{\boldsymbol{W}}_i, \boldsymbol{x}_i\rangle \le -y_i\langle \widehat{\boldsymbol{W}}_\star, \boldsymbol{x}_i\rangle + \frac{\eta}{\log 2}\|\boldsymbol{x}_i\|^2$; the slack inside the exponent cancels this stability cost exactly, and

monotonicity of $z \mapsto \log_2(1+e^z)$ gives the bound. The crucial observation made by Koren and Levy [2015] is that $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{W}}_{T+1}$ has the same distribution as $\widehat{\boldsymbol{W}}_i$ for all $i \in [T+1]$. Using this and taking expectations yields:

$$\mathbb{E}_{S,y'}\left[\mathbb{1}[y'(\boldsymbol{x};\widehat{\boldsymbol{W}}) \neq y]\right] = \frac{1}{T+1}\sum_{i=1}^{T+1}\mathbb{E}_{S,y}\left[\mathbb{1}[y'(\boldsymbol{x}_i;\widehat{\boldsymbol{W}}_i) \neq y_i]\right] \leq \mathbb{E}_S\left[\frac{1}{T+1}\sum_{i=1}^{T+1}\ell_i(\widehat{\boldsymbol{W}}_\star)\right].$$

By optimality of $\widehat{\boldsymbol{W}}_\star$ for ϕ,

$$\sum_{i=1}^{T+1}\ell_i(\widehat{\boldsymbol{W}}_\star) + \frac{1}{2\eta}\|\widehat{\boldsymbol{W}}_\star\|^2 \;\leq\; \sum_{i=1}^{T+1}\ell_i(\boldsymbol{W}_\star) + \frac{1}{2\eta}\|\boldsymbol{W}_\star\|^2.$$

Dropping $\|\widehat{\boldsymbol{W}}_\star\|^2 \geq 0$, dividing by $T+1$, taking expectations, and using that $(\boldsymbol{x}_i, y_i)$ are i.i.d. yields:

$$\mathbb{E}_{S,y'}\left[\mathbb{1}[y'(\boldsymbol{x};\widehat{\boldsymbol{W}}) \neq y]\right] \;\leq\; \mathbb{E}_{(\boldsymbol{x},y)}[\ell(\boldsymbol{W}_\star,\boldsymbol{x},y)] + \frac{\|\boldsymbol{W}_\star\|^2}{2\eta(T+1)}.$$

With $\eta = \frac{(\log 2)\log(\sqrt{2}+1)}{2X^2}$ this gives the rate $\frac{X^2\|\boldsymbol{W}_\star\|^2}{(\log 2)\log(\sqrt{2}+1)\,(T+1)}$. □

E Details of Simulations (Section 6)

We run experiments on synthetic data to facilitate comparisons across feedback settings, noise levels, and values of K. Experiments were run on a single desktop (Intel Core i3-8350K @ 4.0 GHz) with Python 3.12.4. One experiment with 1 repetition for all depicted algorithms took roughly 65 seconds.

We follow the SynSep procedure of van der Hoeven et al. [2021] and Shibukawa et al. [2025]. Each input vector is a binary vector composed of two parts: a label-correlated *signal* and a label-independent *background*. To produce the signal, we generate K unique feature vectors of length $10n'$ by drawing an integer $s \in [n', 5n']$ uniformly and setting s entries to 1 in an otherwise zero vector. The input vector for round t is the signal vector of the round's label concatenated with a length-$30n'$ background vector that has $5n'$ entries set to 1 uniformly at random. The label is then replaced by a uniformly random label with probability r, introducing label noise.

To stress-test feature-norm robustness, we additionally apply a per-round multiplicative rescaling. Let $\rho \geq 0$ denote the *scale log-range* parameter. Each pre-generated feature vector $\bar{\boldsymbol{x}}_t$ is rescaled by an independent random factor

$$\sigma_t = 10^{e_t}, \qquad e_t \overset{\text{i.i.d.}}{\sim} \text{Uniform}(-\rho, \rho),$$

producing a stream whose feature norms span $10^{2\rho}$ orders of magnitude. The stream is then renormalized once, offline, by $\max_t \|\boldsymbol{x}_t\|$ so that $\|\boldsymbol{x}_t\| \leq 1$ for all t, matching the convention used by van der Hoeven et al. [2021] and Shibukawa et al. [2025]. This benchmarking convention does not give the algorithms online access to the future maximum; it only ensures that all algorithms see a stream consistent with the bound assumed by their published tunings (in X). We use $\rho = 1.5$ throughout the drift experiments, so feature norms span three orders of magnitude.

In the full-information setting without drift, `GAPTRON`-Hinge, `Stabletron`-Hinge, and `Stabletron`-Logistic incur the fewest mistakes, with `GAPTRON`-Hinge performing very similarly to `Stabletron`-Logistic. The mechanism behind `Stabletron`'s strong performance is visible in the per-round randomization rates (Figures 3 and 5): `Stabletron` randomizes only on rounds it deems unstable, while the remaining comparator algorithms (`GAPTRON`-Logistic, `Gappletron`, and `OSP`) randomize on a substantial fraction of rounds regardless of the algorithm's confidence. Moreover, even when `Stabletron` does randomize, it does so conservatively: `Stabletron`-Hinge uses the uniform distribution over only the *competitive* classes, and `Stabletron`-Logistic uses the softmax distribution over all classes, both of which place more mass on the argmax than the uniform-over-$[K]$ distributions used by `GAPTRON` and `Gappletron`. Among the remaining comparators, `OSP` fares best, since its softmax-based randomization concentrates as the iterates become confident, but its randomization probability $p_t = \min(1, 2\Delta^\star/\nu)$ does not vanish as the softmax concentrates. Although `GAPTRON`-Hinge matches the `Stabletron` variants in this setting, only `Stabletron` retains its performance under scale drift (see below). In the bandit setting without drift, `GAPTRON`-Hinge achieves the fewest

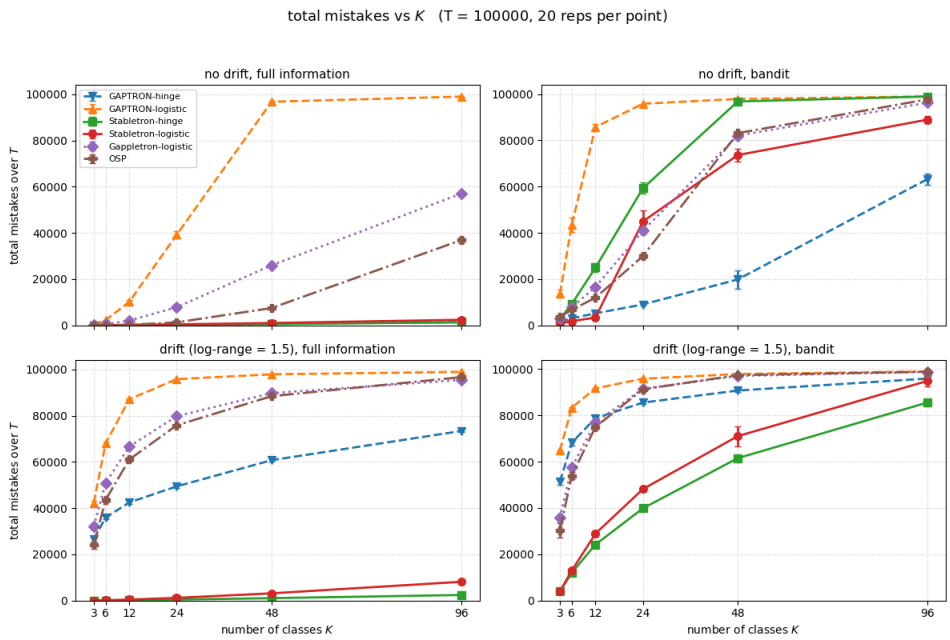


Figure 2: Number of mistakes for each algorithm under non-delayed feedback with label noise $r = 0$.

mistakes, with Stabletron-Logistic and Stabletron-Hinge close behind (Stabletron-Logistic slightly outperforming Stabletron-Hinge under both noise levels). The relative weakness of Stabletron-Hinge is explained by its randomization rate: its instability threshold is inflated in the bandit setting to

$$\beta_t = \frac{K}{\gamma}\,\eta\,\|\boldsymbol{x}_t\|^2,$$

and the tuning prescribed by Remark B.2 sets γ on the order of $1/\sqrt{T}$, which makes β_t grow with $\sqrt{T}$. The threshold therefore exceeds m_t^* on the vast majority of rounds, and the algorithm samples from the competitive set $\mathcal{U}_t$ rather than playing $y_t^\star$ deterministically. GAPTRON-Hinge avoids this issue by using a constant threshold of $1/K$ in both feedback settings, so it plays its argmax more often.

The same property that hurts Stabletron-Hinge in the bandit setting is what gives it strong robustness to scale drift: because its threshold $\beta_t = (K/\gamma)\,\eta\|\boldsymbol{x}_t\|^2$ scales with the feature norm, the ratio $m_t^\star/\beta_t$, and hence the stability decision, is invariant under any rescaling of $\boldsymbol{x}_t$. The comparator algorithms use absolute thresholds: GAPTRON-Hinge uses the constant $1/K$, while GAPTRON-Logistic and OSP use thresholds derived from softmax probabilities that scale slowly when $\|\boldsymbol{x}_t\|$ shrinks. Consequently, when we introduce scale drift ($\rho = 1.5$), the ranking shifts: Stabletron-Hinge incurs the fewest mistakes, with Stabletron-Logistic close behind, while the comparator algorithms deteriorate substantially. Their randomization rates rise toward 1 as feature norms shrink, since they cannot distinguish a confident round on a small-norm $\boldsymbol{x}_t$ from a genuinely uncertain one (Figures 3 and 5).

The delayed-feedback experiments amplify the patterns observed in the non-delayed setting (Figures 6, 7, 8, and 9). Each round the feedback is delayed with a geometric distribution with parameter indicated on the x-axis. Stabletron-Hinge randomizes on nearly every round across all delays, since the bandit-mode threshold inflation $\beta_t = (K/\gamma)\,\eta\,\|\boldsymbol{x}_t\|^2$ is further enlarged by the delay cross-term $2\eta\sum_{s\in n_t}|\langle \boldsymbol{x}_s, \boldsymbol{x}_t\rangle|$, pushing it well above the realized value of m_t^* on essentially every round. Consequently, its mistake count tracks the random-guessing baseline. Stabletron-Logistic plays its argmax more often, since its self-term carries only a single $1/\log 2$ factor against the cross-term's $1/(\log 2)^2$, and its softmax-based unstable-round distribution places more mass on the argmax than the uniform-over-$\mathcal{U}_t$ distribution used by hinge. As a result, Stabletron-Logistic incurs fewer mistakes than Stabletron-Hinge throughout the delay sweep. The drift comparison continues to favor Stabletron-Logistic over ODAFTRL-OSP: the latter's softmax-based randomization fails to concentrate when feature norms shrink, leaving its randomization probability $p_t = \min(1, 2\Delta^\star/\nu)$

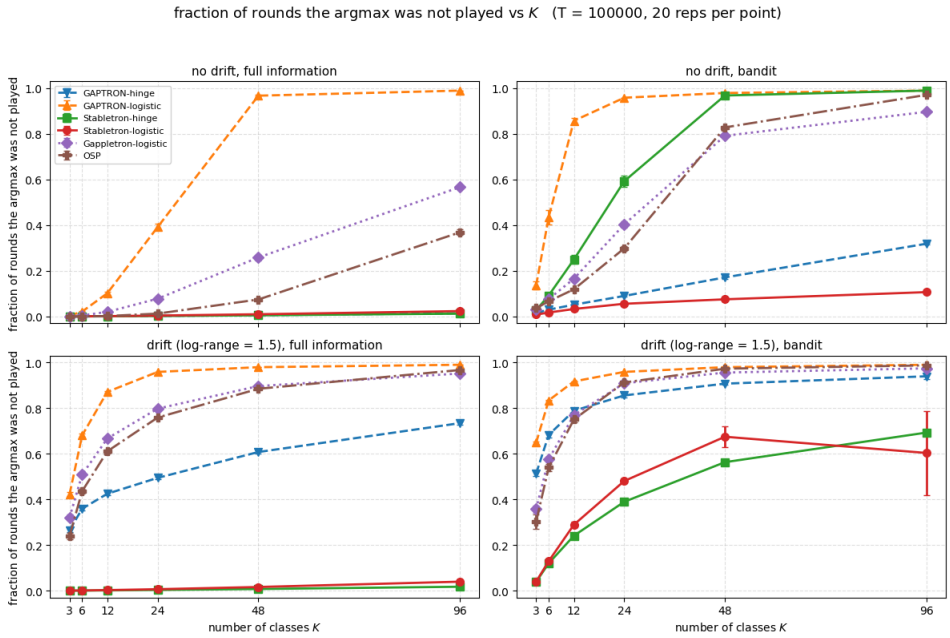


Figure 3: Frequency with which each algorithm did not play its argmax under non-delayed feedback with label noise $r = 0$.

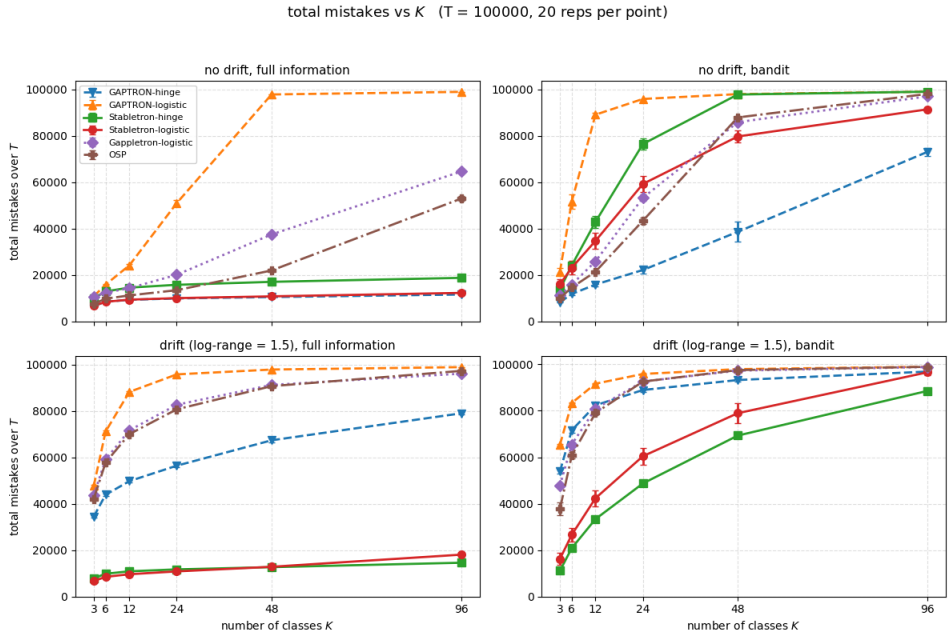


Figure 4: Number of mistakes for each algorithm under non-delayed feedback with label noise $r = 0.1$.

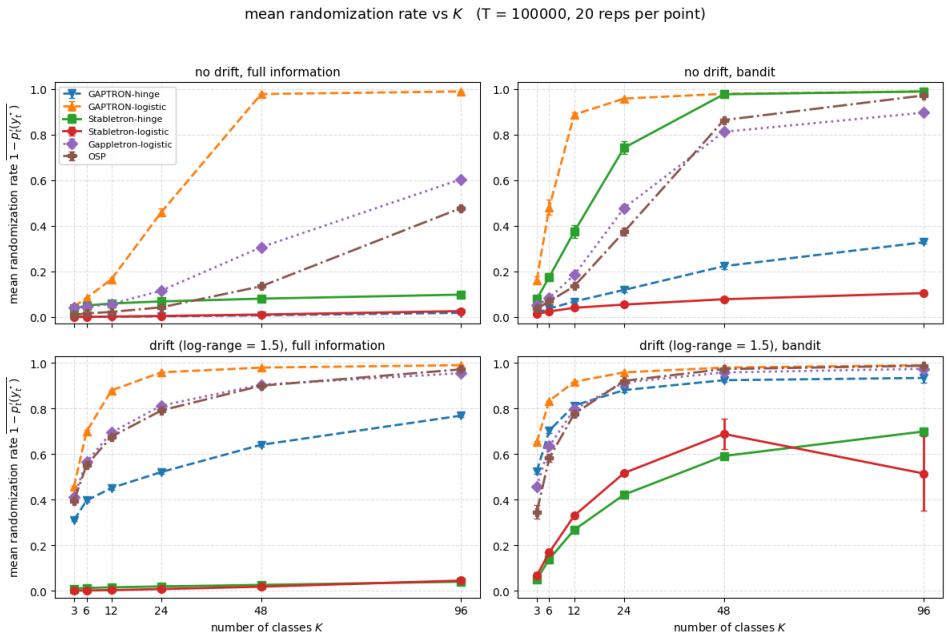


Figure 5: Frequency with which each algorithm did not play its argmax under non-delayed feedback with label noise $r = 0.1$.

near 1 throughout the run. All algorithms perform poorly under combined drift and delay, but `Stabletron`-Logistic does so with the smallest margin of degradation. The same pattern holds at both label-noise levels $r \in \{0, 0.1\}$.

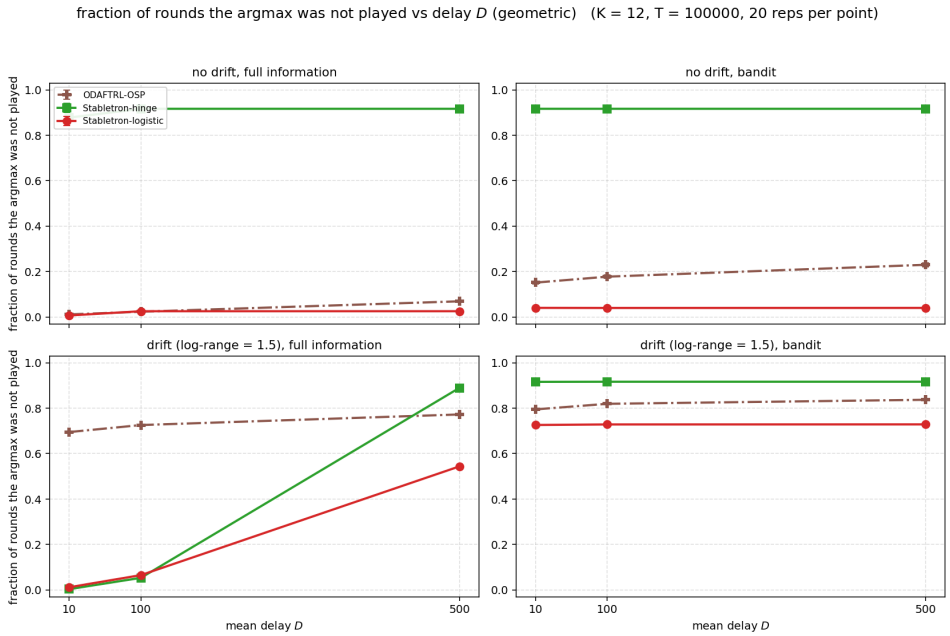


Figure 6: Frequency with which each algorithm did not play its argmax under delayed feedback with label noise $r = 0.0$.

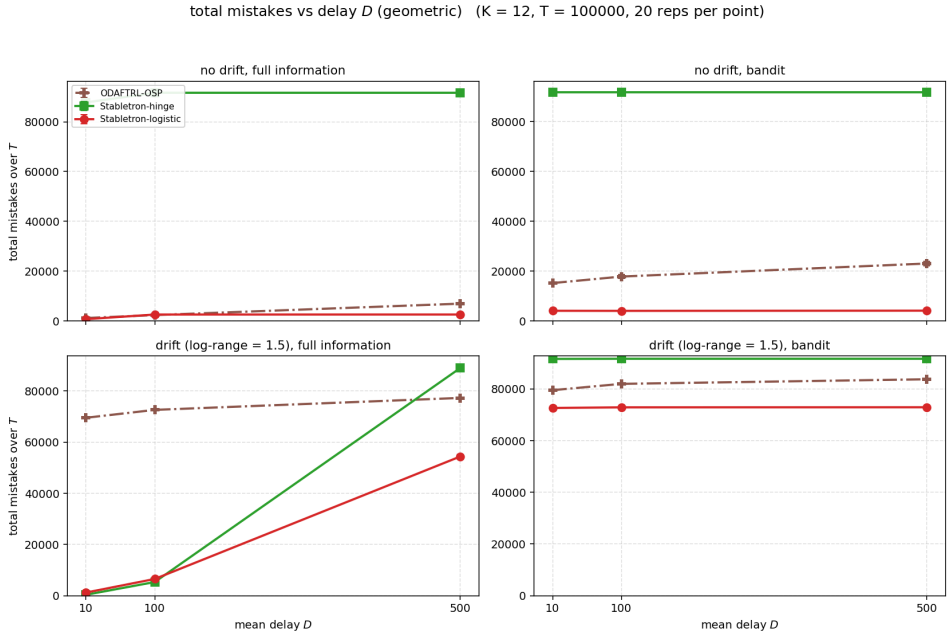


Figure 7: Number of mistakes for each algorithm under non-delayed feedback with label noise $r = 0.0$.

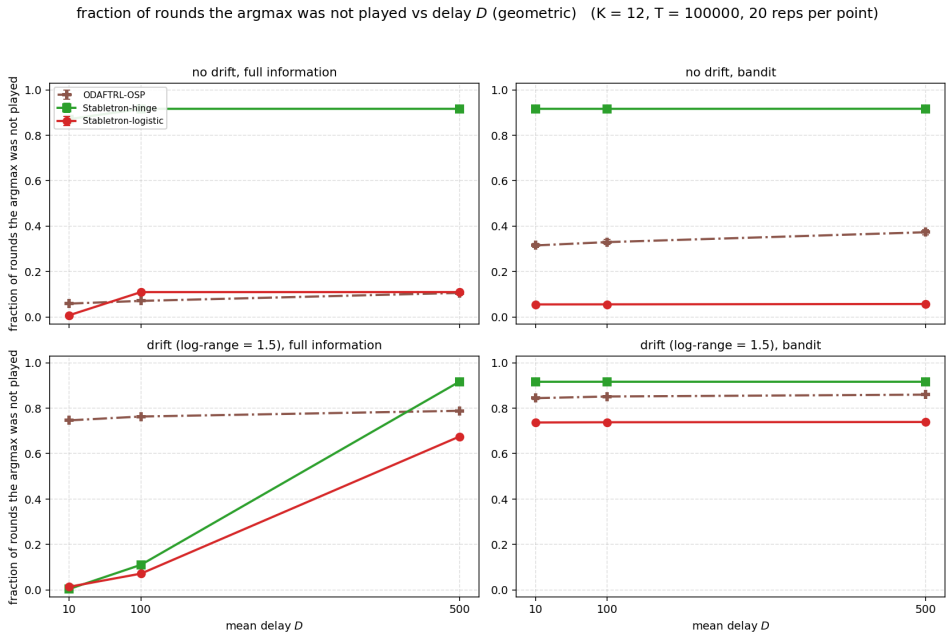


Figure 8: Frequency with which each algorithm did not play its argmax under delayed feedback with label noise $r = 0.1$.

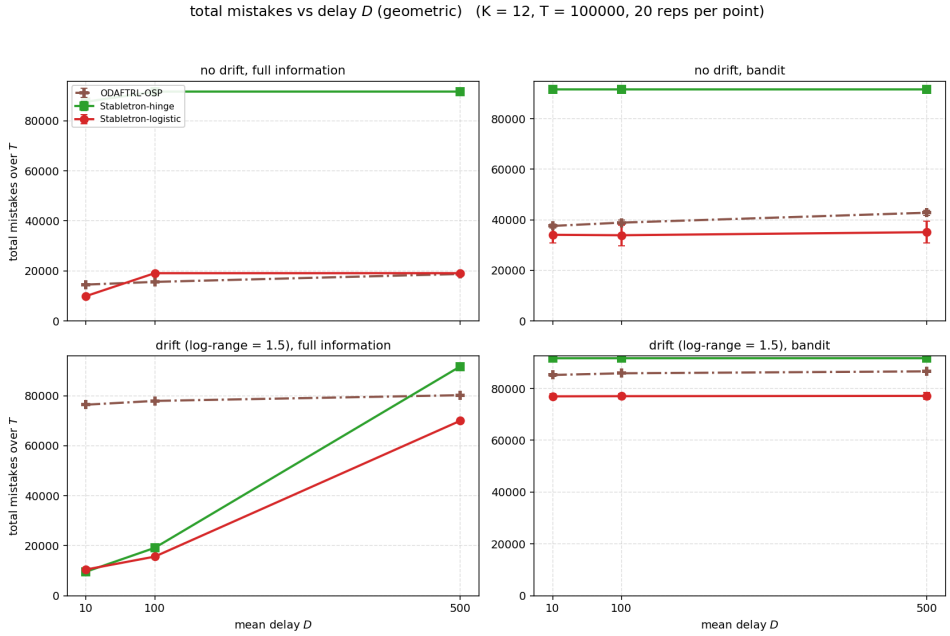


Figure 9: Number of mistakes for each algorithm under non-delayed feedback with label noise $r = 0.1$.

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: We have included references to theorem numbers for the results we claim to achieve in the introduction (the abstract is just a shorter version of the introduction). The proofs are included in the appendix, and throughout the paper the assumptions and limitations are expressed.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: In Section 4 we highlight that our approach works for online multiclass classification, whereas the work of Sakaue et al. [2024] and Shibukawa et al. [2025] applies to the more general online structured prediction setting. The empirical results 6 and Appendix E show that our algorithm for the multiclass hinge loss performs poorly in the delayed bandit setting. These limitations and their causes are then discussed in detail.

   Guidelines:

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

3. **Theory assumptions and proofs**

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

   Answer: [Yes]

   Justification: All theoretical results are stated with their full assumptions (data distribution, boundedness, choice of η in the main text. Proofs of all theorems and supporting lemmas are provided in the appendix, with cross-references in the main body of the text.

   Guidelines:

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

4. **Experimental result reproducibility**

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

   Answer: [Yes]

   Justification: All codes are provided as part of the supplementary material. Data generation happens within the code. Appendix E specifies how the data are generated, and the experiments that are performed.

   Guidelines:

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

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

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

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

   Answer: [Yes]

   Justification: Code is included in the supplementary material, along with a README specifying dependencies, environment setup and commands to reproduce the experimental results in this paper. The datasets are all synthetic and generated with fixed seeds (set in the code).

   Guidelines:

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

6. **Experimental setting/details**

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

   Answer: [Yes]

   Justification: Our experiments use synthetic data generated by procedures specified in Appendix E. Our setting is online learning, which doesn't involve train/test splits, however all algorithm (hyper)parameters are specified according to their theoretical tunings and available throughout the theoretical results and in the code itself.

   Guidelines:

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

7. **Experiment statistical significance**

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

   Answer: [Yes]

Justification: All experimental results are averaged over 20 repetitions and figures display error bars corresponding to +/-1 standard deviation across repetitions. They are however of a much smaller order than the horizon T of 10^5, and therefore not visible in all images. The sources of randomness are those of the data generation and the algorithms' internal randomization.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

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

   Answer: [Yes]

   Justification: As stated in Section 6: All experiments were run on a single desktop machine with an Intel Core i3-8350K CPU @ 4.0 GHz and 8 GB of RAM, using Python 3.12.4. No GPU is required. Each individual run (one algorithm, one configuration, $T = 10^5$ rounds) completes in under 2 minutes; the full experimental suite reported in Section 6 and Appendix E (all algorithms × all values of K × full-information and bandit settings × drift conditions × 20 repetitions) takes approximately 24 hours of time end-to-end. No additional preliminary or failed experiments beyond those reported contributed substantially to the compute budget.

   Guidelines:

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

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: The research is purely theoretical, supplemented by experiments on synthetic data.

Guidelines:

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

10. **Broader impacts**

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

    Answer: [N/A]

    Justification: This paper presents theoretical results on regret bounds for online classification, together with synthetic data experiments. The contributions are at a level of generality, and don't make specific societal impacts. No application or deployment is proposed or studied in this work.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

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

    Answer: [N/A]

    Justification: The paper releases no artifacts with high misuse risks. The contributions are theoretical, supplemented by code for reproducing synthetic-data experiments.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.

- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: Where we adapt or build on code from van der Hoeven et al. [2021] and Shibukawa et al. [2025] we cite the original authors in the experimental section and acknowledge the source in code comments (and README). All reused code is from publicly released artifacts accompanying published papers used in a manner consistent with their accompanying licenses.

    Guidelines:

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

13. **New assets**

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

    Answer: [Yes]

    Justification: Code accompanying the paper is included in the supplementary material with a README documenting dependencies, installation, structure of the codebase, and commands to reproduce the results in the paper. Each script includes a header describing its purpose and input/output.

    Guidelines:

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

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

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

    Answer: [N/A]

Justification: There is no crowdsourcing nor research with human subjects included in this work.

Guidelines:

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

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

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

Answer: [N/A]

Justification: The research involves no user studies or human subjects. Experiments are conducted on synthetic data.

Guidelines:

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

16. **Declaration of LLM usage**

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

Answer: [N/A]

Justification: LLMs were not used as a component of the methodology. Any use was limited to writing assistance (proofreading, formatting, polishing).

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.