
Online Continual Adaptation with Active Self-Training

Shiji Zhou
Tsinghua University

Han Zhao
University of Illinois at Urbana-Champaign

Shanghang Zhang*
UC Berkeley

Lianzhe Wang
Tsinghua University

Heng Chang
Tsinghua University

Zhi Wang
Tsinghua University

Wenwu Zhu*
Tsinghua University

Abstract

Models trained with offline data often suffer from continual distribution shifts and expensive labeling in changing environments. This calls for a new online learning paradigm where the learner can continually adapt to changing environments with limited labels. In this paper, we propose a new online setting – Online Active Continual Adaptation, where the learner aims to continually adapt to changing distributions using both unlabeled samples and active queries of limited labels. To this end, we propose Online Self-Adaptive Mirror Descent (OSAMD), which adopts an online teacher-student structure to enable online self-training from unlabeled data, and a margin-based criterion that decides whether to query the labels to track changing distributions. Theoretically, we show that, in the separable case, OSAMD has an $O(T^{2/3})$ dynamic regret bound under mild assumptions, which is aligned with the $\Omega(T^{2/3})$ lower bound of online learning algorithms with full labels. In the general case, we show a regret bound of $O(T^{2/3} + \alpha^*T)$, where α^* denotes the separability of domains and is usually small. Our theoretical results show that OSAMD can fast adapt to changing environments with active queries. Empirically, we demonstrate that OSAMD achieves favorable regrets under changing environments with limited labels on both simulated and real-world data, which corroborates our theoretical findings.

1 Introduction

Machine learning models, trained with data collected from closed environments, often suffer from continual distribution shift and expensive labeling in open environments. For example, a self-driving recognition system trained with data collected in the daytime may continually degrade when going towards nightfall (Bobu et al., 2018; Wu et al., 2019). The problem can be avoided by collecting and annotating sufficient training data to cover all the possible distributions at the test time. However, such data annotation is prohibitively expensive in many applications (Zhang et al., 2020a). In particular, for many scenarios, the distribution shifts constantly appear over time (Kumar et al., 2020), making it impossible to collect and annotate sufficient training data for a certain domain. This calls for a new online system that can continually adapt to the changing domain using limited labels.

The continual domain shift severely challenges the conventional domain adaptation methods (Tzeng et al., 2014; Ganin and Lempitsky, 2015; Hoffman et al., 2018), for most of them are designed to adapt to a fixed target domain (Su et al., 2020a; Prabhu et al., 2021) (Figure 1 bottom). Some previous works consider gradual domain shift (Bobu et al., 2018; Wu et al., 2019; Kumar et al., 2020), where the data distribution gradually evolves from batch to batch, but it is not realistic to model the continual shift that happens in continuous time. The adaptive online learning (Besbes et al., 2015) provides a classical theoretical framework to deal with changing environments. However, it requires the target data to be fully labeled (Figure 1 middle), which may be infeasible. Furthermore, it remains an open problem for online learning with limited labels (online active learning) under continual domain shift (Lu et al., 2016; Shuji, 2017). Recent work (Chen et al., 2021) studies online active domain adaptation for regression

Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151. Copyright 2022 by the author(s).

*Correspondence Authors: Shanghang Zhang and Wenwu Zhu

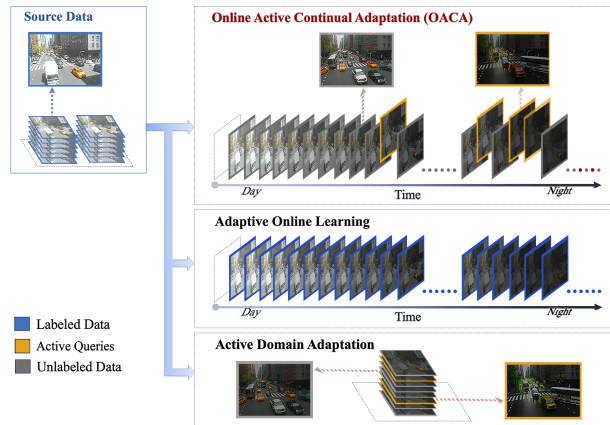


Figure 1: Illustration of OACA and previous works. The environments change continually from daytime to nightfall, introducing a shift of illumination and weather conditions. We model such a problem as an online active continual adaptation, where the online learner adapts to the environment with active queries of limited labels. In contrast, adaptive online learning requires full labels, and active domain adaptation can only adapt to a fixed domain.

problem under covariant (i.e. $P(X)$) shift, but can not deal with classification problem under joint distribution (i.e. $P(X, Y)$) shift, which is more general and realistic (Long et al., 2017, 2018).

To the best of our knowledge, no previous work has considered the online continual adaptation for classification with limited labels¹. To fill this gap, we formulate the Online Active Continual Adaptation (OACA) problem, where learners start with initial models and aim to minimize the dynamic regret caused by the distributional shifts using unlabeled data and active queries.

To resolve this problem, we propose the Online Self-Adaptive Mirror Descent (OSAMD) algorithm, which adopts an *online teacher-student structure* to enable the *self-adaptation* from unlabeled data: an “aggressive” model that updates actively using limited label queries with aggressive stepsizes, to track the max-margin classifier and provide accurate pseudolabels; a “conservative” model that adapts continually using the pseudolabels (taught by the aggressive model) with conservative stepsizes, to track the domain and minimize the dynamic regret. The active queries are given by a margin-based strategy that measures the confidence of the pseudolabels to query uncertain samples.

Theoretically, we show that OSAMD achieves an $O(T^{2/3})$ dynamic regret under mild assumptions in

¹Please refer to the detailed discussion in our literature view in Appendix. B

the separable case, which is aligned with the $\Omega(T^{2/3})$ lower bound of full-label online algorithms. We then extend this result to the general non-separable case, and derive a dynamic regret of $O(T^{2/3} + \alpha^*T)$, where α^* represents the separability of the data distribution. Since α^* is often a small constant by the representation ability of machine learning models (Kumar et al., 2020), the bias $O(\alpha^*T)$ is small such that the regret still approximates the lower bound. The above results lead to algorithmic insights that OSAMD is competitive with the optimal model in hindsight with fast convergence.

Empirically, we establish a simulation with changing environment to corroborate our theoretical findings. The results show that OSAMD performs accurately with limited labeled samples, and the regret aligns with Online Mirror Descent (OMD) with full labels. Furthermore, we extend OSAMD to deep learning on the real-world datasets Portraits (Ginosar et al., 2015) and Cover-Type (Blackard and Dean, 1999) to verify its practical effectiveness. OSAMD attains 93.7%(Portraits)/76.8%(Cover-Type) accuracy using only 3.8%/2.3% labels, comparing with 94.0%/76.8% accuracy of OMD with full labels. While OMD with uniform query and online active learning baseline only obtain 91.8%/75.3% and 92.0%/76.5% with 3.8%/2.3% queries, respectively. Finally, our ablation study shows that both the self-adaptation and active strategy contribute to the remarkable performance.

Our contributions

1. We are the first to formulate the problem of Online Active Continual Adaptation, which models the online continual domain adaptation with limited labels under realistic assumptions;
2. We propose an effective online self-adaptive algorithm named OSAMD with the novel design of the online teacher-student structure supported by strong theoretical guarantees;
3. We provide strict and novel dynamic regret analysis, highlighting the power of the proposed online teacher-student structure to get tight regret bound as in the regime of full-label online learning;
4. We demonstrate the effectiveness of our algorithm on both simulated and real-world datasets with changing environments, corroborating our theory.

2 Preliminaries

We first give a formal definition of the online convex optimization (OCO) with dynamic regret, and then briefly review the classic online mirror descent (OMD) algorithm. Next, we introduce the problem setting of domain adaptation (DA) and a statistical distance that will be used in our analysis.

2.1 Online Convex Optimization with Dynamic Regret

The basic protocol of Online Convex Optimization (OCO) (Hazan et al., 2016) is: at each time step $t = 1, \dots, T$, the online learner takes a decision w_t in a convex set \mathcal{K} . After that, the environment reveals a convex loss function $l_t : \mathcal{K} \rightarrow \mathbb{R}$, and the online learner suffers a loss $l_t(w_t)$. The dynamic regret is a theoretical metric for an online algorithm in changing environments, defined as

$$\text{D-Regret} := \sum_{t=1}^T l_t(w_t) - \sum_{t=1}^T l_t(w_t^*),$$

where $w_t^* = \arg \min_{w \in \mathcal{K}} l_t(w)$. It is well known that a sublinear dynamic regret bound is not possible unless specific constraints are made about the environments (Zinkevich, 2003). The first type of such constraint is the *path-length* (Zinkevich, 2003; Hall and Willett, 2013): $C_T := \sum_{t=1}^{T-1} \|w_t^* - w_{t+1}^*\|$, which measures how the optimal models change with the environment. Others include the *temporal variability* (Besbes et al., 2015; Campolongo and Orabona, 2020): $V_T := \sum_{t=1}^{T-1} \sup_{w \in \mathcal{K}} \|l_t(w) - l_{t+1}(w)\|$, which measures the variation of the loss functions.

Online Mirror Descent (OMD) (Hazan et al., 2016) is a general and classic algorithm, where the decision is updated by

$$w_{t+1} = \arg \min_{w \in \mathcal{K}} \eta \langle \nabla l_t(w_t), w \rangle + D_{\mathcal{R}}(w, w_t)$$

for $t = 1, \dots, T$, where η denotes the stepsize, and $D_{\mathcal{R}}(a, b) := \mathcal{R}(a) - \mathcal{R}(b) - \langle \nabla \mathcal{R}(b), a - b \rangle$ denotes the Bregman divergence with regularizer \mathcal{R} .

2.2 Domain Adaptation

Domain adaptation (DA) (Zhao et al., 2020) is a typical machine learning method to learn a model from a source domain \mathcal{P} that can perform well on a target domain \mathcal{Q} . Researches (Wei et al., 2020) often assume that the source domain and target domain are different but measured by some discrepancies. In this paper, we utilize the celebrated total variation (TV) distance d_{TV} to describe the similarity between two distributions over the same sample space:

Definition 1 (Total Variation). *We use $d_{TV}(\mathcal{P}, \mathcal{Q})$ to denote the total variation distance between distributions \mathcal{P} and \mathcal{Q} :*

$$d_{TV}(\mathcal{P}, \mathcal{Q}) := \sup_E |\mathcal{P}(E) - \mathcal{Q}(E)|,$$

where the supremum is over all the measurable events.

1. The data distribution begins with $P_1(X, Y)$.
2. The learner has enough data samples from $P_1(X, Y)$, and chooses an online algorithm \mathcal{A} .
3. The adversary chooses a sequence of data distribution $\{P_2, \dots, P_T\}$.
4. For each $t = 1, \dots, T$:
 - (a) The data (x_t, y_t) is sampled from joint distribution $P_t(X, Y)$.
 - (b) Instance x_t is revealed to the learner.
 - (c) The learner then chooses action w_t , incurring a loss on the domain $l_t(w_t) = \mathbb{E}_{(x,y) \sim P_t} f(w_t; x, y)$ in hindsight.
 - (d) The active agent decides whether to query the label. If query, true label y_t are revealed.

Figure 2: *Online Active Continual Adaptation setting.*

Several variants of the TV distance have been proposed and used in the domain adaptation (Ben-David et al., 2010; Zhao et al., 2018, 2019). It is also worth pointing out that other metrics can and have been used in the literature, e.g., the Wasserstein infinity (Kumar et al., 2020) distance and the maximum mean discrepancy (Long et al., 2015), as described in our appendix.

3 The Online Active Continual Adaptation Problem

In this section, we first formulate the online active continual adaptation (OACA) problem. We then formally introduce the assumptions used in our analysis and provide justifications for their necessities.

3.1 Problem Formulation

For the purpose of presentation, we consider an online binary classification task² of sequentially predicting labels $y_t \in \{1, -1\}$ from input features $x_t \in \mathcal{X}$ for round $t = 1, \dots, T$. In each round t , assume that our prediction model is parameterized by a vector $w_t \in \mathcal{K}$, and it outputs a soft label prediction³ over instance x_t denoted as $H_t(w_t) = H(w_t; x_t)$. The prediction result suffers a instantaneous loss as $f_t(w_t) = f(w_t; x_t, y_t)$, $x_t \in \mathcal{X}$, $y_t \in \{-1, 1\}$. We assume that H, f (unrelated with x, y) are known by the learner. We present the interaction between the learner and the environment as Figure 2. Specially, we assume each data sample (x_t, y_t) comes from a different distribution by the continual environmental change, which leads to different distributions at different times. To measure how the

²This can be readily extended to the multi-class case, as shown in Appendix. E

³The model output before the sign function, its absolute value is related to the distance to the decision boundary.

learner adapts to the environment, we use the theoretical metric of expected dynamic regret to measure the adaptation performance of the online learner:

$$\text{D-Regret}^{\mathcal{A}}(\{P_t\}, T) := \mathbb{E}^{\mathcal{A}}\left[\sum_{t=1}^T l_t(w_t)\right] - \sum_{t=1}^T l_t(w_t^*),$$

where $l_t(w_t) = \mathbb{E}_{(x,y) \sim P_t} f(w_t; x, y)$ is the expected loss, and the optimal action in hindsight is defined as $w_t^* = \arg \min_{w \in \mathcal{K}} l_t(w)$. The rest expectation $\mathbb{E}^{\mathcal{A}}$ is on the online decision w_t provided by a potentially randomized algorithm \mathcal{A} .

In particular, we here use the expected loss that reflects the performance on the distribution P_t not the instantaneous sample (x_t, y_t) . Intuitively, we study the continual domain adaptation problem, where the expected loss reflects the performance on the domain (environment), while the instantaneous loss only reflects the performance on individual samples. On the technical side, although the distributional change is continual, the change between consecutive samples could be large due to the randomness in sampling, leading to an unbounded dynamic regret (Besbes et al., 2015).

3.2 Assumptions

First, we assume the domain shift is continual and bounded.

Assumption 1 (Continual Domain Shift). *There exists a constant V_T , s.t. $\sum_{t=1}^{T-1} d_{TV}(P_t, P_{t+1}) \leq V_T$. In other words, the total domain shift is bounded.*

This assumption is closed related to the temporal variability constrain (Besbes et al., 2015), where we specify the expectation change as domain shift.

Next, we assume a niceness condition of the environments by its separability.

Assumption 2 (Separation). *For each time step $t \in [T]$, the data distribution P_t can be classified almost correctly with a margin R , i.e., there exists $v_t \in \mathcal{K}$ and a constant α^* such that $\mathbb{E}_{(x_t, y_t) \sim P_t} [\max\{0, R - y_t H_t(v_t)\}] \leq \alpha^*$. Furthermore, there exists a constant C_T such that $\sum_{t=1}^{T-1} \|v_t - v_{t+1}\| \leq C_T$, i.e., the classifiers with margin R change continually.*

Note that the coefficient α^* represents the optimal hinge loss for the classifier space. It is commonly assumed to be small by previous works (Kumar et al., 2020; Wei et al., 2020). The constraint of $\sum_{t=1}^{T-1} \|v_t - v_{t+1}\|$ is similar to the path-length regularity (Zinkevich, 2003) in online learning. Intuitively, v_t can be viewed as a max-margin classifier, and the continual rotation is bounded by C_T .

Finally, we present the following standard assumptions in online learning.

Assumption 3 (Convexity). *We assume that $f(\cdot)$, $-yH(\cdot)$ are all convex functions.*

Assumption 4 (Smoothness). *We assume that f and H are differentiable and G -Lipschitz, i.e. $\|\nabla f(w; x, y)\|_*$, $\|\nabla H(\theta; x)\|_* \leq G$, $\forall x, y, w$, where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. Furthermore, f is L -smooth, i.e. $\|\nabla f(w) - \nabla f(w')\| \leq L\|w - w'\|$.*

Assumption 5 (Bounded Decision Space and Function). *The diameter of decision space \mathcal{K} (convex set in \mathbb{R}^n) is bounded, i.e. there exists $D > 0$ such that $\max_{w, w' \in \mathcal{K}} \|w - w'\| \leq D$. The function value is bounded, i.e. there exists $F > 0$ such that $f(w; x, y) \leq F$, $\forall w, x, y$.*

4 The Online Self-Adaptive Mirror Descent Algorithm

Here we describe the proposed Online Self-Adaptive Mirror Descent (OSAMD) in Algorithm 1. To make our description easier to follow, we first introduce the following procedures.

1. **Pseudolabel:** Pseudolabel the example x_t by an aggressive model (parameterized by θ).
2. **Self-adaptation:** Before making the decision, the learner trusts the pseudolabel and self adapts the conservative model (parameterized by w, \hat{w}) by implicit mirror descent.
3. **Active query:** The active agent decides whether to query the label based on the margin measured by $|H_t(\theta)|$. If query, update the aggressive model by mirror descent with the true label by adaptive stepsizes. The conservative model is updated with pseudolabel or queried label by mirror descent.

Intuitive description At a high level, we design an “aggressive” model to track the max-margin classifier in order to produce correct pseudolabels. On the other hand, the conservative model is updated with pseudolabels with the goal of minimizing the dynamic regret. Intuitively, the aggressive model is updated with an aggressive stepsize, thus can track the max-margin classifier by limited updates with the true labels and provide trustful pseudolabels, although its regret might be large. The trustful pseudolabels enable the learner to “look ahead” with the incoming label and self-adapt the conservative model before making the final decision, leading to a lower regret. Our active query agent measures the uncertainty of pseudolabels by the margin between the data samples and the decision boundary, i.e., $|H_t(\theta)|$, and tends to query the uncertain samples. Then update the aggressive model with the active queries in time. Finally, the conservative model is updated with highly confident pseudolabels or queried real labels by a fixed conservative stepsize, and thus

Algorithm 1 Online Self-Adaptive Mirror Descent

Input: Active probability controller σ , aggressive step size τ_t , conservative step size η , initial data.
Initial: Learn from initial data, get aggressive model θ_1 and conservative model \hat{w}_1 .
for $t = 1, \dots, T$ **do**
 observe data sample x_t
 pseudolabel:
 give the pseudolabel provided by the aggressive model $\hat{y}_t = \text{sign}(H_t(\theta_t))$
 self-adaptation:
 adapt the conservative model $w_t = \arg \min_{w \in \mathcal{K}} \eta f(w; x_t, \hat{y}_t) + D_{\mathcal{R}}(w, \hat{w}_t)$, and then make the decision
 active query:
 draw a Bernoulli random variable with probability $Z_t \sim \text{Bernoulli}(\sigma/(\sigma + |H_t(\theta_t)|))$
 if $Z_t = 1$ **then**
 query label y_t , and let $\tilde{y}_t = y_t$
 update the aggressive model by $\theta_{t+1} = \arg \min_{\theta \in \mathcal{K}} -\tau_t \langle y_t \nabla H_t(\theta_t), \theta \rangle + D_{\mathcal{R}}(\theta, \theta_t)$
 else
 let $\theta_{t+1} = \theta_t$ and $\tilde{y}_t = \hat{y}_t$
 end if
 update the conservative model by $\hat{w}_{t+1} = \arg \min_{w \in \mathcal{K}} \eta \langle \nabla f(w_t; x_t, \tilde{y}_t), w \rangle + D_{\mathcal{R}}(w, \hat{w}_t)$
end for

keeps tracking the continual domain shift.

Online teacher-student structure Our novel design of running two models θ and w , where the aggressive model θ teaches the adaptation of the conservative model w , is motivated by the special property of continual domain shifts. Specifically, the max-margin classifier shifts continually and will not “cross over” the margin in a short time, thus the aggressive model does not need to update frequently, since the old model still can give correct pseudolabels. When the max-margin classifier “crosses over” the margin, the active agent will detect the uncertainty, then the aggressive model θ needs to track the max-margin classifier shift with an “aggressive” stepsize. On the other hand, as the continual domain shift leads to the continual change of the expected loss (on the distribution) l_t , the optimal minimizer w_t^* evolves continually, leading to the need for a conservative model w that frequently adapts to the shift using a “conservative” stepsize.

Remark Our pseudolabel and self-adaptation in the online setting, where we adopt a teacher-student structure, are different from the existing works in the offline setting (Zhao et al., 2020). Also, there is no previous online learning algorithm that uses the same ideas to the best of our knowledge. In fact, this is an open question for online active learning under non-stationary scenarios (Lu et al., 2016; Shuji, 2017).

5 Theoretical Analysis

In this section, we analyze the dynamic regret bound of the proposed algorithm. We first begin with the

separable case, and theoretically show the necessity of active query and the efficacy of the online teacher-student structure. Finally, we generalize the results to the non-separable case.

5.1 Separable Case

We begin with analyzing the regret bound in the separable case, i.e., $\alpha^* = 0$. Since the pseudolabel errors are highly related to the regret bound, we first present the following lemma:

Lemma 1 (Pseudolabel Errors). *Let the regularizer $\mathcal{R} : \mathcal{K} \mapsto \mathbb{R}$ be a 1-strongly convex function on \mathcal{K} with respect to a norm $\|\cdot\|$. Assume that $D_{\mathcal{R}}(\cdot, \cdot)$ satisfies $D_{\mathcal{R}}(x, z) - D_{\mathcal{R}}(y, z) \leq \gamma \|x - y\|, \forall x, y, z \in \mathcal{K}$. Set $\tau_t = \frac{\max\{0, \sigma - y_t H_t(\theta_t)\}}{\|\nabla H_t(\theta_t)\|_*^2}$, and $\sigma \leq R$. If $\alpha^* = 0$, the expected number of pseudolabel errors made by Algorithm 1 is bounded by*

$$\mathbb{E}[\sum_{t=1}^T M_t] \leq \frac{2G^2}{\sigma^2} (\gamma C_T + \epsilon_v),$$

where $M_t = \mathbf{1}_{\hat{y}_t \neq y_t}$ is the instantaneous mistake indicator, and $\epsilon_v = D_{\mathcal{R}}(\theta_1, v_1)$.

We provide detailed proof in the appendix, where we refer to and generalize the technique of online active learning (Cesa-Bianchi et al., 2006; Lu et al., 2016) in stationary settings with l_2 regularizer. Our results hold for non-stationary scenarios with any regularizers \mathcal{R} , which solves the open question proposed in (Lu et al., 2016; Shuji, 2017).

As illustrated in Lemma 1, we have three observations: 1) *Higher query rate leads to lower errors bound.* From

the upper bound, the expected mistakes bound is inversely proportional to query probability controller σ ; 2) The *pseudolabel mistakes are bounded by the classifier shift* C_T . This is aligned with the intuition that if the max-margin classifier shifts severely, then the pseudolabel agent is more likely to make mistakes; 3) *Better initialization implies fewer errors*. Better initialization leads to small ϵ_v , and thus implies a tighter upper bound, which shows the importance of the pre-trained model. As we assume both C_T and ϵ_v are constants, the expected errors are small and controllable.

It should be noted that $\mathbb{E}[f(w_t; x_t, y_t)|w_t] \neq l_t(w_t)$, since w_t depends on x_t , such that fixing w_t changes the distribution of (x_t, y_t) . This leads to biased gradients that complicate the regret analysis. Thus, before presenting the regret, we first measure the impact of the bias brought by this dependency.

Lemma 2. *For algorithm 1. We have the following inequality for $u_t \in \mathcal{K}, t = 1, \dots, T$*

$$\mathbb{E}[l_t(w_t) - l_t(u_t)] \leq \mathbb{E}[\langle \nabla f(w_t; x_t, y_t), w_t - u_t \rangle + \mathbb{E}[2(LD + G)\|w_t - \hat{w}_t\|].$$

Lemma 2 shows that the impact of the gradient bias is bounded by $\|w_t - \hat{w}_t\|$, which is controlled by the choice of the stepsize η . Now, we are ready to analyze the dynamic regret bound.

Theorem 1 (Dynamic Regret). *Under the same conditions and parameters in Lemma 1, Algorithm 1 achieves the following dynamic regret bound*

$$\begin{aligned} \text{D-Regret}^{OSAMD}(\{P_t\}, T) &\leq \frac{4(\eta G^4 + G^3 D)}{\sigma^2} (\gamma C_T + \epsilon_v) \\ &\quad + 2(LD + G)^2 \eta T + \frac{\epsilon_w + \gamma D}{\eta} + 4\sqrt{\frac{\gamma D T F V_T}{\eta}}, \end{aligned}$$

where $\epsilon_v = D_{\mathcal{R}}(\theta_1, v_1)$, $\epsilon_w = D_{\mathcal{R}}(\hat{w}_1, w_1^*)$.

Proof Sketch. By Lemma 2, the path-length version of instantaneous regret can be decomposed as

$$\begin{aligned} &\mathbb{E}[l_t(w_t) - l_t(u_t)] \\ &\leq \mathbb{E}[\langle \nabla f(w_t; x_t, y_t), w_t - u_t \rangle + 2(LD + G)\|w_t - \hat{w}_t\|] \\ &= \underbrace{\mathbb{E}[\langle \nabla f(w_t; x_t, y_t) - \nabla f(w_t; x_t, \hat{y}_t), w_t - \hat{w}_{t+1} \rangle]}_{\text{term A}} \\ &\quad + \underbrace{\mathbb{E}[\langle \nabla f(w_t; x_t, \hat{y}_t), w_t - \hat{w}_{t+1} \rangle]}_{\text{term B}} \\ &\quad + \underbrace{\mathbb{E}[\langle \nabla f(w_t; x_t, \hat{y}_t), \hat{w}_{t+1} - u_t \rangle]}_{\text{term C}} \\ &\quad + \underbrace{\mathbb{E}[\langle \nabla f(w_t; x_t, y_t) - \nabla f(w_t; x_t, \hat{y}_t), \hat{w}_{t+1} - u_t \rangle]}_{\text{term D}} \\ &\quad + \mathbb{E}[2(LD + G)\|w_t - \hat{w}_t\|], \end{aligned}$$

where $u_1, \dots, u_T \in \mathcal{K}$. Since the pseudolabel errors are bounded by Theorem 1, we know the bias $\|\hat{y}_t - y_t\| = 2M_t$. From the definition of algorithm, the bias $\|\tilde{y}_t - y_t\|$ is not larger than $\|\hat{y}_t - y_t\|$, we then have $\sum_{t=1}^T \|\tilde{y}_t - y_t\| \leq \sum_{t=1}^T \|\hat{y}_t - y_t\| \leq \sum_{t=1}^T 2M_t$. From this, as we assume the gradient and decision space are bounded, term A and term D can be bounded in terms of the errors bound, which is small and controllable by Lemma 1. Term C is bounded in terms of a recursive term by a proposition of mirror descent (Beck and Teboulle, 2003). We prove that the implicit gradient update rule has a similar property with explicit gradient mirror descent, then bound term B in terms of a recursive term. By adding all the terms up and setting a suitable stepsize, we obtain a sum of recursion and get a path-length version of dynamic regret bound. Finally, using the similar technique with (Besbes et al., 2015; Zhang et al., 2020b), we could transfer from path-length bound to temporal variability bound, i.e., bound the regret in terms of the continual domain shift. ■

We then carefully choose the parameters to obtain the detailed result.

Corollary 1. *In Theorem 1, set the stepsize $\eta = V_T^{1/3} T^{-1/3}$. We have for $\sigma \leq R$*

$$\text{D-Regret}^{OSAMD}(\{P_t\}, T) \leq O(\sigma^{-2} C_T + V_T^{1/3} T^{2/3}).$$

Corollary 1 implies that the dynamic regret of OSAMD is controlled by active probability controller σ , classifier shift C_T , and continual domain shift V_T . Since we assume C_T, V_T are constants, the regret is dominated by $O(T^{2/3})$, i.e., $\text{D-Regret}(\{P_t\}, T) \leq O(T^{2/3})$, leading to the insight that OSAMD can compete with the optimal competitor in hindsight with fast convergence.

Remark We do not provide the regret bound for $\sigma > R$. Since in the worst case (e.g., all the samples are on the margin, then $|H(v_t)| = R$), the algorithm would query at least half of the labels in expectation, which is contrary to the assumption of limited labels.

We then begin to discuss the advantage of our design from a theoretical review.

The necessity of query We have the following lower bound for any unsupervised (i.e., without query) online algorithms.

Theorem 2. *Assume $\alpha^* = 0$. For any unsupervised online algorithm \mathcal{A} , there exists a sequence of $\{P_t\}$ satisfying our assumptions such that*

$$\text{D-Regret}^{\mathcal{A}}(\{P_t\}, T) \geq \Omega(T).$$

We provide detailed proof in the appendix, where we create a special example. Theorem 2 implies that it is

impossible for unsupervised algorithms to keep adaptive if the environment is changing over time, which shows the necessity of active query.

The efficacy of the online teacher-student structure It should be noted that the online teacher-student structure leads to tight regret bound, which is even aligned with the lower bound for online algorithms with full labels.

Theorem 3. (Besbes et al. (2015)) Assume $\alpha^* = 0$. Even if we have all the labels during the process, for any online learning algorithm \mathcal{A} , there exists a sequence of $\{P_t\}$ satisfying our assumptions such that

$$\text{D-Regret}^{\mathcal{A}}(\{P_t\}, T) \geq \Omega(V_T^{1/3}T^{2/3}).$$

This result is immediate from Theorem 2 of (Besbes et al., 2015), where OMD with suitable stepsize attains the lower bound. As illustrated in Theorem 3, all algorithms suffer from an $\Omega(V_T^{1/3}T^{2/3})$ dynamic regret lower bound. Recall our result of OSAMD is upper bounded by $O(V_T^{1/3}T^{2/3})$. From this, we conclude that the online teacher-student structure is effective in attaining tight regret bound, which shows that OSAMD can adapt well to the changing environments with an optimal convergence.

Remark Note that the query strategy in Algorithm 1 is probabilistic, given by a Bernoulli distribution with probability $\frac{\sigma}{\sigma + |H_t(\theta_t)|}$, indicating that the query rate is relatively low when σ is small. Notice that we also provide a regret bound for $\sigma \leq R$ that can be arbitrarily small in Corollary 1, which means the tight regret bound holds even in the limited-labels case, showing the advantage of the online teacher-student structure compared with previous works with full labels.

5.2 General Case

In this subsection, we extend the results to the non-separable case, i.e., $\alpha^* > 0$ but is a small or negligible constant (typical assumption in previous works (Kumar et al., 2020; Wei et al., 2020)). We still begin with the analysis of the pseudolabel errors, as follows.

Lemma 3 (Pseudolabel Errors). Under the same conditions in Lemma 1. Set $\tau_t = \min\{\frac{\sigma}{G^2}, \frac{\max\{0, \sigma - y_t H_t(\theta_t)\}}{\|\nabla H_t(\theta_t)\|_2^2}\}$, $\sigma \leq R$. The expected number of pseudolabel errors made by Algorithm 1 is bounded by

$$\mathbb{E}\left[\sum_{t=1}^T M_t\right] \leq \frac{2G^2}{\sigma^2}(\gamma C_T + \epsilon_v + \frac{\sigma}{G^2}T\alpha^*).$$

where $M_t = \mathbf{1}_{\hat{y}_t \neq y_t}$ is the instantaneous mistake indicator, and $\epsilon_v = D_{\mathcal{R}}(\theta_1, v_1)$.

We provide detailed proof in the appendix, where we generalize the proof of the separable case. From Lemma 3, we observe that the expected pseudolabel errors are bounded by an $O(\alpha^*T)$ term, which is linear increasing. This cannot be eschewed, because any classifier would make mistakes if the data distribution is not separable. We then present the regret bound in such a case.

Theorem 4 (Regret Bound). Under the same conditions and parameters in Lemma 3. Algorithm 1 achieves the following regret bound

$$\begin{aligned} \text{D-Regret}^{OSAMD}(\{P_t\}, T) &\leq \frac{\epsilon_w + \gamma D}{\eta} + 4\sqrt{\frac{\gamma DTFV_T}{\eta}} + 2(LD + G)^2\eta T \\ &\quad + \frac{4(\eta G^4 + G^3 D)}{\sigma^2}(\gamma C_T + \epsilon_v + \frac{\sigma}{G^2}T\alpha^*), \end{aligned}$$

where $\epsilon_v = D_{\mathcal{R}}(\theta_1, v_1)$, $\epsilon_w = D_{\mathcal{R}}(\hat{w}_1, w_1^*)$.

We provide detailed proof in the appendix, where it is a simple generation of the separable case. We next choose the parameters to obtain the detailed results.

Corollary 2. In Theorem 4, set the parameter $\eta = V_T^{1/3}T^{-1/3}$. We have for $\sigma \leq R$

$$\begin{aligned} \text{D-Regret}^{OSAMD}(\{P_t\}, T) &\leq O(\sigma^{-2}C_T + V_T^{1/3}T^{2/3} + \sigma^{-1}\alpha^*T). \end{aligned}$$

Since C_T is assumed to be a constant. Therefore, from Corollary 2, we show the regret is of $O(V_T^{1/3}T^{2/3} + \alpha^*T)$, suggesting that the OSAMD algorithm is comparable to the optimal in hindsight with only $O(\alpha^*T)$ bias, which is the best we can hope to achieve. Besides, recall the regret lower bound of traditional algorithms is $\Omega(V_T^{1/3}T^{2/3})$. As α^* is often much small and negligible, the convergent rate is still aligned with the theoretical lower bound. Thus, we can claim that OSAMD can still adapt well to the changing environment.

Remark The α^* bias can not be eschewed for any self-adaptive algorithms with limited labeled data. For instance, if all the data are on the margin of v_t , then the mistake probability is at least $\alpha^*/2R$. Since labeled data is limited, we could assume that number of unlabeled data is of $\Omega(T)$. Then the pseudolabel errors are of $\Omega(\alpha^*T/2R)$, which leads to $\Omega(\alpha^*T/2R)$ mistake feedbacks. Therefore, it is easy to prove that the dynamic regret suffers from an $\Omega(\alpha^*T/2R)$ term, which is aligned with the term of $O(\sigma^{-1}\alpha^*T)$ in Corollary 2.

6 Experiments

In this section, we extensively evaluate OSAMD on both synthetic and real-world datasets. We first ver-

ify OSAMD on the synthesis dataset with continually changing distributions for the linear classification task. Then, we evaluate OSAMD on a deep learning model using a real-world dataset and demonstrate that the theoretical intuition can be applied to practical deep learning tasks as well.

6.1 Experimental Setup

We first briefly introduce our experimental setup, and leave the details in the appendix due to the space limit.

Dataset We experiment on two synthesis datasets - Rotating Gaussian (binary) & Rotating MNIST (multi-class), and two real-world datasets - Portraits (Ginosar et al., 2015) & Cover-Type (Blackard and Dean, 1999): 1) *Rotating Gaussian*: We sequentially sample the data from two continually changing Gaussian distributions representing two classes. The center points rotate from 0° to 180° . 2) *Rotating MNIST*: We averagely rotate the images from 0° to 90° counterclockwise to be the target dataset with a continually changing domain. 3) *Portraits*: It contains 37,921 photos of high school seniors labeled by gender. This real dataset suffers from a natural continual domain shift over the years (Kumar et al., 2020). 4) *Cover-Type*: It contains 495141 samples with 54 features labeled by cover types. This dataset suffers from a natural continual domain shift over the sample indexes (Kumar et al., 2020).

Baselines We compare with the following baselines: 1) *PAA* (Lu et al., 2016): To demonstrate the advantage of the online teacher-student structure, we compare with one online active algorithm – passive-aggressive active (PAA) learning; 2) *OMD (all)*: To compare OSAMD and traditional non-stationary online learning with full labels, we use all the labels to update by OMD; 3) *OMD (partial)*: To compare OSAMD and OMD with the same amount of labeled samples, we use uniform sampled labels to update by OMD; 4) *OSAMD w/o Self-adaptation*: To evaluate the self-adaptation method of OSAMD, we use the same active queries as OSAMD to update by OMD; 5) *OSAMD w/o Active-query*: To evaluate the active query strategy of OSAMD, we use uniform sampled labels to update the aggressive pseudolabel model for OSAMD.

We do not compare with the proposed methods in Kumar et al. (2020) and Chen et al. (2021), which study similar setting. For Kumar et al. (2020), their algorithm should learn from batch data, thus can not be directly applied in the online continual adaptation setting that we study. For Chen et al. (2021), the classification algorithm needs to memorize all the active queries during the online training and find the minimal of the received samples in every round. It is computationally

intractable in our deep learning experiments.

Implementation Details For Rotating Gaussian, we set the objective function to be the SVM loss. For other datasets, we follow the deep learning setting as previous work on unsupervised gradual domain adaptation (Kumar et al., 2020). For rotating MNIST and Portraits, we used a 3-layer convolutional network with dropout(0.5) and batchnorm on the last layer. For the Cover-Type dataset we used a 2 hidden layer feedforward neural network with dropout(0.5). Due to space limitations, please refer to the supplementary material for more details.

6.2 Experimental Results

We next present the experimental results and ablation study for the proposed method.

Synthesis data We first investigate whether the simulation can corroborate our theoretical findings, and present the results illustrated in Table 1 (Rotating Gaussian, Rotating MNIST) and Figure 3(a),3(b), from which we can make the following observations: *OSAMD performs remarkably well with limited labels*. There is no accuracy decrease from the full-label OMD with only 18.2% labels in Rotating Gaussian, and marginal decrease in Rotating MNIST with only 6.4% labels. In the above two datasets, OSAMD achieves similar regret/accumulated loss with full-label OMD, showing remarkable adaptation ability. In contrast, the regrets/accumulated losses of other baselines increase dramatically. This experimental result is aligned with our dynamic regret bound in Theorem 4, where OSAMD has similar dynamic regret to full-label OMD with only a small bias in the general case.

Real-world data We then extend OSAMD to work with deep learning models, and observe the performance in practice. As shown in Table 1 (Portraits, Cover-Type) and Figure 3(c),3(d), the practical results are similar to the synthesis data. OSAMD attains 93.7%(Portraits)/76.8%(Cover-Type) accuracy using only 3.8%/2.3% labels compared with 94.0%/76.8% accuracy of OMD (all) with full labels, while PAA and OMD (partial) with 3.8%/2.3% uniform queries only obtains 92.0%/76.5% and 91.8%/75.3%. The accumulated loss of OSAMD is aligned with OMD (full), being a side-information to reflect the consistent of the regret, which demonstrates the remarkable adaptation ability to real-world environments. While the accumulated losses of other baselines increase quickly, showing the practical advantage of our theoretical design.

Ablation study Note that we have two key designs on OSAMD, i.e., self-adaptation and active query. To

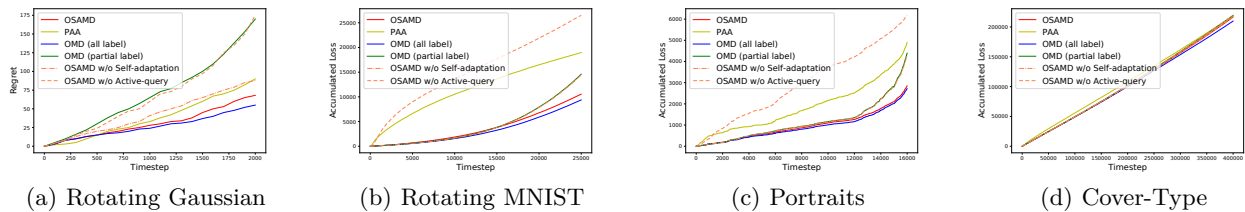


Figure 3: Regret/Accumulated loss v.s. timestep on different datasets.

Table 1: Classification accuracy with 90% confidence intervals for the mean over 10 runs.

	Rotating Gaussian		Rotating MNIST		Portraits		Cover-Type	
	Accuracy	Labels	Accuracy	Labels	Accuracy	Labels	Accuracy	Labels
OSAMD	98.9 ±0.2%	18.2±1.3%	86.8 ±0.8%	6.4±0.8%	93.7 ±0.3%	3.8±0.8%	76.8 ±0.07%	2.3±0.08%
PAA	98.5±0.2%	18.2±1.3%	84.0±0.6%	6.4±0.8%	92.0±1.0%	3.8±0.8%	76.5±0.02%	2.3±0.08%
OMD(all)	98.9 ±0.0%	100.0±0.0%	88.1 ±0.7%	100.0±0.0%	94.0 ±0.4%	100.0±0.0%	76.8 ±0.02%	100.0±0.0%
OMD(partial)	97.0±1.0%	18.2±1.3%	81.7±1.3%	6.4±0.8%	91.8±1.1%	3.8±0.8%	75.3±0.04%	2.3±0.08%
OSAMD w/o Self-adaption	98.2±0.3%	18.2±1.3%	81.7±1.2%	6.4±0.8%	91.8±1.1%	3.8±0.8%	75.3±0.03%	2.3±0.08%
OSAMD w/o Active-query	96.6±1.0%	18.2±1.3%	84.9±1.0%	6.4±0.8%	92.7±0.6%	3.8±0.8%	76.1±0.16%	2.3±0.08%

verify the efficacy of each component, we compare with two baselines: OSAMD w/o Self-adaptation and OSAMD w/o Active query. The experimental results in Table 1 and Figure 3 show that: 1) *The self-adaptation is effective*: OSAMD outperforms OSAMD w/o Self-adaptation, obtains a noticeable accuracy increase, and achieves a significant lower regret, which highlights the power of self-adaptation as in our theoretical findings. 2) *The active query is effective*: OSAMD is more accurate and achieves a lower regret than OSAMD w/o Active-query, which demonstrates that the active queries are more effective than uniform samples.

Sensitivity of Query Rate We test the performance regarding the effect of query number, as shown in Figure 4. We vary the parameter choice of σ that balances the number of queries and performance, and compare it with other baselines on the Rotating Gaussian dataset. From the result, we find that the proposed OSAMD consistently outperforms others under different query rates, suggesting the advantage of OSAMD is not sensitive to the query number.

7 Conclusions and Limitations

This paper studies an open problem for machine learning models to continually adapt to changing environments with limited labels, where previous works show limitations on realistic modeling and theoretical guarantees. To fill this gap, we formulate the OACA problem and propose OSAMD, an effective online active learning algorithm with the novel design of the online teacher-

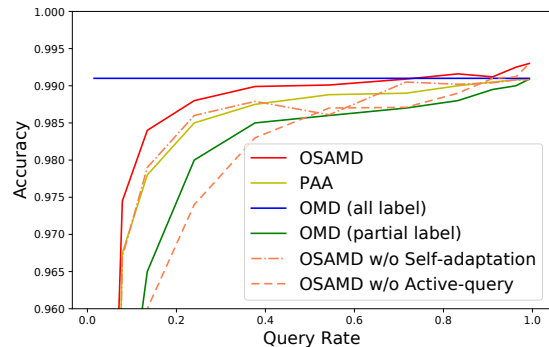


Figure 4: Accuracy v.s. query on Rotating Gaussian.

student structure. We show it can compete with the optimal model in hindsight with optimal convergence order. Experimental evaluations corroborate our theoretical findings and verify the efficacy of OSAMD. Our results take the first step towards online domain adaptation in continually changing environments.

Limitations In this work, we tradeoff the number of queries and regret in an implicit way by setting the parameter σ . When seeking the explicit way, we face the common challenge of estimating the expected queries in online active learning literature (Cesa-Bianchi et al., 2006; Lu et al., 2016), and this becomes even harder under non-stationary settings. Although the limitation does not affect our experimental results and practical usage, it remains a future theoretical interest.

Acknowledgements

This work is supported by the National Key Research and Development Program of China No. 2020AAA0106300, National Natural Science Foundation of China No. 62050110 and No. 61872215, and the Shenzhen Science and Technology Program of Grant No. RCYX20200714114523079. HZ would like to thank support from a Facebook research award. The authors also thank Kuaishou for sponsoring the research. We gratefully thank Tianyu Liu, Chenghao Hu, Jiangjing Yan, Yinan Mao for proof reading.

References

- Balcan, M.-F., Khodak, M., and Talwalkar, A. (2019). Provable guarantees for gradient-based meta-learning. In *International Conference on Machine Learning*, pages 424–433.
- Beck, A. and Teboulle, M. (2003). Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175.
- Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., and Vaughan, J. W. (2010). A theory of learning from different domains. *Machine learning*, 79(1):151–175.
- Besbes, O., Gur, Y., and Zeevi, A. (2015). Non-stationary stochastic optimization. *Operations research*, 63(5):1227–1244.
- Blackard, J. A. and Dean, D. J. (1999). Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables. *Computers and Electronics in Agriculture*, 24(3):131–151.
- Bobu, A., Tzeng, E., Hoffman, J., and Darrell, T. (2018). Adapting to continuously shifting domains. In *International Conference on Learning Representations Workshop*.
- Campolongo, N. and Orabona, F. (2020). Temporal variability in implicit online learning. *Advances in Neural Information Processing Systems*, 33:12377–12387.
- Cesa-Bianchi, N., Gentile, C., and Zaniboni, L. (2006). Worst-case analysis of selective sampling for linear classification. *Journal of Machine Learning Research*, 7(Jul):1205–1230.
- Cesa-Bianchi, N., Lugosi, G., and Stoltz, G. (2005). Minimizing regret with label efficient prediction. *IEEE Transactions on Information Theory*, 51(6):2152–2162.
- Chattopadhyay, R., Fan, W., Davidson, I., Panchanathan, S., and Ye, J. (2013). Joint transfer and batch-mode active learning. In *International Conference on Machine Learning*, pages 253–261.
- Chen, Y., Luo, H., Ma, T., and Zhang, C. (2021). Active online learning with hidden shifting domains. In *International Conference on Artificial Intelligence and Statistics*, pages 2053–2061.
- Finn, C., Rajeswaran, A., Kakade, S., and Levine, S. (2019). Online meta-learning. In *International Conference on Machine Learning*, pages 1920–1930.
- Gadermayr, M., Eschweiler, D., Klinkhammer, B. M., Boor, P., and Merhof, D. (2018). Gradual domain adaptation for segmenting whole slide images showing pathological variability. In *International Conference on Image and Signal Processing*, pages 461–469.
- Ganin, Y. and Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning*, pages 1180–1189.
- Ginosar, S., Rakelly, K., Sachs, S., Yin, B., and Efros, A. A. (2015). A century of portraits: A visual historical record of american high school yearbooks. In *International Conference on Computer Vision Workshops*, pages 1–7.
- Hall, E. and Willett, R. (2013). Dynamical models and tracking regret in online convex programming. In *International Conference on Machine Learning*, pages 579–587.
- Hao, S., Lu, J., Zhao, P., Zhang, C., Hoi, S. C., and Miao, C. (2017). Second-order online active learning and its applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(7):1338–1351.
- Hazan, E. et al. (2016). Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325.
- Hoffman, J., Darrell, T., and Saenko, K. (2014). Continuous manifold based adaptation for evolving visual domains. In *Conference on Computer Vision and Pattern Recognition*, pages 867–874.
- Hoffman, J., Tzeng, E., Park, T., Zhu, J.-Y., Isola, P., Saenko, K., Efros, A., and Darrell, T. (2018). Cycada: Cycle-consistent adversarial domain adaptation. In *International Conference on Machine Learning*, pages 1989–1998.
- Jadbabaie, A., Rakhlin, A., Shahrampour, S., and Sridharan, K. (2015). Online optimization: Competing with dynamic comparators. In *International Conference on Artificial Intelligence and Statistics*, pages 398–406.
- Khodak, M., Balcan, M.-F., and Talwalkar, A. (2019). Adaptive gradient-based meta-learning methods. In *Advances in Neural Information Processing Systems*, pages 5917–5928.

- Kumar, A., Ma, T., and Liang, P. (2020). Understanding self-training for gradual domain adaptation. In *International Conference on Machine Learning*, pages 5468–5479.
- Long, M., Cao, Y., Wang, J., and Jordan, M. (2015). Learning transferable features with deep adaptation networks. In *International Conference on Machine Learning*, pages 97–105.
- Long, M., Cao, Z., Wang, J., and Jordan, M. I. (2018). Conditional adversarial domain adaptation. In *Advances in Neural Information Processing Systems*, pages 1647–1657.
- Long, M., Zhu, H., Wang, J., and Jordan, M. I. (2017). Deep transfer learning with joint adaptation networks. In *International Conference on Machine Learning*, pages 2208–2217.
- Lu, J., Zhao, P., and Hoi, S. C. (2016). Online passive-aggressive active learning. *Machine Learning*, 103(2):141–183.
- Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2018). *Foundations of Machine Learning*. MIT press.
- Mokhtari, A., Shahrampour, S., Jadbabaie, A., and Ribeiro, A. (2016). Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *Conference on Decision and Control*, pages 7195–7201.
- Prabhu, V., Chandrasekaran, A., Saenko, K., and Hoffman, J. (2021). Active domain adaptation via clustering uncertainty-weighted embeddings. In *International Conference on Computer Vision*, pages 8505–8514.
- Rai, P., Saha, A., Daumé III, H., and Venkatasubramanian, S. (2010). Domain adaptation meets active learning. In *NAACL HLT Workshop on Active Learning for Natural Language Processing*, pages 27–32.
- Shuji, H. (2017). *Budget Efficient Online Active Learning and Its Applications*. PhD thesis, Nanyang Technological University.
- Su, J.-C., Tsai, Y.-H., Sohn, K., Liu, B., Maji, S., and Chandraker, M. (2020a). Active adversarial domain adaptation. In *Winter Conference on Applications of Computer Vision*.
- Su, J.-C., Tsai, Y.-H., Sohn, K., Liu, B., Maji, S., and Chandraker, M. (2020b). Active adversarial domain adaptation. In *Winter Conference on Applications of Computer Vision*, pages 739–748.
- Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., and Darrell, T. (2014). Deep domain confusion: Maximizing for domain invariance. *arXiv:1412.3474*.
- Wei, C., Shen, K., Chen, Y., and Ma, T. (2020). Theoretical analysis of self-training with deep networks on unlabeled data. In *International Conference on Learning Representations*.
- Wu, Z., Wang, X., Gonzalez, J. E., Goldstein, T., and Davis, L. S. (2019). Ace: Adapting to changing environments for semantic segmentation. In *International Conference on Computer Vision*, pages 2121–2130.
- Wulfmeier, M., Bewley, A., and Posner, I. (2018). Incremental adversarial domain adaptation for continually changing environments. In *International conference on robotics and automation*, pages 4489–4495.
- Yang, T., Zhang, L., Jin, R., and Yi, J. (2016). Tracking slowly moving clairvoyant: Optimal dynamic regret of online learning with true and noisy gradient. In *International Conference on Machine Learning*, pages 449–457.
- Zhang, Y., Wei, Y., Wu, Q., Zhao, P., Niu, S., Huang, J., and Tan, M. (2020a). Collaborative unsupervised domain adaptation for medical image diagnosis. *IEEE Transactions on Image Processing*, 29:7834–7844.
- Zhang, Y., Zhao, P., Cao, J., Ma, W., Huang, J., Wu, Q., and Tan, M. (2018). Online adaptive asymmetric active learning for budgeted imbalanced data. In *International Conference on Knowledge Discovery & Data Mining*, pages 2768–2777.
- Zhang, Y., Zhao, P., Niu, S., Wu, Q., Cao, J., Huang, J., and Tan, M. (2019). Online adaptive asymmetric active learning with limited budgets. *IEEE Transactions on Knowledge and Data Engineering*, 33(6):2680–2692.
- Zhang, Y.-J., Zhao, P., and Zhou, Z.-H. (2020b). A simple online algorithm for competing with dynamic comparators. In *Conference on Uncertainty in Artificial Intelligence*, pages 390–399.
- Zhao, H., Des Combes, R. T., Zhang, K., and Gordon, G. (2019). On learning invariant representations for domain adaptation. In *International Conference on Machine Learning*, pages 7523–7532.
- Zhao, H., Zhang, S., Wu, G., Costeira, J. P., Moura, J. M., and Gordon, G. J. (2018). Adversarial multiple source domain adaptation. In *Advances in Neural Information Processing Systems*, pages 8568–8579.
- Zhao, S., Yue, X., Zhang, S., Li, B., Zhao, H., Wu, B., Krishna, R., Gonzalez, J. E., Sangiovanni-Vincentelli, A. L., Seshia, S. A., et al. (2020). A review of single-source deep unsupervised visual domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*.
- Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning*, pages 928–936.

Supplementary Material: Online Continual Adaptation with Active Self-Training

A FORMATTING INSTRUCTIONS FOR THE SUPPLEMENTARY MATERIAL

We first provide a brief overview of the appendix. In Section B, we introduce previous works related to our literature. In Section C, we discuss how to connect various discrepancy measures between probability distributions in domain adaptation with the temporal variability condition in online learning, as described in Section 2 of the main paper. In Section D, we provide detailed proof for the analysis in Section 5 of the main paper. We then extend the results to the multiclass case in Section E. We provide the experimental details in Section F.

B Related Work

The topic of this paper sits well in between two amazing bodies of literature: domain adaptation (Tzeng et al., 2014; Ganin and Lempitsky, 2015; Hoffman et al., 2018; Zhao et al., 2020) that is a typical method to improve the generalization of a pre-trained model when testing on new domains without or with limited labels, and online learning (Hazan et al., 2016) that is a basic framework for learning with streaming online data. Our results therefore contribute to both fields and hopefully will inspire more interplay between the two communities.

B.1 Domain Adaptation

In the domain adaptation literature, our setting is related to active domain adaptation that queries additional labels to enable effective adaptation, and gradual domain adaptation that studies the adaptation problem under gradual domain shift. We present a brief summary as follows.

Active Domain Adaptation Active Domain Adaptation (Rai et al., 2010; Chattopadhyay et al., 2013; Su et al., 2020b; Prabhu et al., 2021) aims to actively select the most representative samples from the target domain, and learn a model to maximize performance on the target set. It was first proposed by Rai et al. (2010) with application to sentiment classification from text data, where they embedded an online uncertainty-based sample strategy in domain adaptation. Chattopadhyay et al. (2013) proposed a method that performs transfer and active learning simultaneously by solving a single convex optimization problem. Recently, active adversarial domain adaptation (AADA) (Su et al., 2020b) was proposed to solve the active domain adaptation problem in the context of deep learning, where AADA selects samples based on the uncertainty measured by entropy and targetness measured by the domain discriminator. Prabhu et al. (2021) proposed ADA-CLUE that queried labels based on uncertainty and diversity, then adopts a semi-supervised domain adaptation to transfer the domain knowledge to the target. However, current works are designed to adapt from a fixed source domain to a fixed target domain, and can not be applied to continual domain adaptation in the changing environment.

Gradual Domain Adaptation Gradual domain adaptation (Hoffman et al., 2014; Gadermayr et al., 2018; Wulfmeier et al., 2018; Bobu et al., 2018; Kumar et al., 2020) cares about how to adapt the model to a changing environment with unlabeled data. Continuous manifold learning (Hoffman et al., 2014) tried to adapt to evolving visual domains by learning a sequence of transformations on a fixed source representation. Gadermayr et al. (2018) extended previous approaches by adding two methods for regularization of the fully-unsupervised adaptation process. Wulfmeier et al. (2018) presented an adversarial approach benefiting from unsupervised alignment to a series of intermediate domains. Bobu et al. (2018) proposed a continuous replay model that enforced the past prediction to be matched. Kumar et al. (2020) first developed a theory, and proposed a gradual self-training method, which self-trains on the finite unlabeled examples from each batch successively. However, the generalization bound in (Kumar et al., 2020) suffers from an exponential error blow-up in time horizon T . Our

analysis further shows that unsupervised methods suffer from linear regret even in the separable case, implying the necessity of additional labels in the dynamic online setting.

B.2 Online Learning

In the online learning literature, our setting is related to adaptive online learning that aims to achieve optimal bound in dynamic environments, online active learning that studies online classification with active queries, online meta learning that provides a framework for adapting to a new domain with few-shot samples. We present a brief summary as follows.

Adaptive Online Learning Adaptive online learning (Besbes et al., 2015; Mokhtari et al., 2016; Jadbabaie et al., 2015) extends the traditional online learning setting to deal with dynamic problems, by introducing the dynamic regret that measures the online performance in dynamic environments. Under the path-length (Zinkevich, 2003; Hall and Willett, 2013) or temporal variability (Besbes et al., 2015; Campolongo and Orabona, 2020) conditions, sublinear regret is achieved by online algorithms with suitable stepsizes (Yang et al., 2016). However, practical deployments of fully online learning systems have been somewhat limited and impractical, partly due to the expense of annotations.

Online Active Learning Previous works (Cesa-Bianchi et al., 2005; Lu et al., 2016; Hao et al., 2017) study online active learning for classification. However, online classification with limited labels in changing environments remains an open question (Shuji, 2017). Recent work (Chen et al., 2021) considers online active learning with hidden covariate shift for regression tasks. However, both the algorithm and theory can not be generalized to online classification with joint distribution shift. In this paper, we tackle this problem and resolve the open question proposed in (Lu et al., 2016; Shuji, 2017).

Online Meta Learning Online meta learning (Finn et al., 2019; Balcan et al., 2019; Khodak et al., 2019) provides a framework for online few shot adaptation by learning the meta regularization. It studies how the model can fast adapt to a new environment using only a few samples by capturing the optimal initialization. However, online meta-learning focuses on “few-samples learning” using passively received labeled samples (usually not sufficient to achieve sublinear regret), while our setting focuses on “few-labels learning”, where the active queries and unlabeled samples also help the adaptation.

C Domain Discrepancy to Temporal Variability

In this section, we discuss how to connect classic distance metrics between probability distributions to the temporal variability condition used in the online learning literature. We present all the results in Table 2, where we provide the conditions for connecting these two.

Table 2: The conditions for connecting different domain discrepancy with the temporal variability condition

	Temporal Variability
Bounded sum of Total Variation	Bounded function f
Bounded sum of Wasserstein Infinity	f is Lischitz continuous on x ; No label shift
Bounded sum of Maximum Mean Discrepancy	Bounded reproducing kernel Hilbert space \mathcal{K} ; Linear function f

C.1 Bounded Sum of Total Variation

We first show that the bounded sum of total variation (Ben-David et al., 2010; Zhao et al., 2018, 2019) (as Assumption 1) with bounded function (as Assumption 5) can lead to the temporal variability condition in the online learning literature.

Proposition 1. *Assume the sum of total variation between P_t, P_{t+1} is bounded, i.e., satisfying Assumption 1. If the function value $f(w; x, y)$ is bounded for all $w \in \mathcal{K}, x \in \mathcal{X}, y \in \{-1, 1\}$, i.e., satisfying Assumption 5. Then the temporal variability is bounded as following*

$$\sum_{t=1}^{T-1} \sup_{w \in \mathcal{K}} |l_t(w) - l_{t+1}(w)| \leq 2FV_T.$$

Proof. First, by the definition of l_t and bounded f , we have

$$\begin{aligned} \sup_{w \in \mathcal{K}} |l_t(w) - l_{t+1}(w)| &= \sup_{w \in \mathcal{K}} \left| \mathbb{E}_{x,y \sim P_t(x,y)} [f(w; x, y)] - \mathbb{E}_{x,y \sim P_{t+1}(x,y)} [f(w; x, y)] \right| \\ &\leq \sup_{w \in \mathcal{K}} \int_{x,y} |f(w; x, y) dP_t - f(w; x, y) dP_{t+1}| \\ &\leq \sup_{w \in \mathcal{K}} \int_{x,y} |f(w; x, y)| |dP_t - dP_{t+1}| \\ &\leq F \int_{x,y} |dP_t - dP_{t+1}|. \end{aligned}$$

The last inequality is from Assumption 5. Then, sum this term from 1 to $T - 1$, and by the definition of d_{TV} , we obtain

$$\begin{aligned} \sum_{t=1}^{T-1} \sup_{w \in \mathcal{K}} |l_t(w) - l_{t+1}(w)| &\leq F \sum_{t=1}^{T-1} \int_{x,y} |dP_t - dP_{t+1}| \\ &\leq F \cdot 2 \sum_{t=1}^{T-1} \sup_E |P_t(E) - P_{t+1}(E)| \\ &= 2F \sum_{t=1}^{T-1} d_{TV}(P_t, P_{t+1}) \\ &\leq 2FV_T. \end{aligned}$$

The last inequality is from Assumption 1. We thus end the proof. \blacksquare

The Proposition 1 also holds for the multiclass case, since we define the total variation by measurable events, which do not depend on the class set.

C.2 Bounded Sum of Wasserstein Infinity Distance

We next present the bounded sum of wasserstein infinity distance that also leads to the temporal variability condition in the online learning literature, under conditions that f is Lischitz continuous over x and P_t has no label shift. We begin with the definition of wasserstein infinity distance.

Definition 2 (Wasserstein Infinity Distance). *We use $W_\infty(\mathcal{P}, \mathcal{Q})$ to denote the Wasserstein-infinity distance between distributions \mathcal{P} and \mathcal{Q} :*

$$W_\infty(\mathcal{P}, \mathcal{Q}) := \inf \left\{ \sup_{x \in \mathcal{X}} \|h(x) - x\| : h : \mathcal{X} \rightarrow \mathcal{X}, h_\# \mathcal{P} = \mathcal{Q} \right\},$$

where $\#$ denotes the push-forward of a measure, that is, for every set $A \subseteq \mathcal{X}$, $h_\# P(A) = P(h^{-1}(A))$.

Remark Note that in the definition above we use the Monge formulation of the Wasserstein distance. Under mild assumptions, e.g., both \mathcal{P} and \mathcal{Q} have densities, the Monge formulation is well-defined. This formulation has also been used in a previous work (Kumar et al., 2020) to measure the distributional shift.

In particular, the authors (Kumar et al., 2020) assume that the conditional distributions do not shift too much, i.e.,

$$\rho(\mathcal{P}, \mathcal{Q}) := \max(W_\infty(\mathcal{P}_{X|Y=1}, \mathcal{Q}_{X|Y=1}), W_\infty(\mathcal{P}_{X|Y=-1}, \mathcal{Q}_{X|Y=-1}))$$

is bounded, and there is no label shift, i.e., $\mathcal{P}(Y) = \mathcal{Q}(Y)$. We adopt similar assumptions and further assume that $f(w; x, y)$ is Lipschitz continuous on x , a general assumption on the loss function, which leads to temporal variability:

Proposition 2. *Let the function value $f(w; x, y)$ be Lipschitz continuous on x , i.e., there exists a constant $L \geq 0$, such that*

$$|f(w; x_1, y) - f(w; x_2, y)| \leq L \|x_1 - x_2\|, \forall x_1, x_2 \in \mathcal{X}, w \in \mathcal{K}.$$

Assume the sum of Wasserstein Infinity Distance between each consequent pair of conditional distribution is bounded, i.e.

$$\sum_{t=1}^{T-1} \rho(P_t, P_{t+1}) \leq V_T.$$

Further assume there is no label shift, i.e., $\forall t \in [T], P_t(Y) = P_{t+1}(Y) = P(Y)$. Then the temporal variability is bounded, as follows

$$\sum_{t=1}^{T-1} \sup_{w \in \mathcal{K}} |l_t(w) - l_{t+1}(w)| \leq LV_T.$$

Proof. First, by the definition of Wasserstein Infinity Distance, we know that there exist $h_t^{(y)}$, $y \in \{-1, 1\}$ such that

$$h_t^{(y)} \# P_t(\cdot | y) = P_{t+1}(\cdot | y)$$

and

$$\sup_{x \in \mathcal{X}} \left\| h_t^{(y)}(x) - x \right\| \leq \rho(P_t, P_{t+1}) + \epsilon, \quad \forall \epsilon > 0.$$

Then, by the definition of l_t , we have

$$\begin{aligned} |l_t(w) - l_{t+1}(w)| &\leq \sum_{y=-1,1} |P(Y=y) \mathbb{E}_{x \sim P_t(x|y)}[f(w; x, y)] - P(Y=y) \mathbb{E}_{x \sim P_{t+1}(x|y)}[f(w; x, y)]| \\ &= \sum_{y=-1,1} \left| \int_x f(w; x, y) P(y) dP_t(X|Y=y) - \int_x f(w; x, y) P(y) dP_{t+1}(X|Y=y) \right| \\ &= \sum_{y=-1,1} \left| \int_x f(w; x, y) P(y) dP_t(X|Y=y) - \int_x f(w; h_t^{(y)}(x), y) P(y) dP_t(X|Y=y) \right| \\ &\leq \sum_{y=-1,1} \int_x |f(w; x, y) P(y) - f(w; h_t^{(y)}(x), y) P(y)| dP_t(X|Y=y) \\ &\leq L \int_x \|x - h_t^{(1)}(x)\| P(Y=1) dP_t(X|Y=1) + L \int_x \|x - h_t^{(-1)}(x)\| P(Y=-1) dP_t(X|Y=-1) \\ &\leq L \max_{h=h_t^{(1)}, h_t^{(-1)}} \sup_{x \in \mathcal{X}} \|h(x) - x\|_2 \left(\int_x P(Y=1) dP_t(X|Y=1) + \int_x P(Y=-1) dP_t(X|Y=-1) \right) \\ &\leq L(\rho(P_t, P_{t+1}) + \epsilon), \forall \epsilon > 0. \end{aligned}$$

There is no w in right side, and thus the inequality still hold if we take \sup_w in the left side. Then

$$\sup_{w \in \mathcal{K}} |l_t(w) - l_{t+1}(w)| \leq \inf_{\epsilon > 0} L(\rho(P_t, P_{t+1}) + \epsilon) = L\rho(P_t, P_{t+1}).$$

By summing up, we get

$$\sum_{t=1}^{T-1} \sup_{w \in \mathcal{K}} |l_t(w) - l_{t+1}(w)| \leq L \sum_{t=1}^{T-1} \rho(P_t, P_{t+1}) \leq LV_T. \quad \blacksquare$$

C.3 Bounded Sum of Maximum Mean Discrepancy

We finally show that under the conditions that the decision space \mathcal{K} is a bounded reproducing kernel Hilbert space and f is linear on the representation space, the bounded sum of maximum mean discrepancy (Long et al., 2015) can lead to the temporal variability condition in the online learning.

Definition 3 (Maximum Mean Discrepancy). We use $MMD(\mathcal{P}, \mathcal{Q})$ to denote the maximum mean discrepancy between distributions \mathcal{P} and \mathcal{Q} :

$$MMD_\phi(\mathcal{P}, \mathcal{Q}) := \|\mathbb{E}_{x \sim \mathcal{P}}[\phi(x)] - \mathbb{E}_{x \sim \mathcal{Q}}[\phi(x)]\|_{\mathcal{H}},$$

where feature map $\phi : \mathcal{X} \rightarrow \mathcal{H}$, and \mathcal{H} is a reproducing kernel Hilbert space. In binary class, the distance between conditional distribution

$$d_{MMD}^\phi(\mathcal{P}, \mathcal{Q}) := \max\{MMD_\phi(\mathcal{P}_{X|Y=1}, \mathcal{Q}_{X|Y=1}), MMD_\phi(\mathcal{P}_{X|Y=-1}, \mathcal{Q}_{X|Y=-1})\}.$$

Proposition 3. Let \mathcal{K} to be a reproducing kernel Hilbert space. Assume the sum of Maximum Mean Discrepancy between conditional P_t, P_{t+1} is bounded, i.e.

$$\sum_{t=1}^{T-1} d_{MMD}^\phi(P_t, P_{t+1}) \leq V_T,$$

where $\phi : \mathcal{X} \rightarrow \mathcal{K}$. Let $f(w; x, y) = y\langle w, \phi(x) \rangle$ linear on the representation space. Assume \mathcal{K} is bounded by $\|w\|_{\mathcal{H}} \leq F, \forall w \in \mathcal{K}$, then the temporal variability is bounded, as following

$$\sum_{t=1}^{T-1} \sup_{w \in \mathcal{K}} |l_t(w) - l_{t+1}(w)| \leq FV_T.$$

Proof. From the linear property of f , and the definition of l_t , we have

$$l_t(w) = \sum_{y=-1,1} P(Y=y) \mathbb{E}_{x \sim P_t(x|y)} f(w; x, y) = \sum_{y=-1,1} P(Y=y) y \langle w, \mathbb{E}_{x \sim P_t(x|y)} \phi(x) \rangle.$$

Then, by the definition of Maximum Mean Discrepancy

$$\begin{aligned} & \sup_{w \in \mathcal{K}} |l_t(w) - l_{t+1}(w)| \\ &= \sup_{w \in \mathcal{K}} \left| \sum_{y=-1,1} P(Y=y) y \langle w, \mathbb{E}_{x \sim P_t(x|y)} \phi(x) \rangle - \sum_{y=-1,1} P(Y=y) y \langle w, \mathbb{E}_{x \sim P_{t+1}(x|y)} \phi(x) \rangle \right| \\ &\leq \sup_{w \in \mathcal{K}} \sum_{y=-1,1} P(Y=y) \|w\|_{\mathcal{H}} \|\mathbb{E}_{x \sim P_t(x|y)} \phi(x) - \mathbb{E}_{x \sim P_{t+1}(x|y)} \phi(x)\|_{\mathcal{H}} \\ &\leq F d_{MMD}^\phi(P_t, P_{t+1}). \end{aligned}$$

The first inequality comes from the Hölder inequality. By summing up, we finally get

$$\sum_{t=1}^{T-1} \sup_{w \in \mathcal{K}} |l_t(w) - l_{t+1}(w)| \leq \sum_{t=1}^{T-1} F d_{MMD}^\phi(P_t, P_{t+1}) \leq FV_T. \quad \blacksquare$$

D Missing Proofs

In this section, we provide the detailed proof of the pseudolabel errors bound and the dynamic regret bound for OSAMD.

D.1 Pseudolabel Errors

In this subsection, we analyze the pseudolabel errors for the OSAMD algorithm. We will first present some useful lemmas, then provide the proof of the separable case, where the data distribution can be correctly classified within a margin (i.e., $\alpha^* = 0$ in Assumption 2). Finally, we generalize it to the non-separable case. Here we generalize the proof in Lu et al. (2016); Cesa-Bianchi et al. (2006) to non-stationary cases with mirror descent.

We first introduce the lemma on the property of Bregman divergence.

Lemma 4 (Beck and Teboulle (2003)). *Let \mathcal{K} be a convex set in a Banach space \mathcal{B} , and regularizer $\mathcal{R} : \mathcal{K} \mapsto \mathbb{R}$ be a convex function, and let $D_{\mathcal{R}}(\cdot, \cdot)$ be the Bregman divergence induced by \mathcal{R} . Then, any update of the form*

$$w^* = \arg \min_{w \in \mathcal{K}} \{\langle a, w \rangle + D_{\mathcal{R}}(w, c)\}$$

satisfies the following inequality

$$\langle w^* - d, a \rangle \leq D_{\mathcal{R}}(d, c) - D_{\mathcal{R}}(d, w^*) - D_{\mathcal{R}}(w^*, c)$$

for any $d \in \mathcal{K}$.

Denote the instantaneous hinge loss with margin r by $f_t^r(\theta) = \max\{0, r - y_t H(\theta; x_t)\}$, where x_t, y_t is sampled from P_t . We then present a useful lemma to get the recurrence.

Lemma 5. *When regularizer $\mathcal{R} : \mathcal{K} \mapsto \mathbb{R}$ is a 1-strongly convex function on \mathcal{K} with respect to a norm $\|\cdot\|$. Then for algorithm 1, the following inequality holds*

$$\tau_t r - \tau_t y_t H_t(\theta_t) - \frac{\tau_t^2}{2} \|\nabla H_t(\theta_t)\|_*^2 \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t).$$

for $r > 0$.

Proof. First, by the definition of f_t^r , we have

$$\begin{aligned} r - f_t^r(v_t) &= r - \max\{0, r - y_t H_t(v_t)\} \leq y_t H_t(v_t) \\ &= y_t H_t(v_t) - y_t H_t(\theta_t) + y_t H_t(\theta_t). \end{aligned}$$

By the convexity of $-yH(\cdot)$, we have

$$\begin{aligned} y_t H_t(v_t) - y_t H_t(\theta_t) &= -y_t H_t(\theta_t) - (-y_t H_t(v_t)) \\ &\leq \langle -y_t \nabla H_t(\theta_t), \theta_t - v_t \rangle \\ &\leq \langle -y_t \nabla H_t(\theta_t), \theta_{t+1} - v_t \rangle + \langle -y_t \nabla H_t(\theta_t), \theta_t - \theta_{t+1} \rangle. \end{aligned}$$

By the update rule of θ and Lemma 4, the first term can be bounded by

$$\langle -y_t \nabla H_t(\theta_t), \theta_{t+1} - v_t \rangle \leq \frac{1}{\tau_t} (D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) - D_{\mathcal{R}}(\theta_{t+1}, \theta_t)).$$

Due to Hölder inequality and the fact that $ab \leq \frac{\eta}{2} a^2 + \frac{1}{2\eta} G^2$ for $\eta > 0$, for the second term, we have

$$\begin{aligned} \langle -y_t \nabla H_t(\theta_t), \theta_t - \theta_{t+1} \rangle &\leq \|\nabla H_t(\theta_t)\|_* \|\theta_{t+1} - \theta_t\| \\ &\leq \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2 + \frac{1}{2\tau_t} \|\theta_{t+1} - \theta_t\|^2. \end{aligned}$$

Due to the strong convexity of regularizer \mathcal{R} , we have $D_{\mathcal{R}}(x, y) \geq \frac{1}{2} \|x - y\|^2$ for any $x, y \in \mathcal{X}$ (Mohri et al., 2018). Therefore, by plugging the above term, we obtain that

$$\begin{aligned} r - f_t^r(v_t) &\leq \frac{1}{\tau_t} (D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) - \frac{1}{2} \|\theta_{t+1} - \theta_t\|^2) \\ &\quad + \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2 + \frac{1}{2\tau_t} \|\theta_{t+1} - \theta_t\|^2 + y_t H_t(\theta_t). \end{aligned}$$

By rearranging, we have

$$\tau_t r - \tau_t y_t H_t(\theta_t) - \frac{\tau_t^2}{2} \|\nabla H_t(\theta_t)\|_*^2 \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t). \quad \blacksquare$$

Denote the instantaneous mistake by $M_t(w) = \mathbf{1}_{\hat{y}_t \neq y_t}$, and let $L_t(w) = \mathbf{1}_{\hat{y}_t = y_t, \hat{y}_t H_t(w) \leq r}$ to be the indicator of the right decision but in the margin r , where $\mathbf{1}_{(\cdot)}$ is the indicator function. We then have the following relationship.

Lemma 6. *Take the same assumptions as Lemma 5. For Algorithm 1, let $\tau_t = 0$ if $f_t^r(\theta_t) = 0$, then the following inequality holds for every t*

$$\begin{aligned} & M_t Z_t \tau_t (r + |H_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2) + L_t Z_t \tau_t (r - |H_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2) \\ & \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t) \end{aligned}$$

for $r > 0$.

Proof. From Lemma 5, we know that

$$\tau_t r - \tau_t y_t H_t(\theta_t) - \frac{\tau_t^2}{2} \|\nabla H_t(\theta_t)\|_*^2 \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t).$$

If $M_t = 1$ then $y_t H_t(\theta_t) \leq 0$, and if $L_t = 1$ then $y_t H_t(\theta_t) \geq 0$. Therefore, we can obtain

$$\begin{aligned} & M_t Z_t \tau_t (r + |H_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2) + L_t Z_t \tau_t (r - |H_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2) \\ & \leq M_t Z_t (D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t)) + L_t Z_t (D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t)) \\ & = (M_t + L_t) Z_t (D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t)). \end{aligned}$$

From Algorithm 1, we know that if $Z_t = 0$, then $\tau_t = 0, \theta_t = \theta_{t+1}$. And if $M_t + L_t = 0$, we get $y_t H_t(\theta_t) \geq r$, then $\tau_t = 0, \theta_t = \theta_{t+1}$. Therefore, we have

$$(M_t + L_t) Z_t (D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t)) = D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t).$$

We finally get

$$\begin{aligned} & M_t Z_t \tau_t (r + |H_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2) + L_t Z_t \tau_t (r - |H_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2) \\ & \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t). \end{aligned}$$

■

D.1.1 Separable Case

Here, we analyze pseudolabel errors for the separable case, i.e., $\alpha^* = 0$, where we can easily know that $f_t^r(v_t) = 0$ if $r \leq R$. Before proving the theorem, we first present the following lemma.

Lemma 7. *Take the same assumptions as Lemma 1. Let $\tau_t = f_t^r(\theta_t) / \|\nabla H_t(\theta_t)\|_*^2$. Then for Algorithm 1, the following inequality holds*

$$\frac{r}{2G^2} M_t Z_t (r + |H_t(\theta_t)|) \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}),$$

for $r \leq R$.

Proof. By the separability, we know $f_t^r(v_t) = 0, r \leq R$. According to Lemma 6, we have

$$\begin{aligned} & M_t Z_t \tau_t (r + |H_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2) + L_t Z_t \tau_t (r - |H_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2) \\ & \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}). \end{aligned}$$

By taking $\tau_t = f_t^r(\theta_t) / \|\nabla H_t(\theta_t)\|_*^2$, we can obtain

$$\begin{aligned} & M_t Z_t \tau_t (r + |H_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2) + L_t Z_t \tau_t (r - |H_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2) \\ & = M_t Z_t \tau_t (r + |H_t(\theta_t)| - \frac{1}{2} (r + |H_t(\theta_t)|)) + L_t Z_t \tau_t (r - |H_t(\theta_t)| - \frac{1}{2} (r - |H_t(\theta_t)|)) \\ & = \frac{1}{2} M_t Z_t \tau_t (r + |H_t(\theta_t)|) + \frac{1}{2} L_t Z_t (r - |H_t(\theta_t)|) \\ & \geq \frac{1}{2} M_t Z_t \tau_t (r + |H_t(\theta_t)|). \end{aligned}$$

The last inequality comes from the definition of L_t . Therefore

$$\frac{1}{2}M_t Z_t \tau_t (r + |H_t(\theta_t)|) \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}).$$

From Assumption 4, we know that

$$M_t \tau_t = M_t \frac{f_t^r(\theta_t)}{\|\nabla H_t(\theta_t)\|_*^2} \geq M_t \frac{f_t^r(\theta_t)}{G^2} \geq M_t \frac{r}{G^2},$$

we thus have

$$\frac{r}{2G^2}M_t Z_t (r + |H_t(\theta_t)|) \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}).$$

■

With the above Lemmas, we are now ready to proof the Lemma 1.

Proof of Lemma 1. First, by the condition $D_{\mathcal{R}}(x, z) - D_{\mathcal{R}}(y, z) \leq \gamma \|x - y\|, \forall x, y, z \in \mathcal{K}$, we have

$$\begin{aligned} \sum_{t=1}^T D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) &\leq D_{\mathcal{R}}(v_1, \theta_1) + \sum_{t=1}^{T-1} (D_{\mathcal{R}}(v_{t+1}, \theta_{t+1}) - D_{\mathcal{R}}(v_t, \theta_{t+1})) \\ &= \epsilon_v + \gamma \sum_{t=1}^{T-1} \|v_{t+1} - v_t\| \\ &= \epsilon_v + \gamma C_T \end{aligned}$$

Then, by the definition of OSAMD algorithm and Lemma 7, we have

$$\begin{aligned} \mathbb{E}[\sum_{t=1}^T M_t] &= \frac{1}{r} \mathbb{E}[\sum_{t=1}^T M_t Z_t (r + |H_t(\theta_t)|)] \\ &= \frac{2G^2}{r^2} \mathbb{E}[\sum_{t=1}^T \frac{r}{2G^2} M_t Z_t (r + |H_t(\theta_t)|)] \\ &\leq \frac{2G^2}{r^2} \mathbb{E}[\sum_{t=1}^T D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1})] \\ &\leq \frac{2G^2}{r^2} (\epsilon_v + \gamma C_T) \\ &= \frac{2G^2}{\sigma^2} (\epsilon_v + \gamma C_T), \end{aligned}$$

where $r = \sigma$. We thus end the proof. ■

D.1.2 General Case

Here, we provide the analysis for the pseudolabel errors of the general case, where we do not assume that the data distribution P_t is 100% separated within a margin. We first present the following lemma.

Lemma 8. *Take the same assumptions as Lemma 1. Then for Algorithm 1, let $\tau_t = \min\{C, \frac{f_t^r(\theta_t)}{\|\nabla H_t(\theta_t)\|_*^2}\}$, the following inequality holds*

$$\min\{C, \frac{r}{G^2}\} \frac{1}{2} M_t Z_t (r + |H_t(\theta_t)|) \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + C f_t^r(v_t),$$

for $r \leq R$.

Proof. First, according to Lemma 6, we have

$$\begin{aligned} & M_t Z_t \tau_t (r + |H_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2) + L_t Z_t \tau_t (r - |H_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla H_t(\theta_t)\|_*^2) \\ & \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t). \end{aligned}$$

Since we take

$$\tau_t = \min\left\{C, \frac{f_t^r(\theta_t)}{\|\nabla H_t(\theta_t)\|_*^2}\right\} \leq f_t^r(\theta_t) / \|\nabla H_t(\theta_t)\|_*^2.$$

Similar to Lemma 7, we have

$$\frac{\tau_t}{2} M_t Z_t (r + |H_t(\theta_t)|) \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t).$$

Since we know that

$$M_t \tau_t = M_t \min\left\{C, \frac{f_t^r(\theta_t)}{\|\nabla H_t(\theta_t)\|_*^2}\right\} \leq M_t \min\left\{C, \frac{r}{G^2}\right\}.$$

Further, by $\tau_t \leq C$. We therefore have

$$\min\left\{C, \frac{r}{G^2}\right\} \frac{1}{2} M_t Z_t (r + |H_t(\theta_t)|) \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + C f_t^r(v_t).$$

■

We are now ready to prove the Lemma 3.

Proof of Lemma 3. First by the proof of Lemma 1, we have

$$\sum_{t=1}^T D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) \leq \epsilon_v + \gamma C_T.$$

Then

$$\begin{aligned} \mathbb{E}\left[\sum_{t=1}^T M_t\right] &= \frac{1}{r} \mathbb{E}\left[\sum_{t=1}^T M_t Z_t (r + |H_t(\theta_t)|)\right] \\ &= \frac{2}{r^2} \max\left\{\frac{r}{C}, G^2\right\} \mathbb{E}\left[\sum_{t=1}^T \min\left\{C, \frac{r}{G^2}\right\} \frac{1}{2} M_t Z_t (r + |H_t(\theta_t)|)\right] \\ &\leq \frac{2}{r^2} \max\left\{\frac{r}{C}, G^2\right\} \mathbb{E}\left[\sum_{t=1}^T D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \sum_{t=1}^T C f_t^r(v_t)\right] \\ &\leq \frac{2}{r^2} \max\left\{\frac{r}{C}, G^2\right\} (\epsilon_v + \gamma C_T + C \sum_{t=1}^T l_t^r(v_t)) \\ &\leq \frac{2}{r^2} \max\left\{\frac{r}{C}, G^2\right\} (\epsilon_v + \gamma C_T + C T \alpha^*) \\ &= \frac{2G^2}{\sigma^2} (\epsilon_v + \gamma C_T + \frac{\sigma}{G^2} T \alpha^*), \end{aligned}$$

where $r = \sigma$ and $C = \sigma/G^2$. The second inequality comes from $l_t^r(v_t) = \mathbb{E}[f_t^r(v_t)]$, and the last inequality comes from $l_t^r(v_t) \leq l_t^R(v_t) \leq \alpha^*$. We thus end the proof. ■

D.2 Dynamic Regret Bound

In this subsection, we begin to bound the dynamic regret. We will first provide necessary lemmas, and then use these lemmas to give the final proof.

We here give similar result as Lemma 4 for property of the implicit gradient mirror descent.

Lemma 9. Let \mathcal{K} be a convex set in a Banach space \mathcal{B} , and regularizer $\mathcal{R} : \mathcal{K} \mapsto \mathbb{R}$ be a convex function, and let $D_{\mathcal{R}}(\cdot, \cdot)$ be the Bregman divergence induced by \mathcal{R} . Then, any update of the form for convex function f

$$w^* = \arg \min_{w \in \mathcal{K}} \{f(w) + D_{\mathcal{R}}(w, c)\}$$

satisfies the following inequality

$$\langle w^* - d, \nabla f(w^*) \rangle \leq D_{\mathcal{R}}(d, c) - D_{\mathcal{R}}(d, w^*) - D_{\mathcal{R}}(w^*, c)$$

for any $d \in \mathcal{K}$.

Proof. By the convexity of f and \mathcal{R} , it is easy to verify the convexity of $D_{\mathcal{R}}$. Then $f(w) + D_{\mathcal{R}}(w, c)$ is convex, and by the optimality of w^* and KKT condition (Theorem 2.2 (Hazan et al., 2016)), we have

$$\langle d - w^*, \nabla_{w^*}(f(w^*) + D_{\mathcal{R}}(w^*, c)) \rangle \geq 0, \forall d \in \mathcal{K}.$$

By the definition of Bregman divergence, we can see that

$$\langle d - w^*, \nabla f(w^*) + \nabla R(w^*) - \nabla R(c) \rangle \geq 0, \forall d \in \mathcal{K}.$$

Thus we obtain

$$\langle w^* - d, \nabla f(w^*) \rangle \leq \langle d - w^*, \nabla R(w^*) - \nabla R(c) \rangle, \forall d \in \mathcal{K}.$$

The rest is the same with Lemma 4. For completeness, we present the proof here. By the definition of Bergman divergence, we know that

$$\begin{aligned} & D_{\mathcal{R}}(d, c) - D_{\mathcal{R}}(d, w^*) - D_{\mathcal{R}}(w^*, c) \\ &= R(d) - R(c) - \nabla R(c)(d - c) - (R(d) - R(w^*) - \nabla R(w^*)(d - w^*)) \\ &\quad - (R(w^*) - R(c) - \nabla R(c)(w^* - c)) \\ &= -\langle \nabla R(c), d \rangle + \langle \nabla R(w^*), d - w^* \rangle + \langle \nabla R(c), w^* \rangle \\ &= \langle d - w^*, \nabla R(w^*) - \nabla R(c) \rangle. \end{aligned}$$

Therefore, we finally conclude for $\forall d \in \mathcal{K}$

$$\langle w^* - d, \nabla f(w^*) \rangle \leq D_{\mathcal{R}}(d, c) - D_{\mathcal{R}}(d, w^*) - D_{\mathcal{R}}(w^*, c). \quad \blacksquare$$

Usually, previous works handle the noisy gradient by the property of $\mathbb{E}[f_t(w_t)|w_t] = l_t(w_t)$. However, since x_t and w_t are mutually depended, we have $\mathbb{E}[f_t(w_t)|w_t] \neq l_t(w_t)$. Fortunately, using the linearity of expectation and by the law of total expectation, we wouldn't need to handle the conditional expectation directly, as the following Lemma.

Lemma 10 (Restatement of Lemma 2). For algorithm 1. We have for $t = 1, \dots, T$

$$\mathbb{E}[l_t(w_t) - l_t(u_t)] \leq \mathbb{E}[\langle \nabla f_t(w_t), w_t - u_t \rangle] + \mathbb{E}[2(LD + G)\|w_t - \hat{w}_t\|]. \quad (1)$$

Proof. First, from the above condition and by the convexity, we have

$$\begin{aligned} \mathbb{E}[l_t(w_t) - l_t(u_t)] &\leq \mathbb{E}[\langle \nabla l_t(w_t), w_t - u_t \rangle] \\ &= \mathbb{E}[\langle \mathbb{E}[\nabla f_t(w_t)|w_t], w_t - u_t \rangle] + \mathbb{E}[\langle \nabla l_t(w_t) - \mathbb{E}[\nabla f_t(w_t)|w_t], w_t - u_t \rangle] \\ &= \mathbb{E}[\mathbb{E}[\langle \nabla f_t(w_t), w_t - u_t \rangle | w_t]] + \mathbb{E}[\langle \nabla l_t(w_t) - \mathbb{E}[\nabla f_t(w_t)|w_t], w_t - u_t \rangle] \\ &= \mathbb{E}[\langle \nabla f_t(w_t), w_t - u_t \rangle] + \mathbb{E}[\langle \nabla l_t(w_t) - \mathbb{E}[\nabla f_t(w_t)|w_t], w_t - u_t \rangle]. \end{aligned}$$

Since \hat{w}_t does not depend on x_t , we know that $\nabla l_t(\hat{w}_t) = \mathbb{E}[\nabla f_t(\hat{w}_t)|\hat{w}_t]$. We then begin to estimate the second term, which can be decomposed as

$$\begin{aligned} \mathbb{E}[\langle \nabla l_t(w_t) - \mathbb{E}[\nabla f_t(w_t)|w_t], w_t - u_t \rangle] &= \underbrace{\mathbb{E}[\langle \nabla l_t(w_t) - \nabla l_t(\hat{w}_t), w_t - u_t \rangle]}_{\text{term A}} \\ &\quad + \underbrace{\mathbb{E}[\langle \mathbb{E}[\nabla f_t(\hat{w}_t)|\hat{w}_t] - \nabla f_t(\hat{w}_t), w_t - u_t \rangle]}_{\text{term B}} \\ &\quad + \underbrace{\mathbb{E}[\langle \nabla f_t(\hat{w}_t) - \mathbb{E}[\nabla f_t(w_t)|w_t], w_t - u_t \rangle]}_{\text{term C}}. \end{aligned}$$

We next bound each term step by step. First, by the assumption of L -smoothness (Assumption 4), we can bound the term A by

$$\begin{aligned} \text{term A} &= \mathbb{E}[\langle \nabla l_t(w_t) - \nabla l_t(\hat{w}_t), w_t - u_t \rangle] \\ &\leq \mathbb{E}[\|\nabla l_t(w_t) - \nabla l_t(\hat{w}_t)\| \|w_t - u_t\|] \\ &\leq \mathbb{E}[L\|w_t - \hat{w}_t\| \|w_t - u_t\|] \\ &\leq \mathbb{E}[LD\|w_t - \hat{w}_t\|]. \end{aligned}$$

The first inequality comes from Hölder inequality, and the last one is from the bounded space (Assumption 5). Second, by the law of total expectation, we have $\mathbb{E}[\langle \mathbb{E}[\nabla f_t(\hat{w}_t)|\hat{w}_t], \hat{w}_t - u_t \rangle] = \mathbb{E}[\langle \nabla f_t(\hat{w}_t), \hat{w}_t - u_t \rangle]$, hence the term B can be bounded as

$$\begin{aligned} \text{term B} &= \mathbb{E}[\langle \mathbb{E}[\nabla f_t(\hat{w}_t)|\hat{w}_t] - \nabla f_t(\hat{w}_t), w_t - u_t \rangle] \\ &= \mathbb{E}[\langle \mathbb{E}[\nabla f_t(\hat{w}_t)|\hat{w}_t] - \nabla f_t(\hat{w}_t), \hat{w}_t - u_t \rangle] + \mathbb{E}[\langle \nabla l_t(\hat{w}_t) - \nabla f_t(\hat{w}_t), w_t - \hat{w}_t \rangle] \\ &= \mathbb{E}[\langle \mathbb{E}[\nabla f_t(\hat{w}_t)|\hat{w}_t], \hat{w}_t - u_t \rangle] - \mathbb{E}[\langle \nabla f_t(\hat{w}_t), \hat{w}_t - u_t \rangle] \\ &\quad + \mathbb{E}[\langle \nabla l_t(\hat{w}_t) - \nabla f_t(\hat{w}_t), w_t - \hat{w}_t \rangle] \\ &\leq 0 + \mathbb{E}[\|\nabla l_t(\hat{w}_t) - \nabla f_t(\hat{w}_t)\| \|w_t - \hat{w}_t\|] \\ &\leq 0 + \mathbb{E}[2G\|w_t - \hat{w}_t\|]. \end{aligned}$$

The last inequality is from Hölder inequality and bounded gradient (Assumption 4). Finally, we have for term C

$$\begin{aligned} \text{term C} &= \mathbb{E}[\langle \nabla f_t(\hat{w}_t) - \mathbb{E}[\nabla f_t(w_t)|w_t], w_t - u_t \rangle] \\ &= \mathbb{E}[\langle \nabla f_t(\hat{w}_t), w_t - u_t \rangle] - \mathbb{E}[\mathbb{E}[\nabla f_t(w_t)|w_t], w_t - u_t \rangle] \\ &= \mathbb{E}[\langle \nabla f_t(\hat{w}_t), w_t - u_t \rangle] - \mathbb{E}[\mathbb{E}[\nabla f_t(w_t), w_t - u_t | w_t]] \\ &= \mathbb{E}[\langle \nabla f_t(\hat{w}_t) - \nabla f_t(w_t), w_t - u_t \rangle] \\ &\leq \mathbb{E}[L\|w_t - u_t\| \|w_t - \hat{w}_t\|] \\ &\leq \mathbb{E}[LD\|w_t - \hat{w}_t\|]. \end{aligned}$$

The last two inequality is from Hölder inequality and L -smoothness, and the last one is from bounded space (Assumption 5). By summing up, we get

$$\mathbb{E}[l_t(w_t) - l_t(u_t)] \leq \mathbb{E}[\langle \nabla f_t(w_t), w_t - u_t \rangle] + \mathbb{E}[2(LD + G)\|w_t - \hat{w}_t\|]. \quad (2)$$

■

The continual domain shift (Assumption 1) and bounded function (Assumption 5) lead to the temporal variability condition in the online learning (as shown in Proposition 1). It is not easy to analyze the dynamic regret (temporal variability form) directly, thus we first provide the path-length version as the following lemma.

Lemma 11. *Under the same assumption as Lemma 1. If we choose $\eta \leq \frac{1}{4(LD+G)}$, Algorithm 1 has the following bound*

$$\mathbb{E}\left[\sum_{t=1}^T l_t(w_t)\right] - \sum_{t=1}^T l_t(u_t) \leq (2\eta G^2 + 2GD)\mathbb{E}\left[\sum_{t=1}^T M_t\right] + \sum_{t=1}^{T-1} \frac{1}{\eta} \gamma \|u_{t+1} - u_t\| + \frac{1}{\eta} D_{\mathcal{R}}(u_1, \hat{w}_1),$$

for all $u_1, \dots, u_T \in \mathcal{K}$.

Proof. Denote $\hat{f}_t(\cdot) = f(\cdot; x_t, \hat{y}_t)$, $\tilde{f}_t(\cdot) = f(\cdot; x_t, \tilde{y}_t)$, $f_t(\cdot) = f(\cdot; x_t, y_t)$ for simplicity. By Lemma 2, we have

$$\begin{aligned}
 \mathbb{E}[\sum_{t=1}^T l_t(w_t)] - \sum_{t=1}^T l_t(u_t) &= \mathbb{E}[\sum_{t=1}^T l_t(w_t) - \sum_{t=1}^T l_t(u_t)] \\
 &\leq \mathbb{E}[\langle \nabla f_t(w_t), w_t - u_t \rangle + 2(LD + G)\|w_t - \hat{w}_t\|] \\
 &= \mathbb{E}[\underbrace{\langle \nabla f_t(w_t) - \nabla \hat{f}_t(w_t), w_t - \hat{w}_{t+1} \rangle}_{\text{term A}} + \underbrace{\langle \nabla \hat{f}_t(w_t), w_t - \hat{w}_{t+1} \rangle}_{\text{term B}} + \underbrace{\langle \nabla \tilde{f}_t(w_t), \hat{w}_{t+1} - u_t \rangle}_{\text{term C}} \\
 &\quad + \underbrace{\langle \nabla f_t(w_t) - \nabla \tilde{f}_t(w_t), \hat{w}_{t+1} - u_t \rangle}_{\text{term D}} + 2(LD + G)\|w_t - \hat{w}_t\|].
 \end{aligned}$$

We next bound each term step by step. First, we can bound term A in terms of the pseudolabel errors.

$$\begin{aligned}
 \text{term A} &= \langle \nabla f_t(w_t) - \nabla \hat{f}_t(w_t), w_t - \hat{w}_{t+1} \rangle \\
 &= M_t \langle \nabla f_t(w_t) - \nabla \hat{f}_t(w_t), w_t - \hat{w}_{t+1} \rangle \\
 &\leq M_t \|\nabla f_t(w_t) - \nabla \hat{f}_t(w_t)\|_* \|w_t - \hat{w}_{t+1}\| \\
 &\leq M_t 2G \|w_t - \hat{w}_{t+1}\| \\
 &\leq 2\eta M_t G^2 + \frac{1}{2\eta} \|w_t - \hat{w}_{t+1}\|^2.
 \end{aligned}$$

The first inequality holds due to Hölder inequality, and the last one holds due to the fact that $2ab \leq 2\eta a^2 + \frac{1}{2\eta} b^2$ for $\eta > 0$ and $M_t^2 = M_t$. By Lemma 9, we could bound term B

$$\begin{aligned}
 \text{term B} &= \langle \nabla \hat{f}_t(w_t), w_t - \hat{w}_{t+1} \rangle \\
 &\leq \frac{1}{\eta} (D_{\mathcal{R}}(\hat{w}_{t+1}, \hat{w}_t) - D_{\mathcal{R}}(\hat{w}_{t+1}, w_t) - D_{\mathcal{R}}(w_t, \hat{w}_t)) \\
 &\leq \frac{1}{\eta} (D_{\mathcal{R}}(\hat{w}_{t+1}, \hat{w}_t) - \frac{1}{2} \|w_t - \hat{w}_{t+1}\|^2 - \frac{1}{2} \|w_t - \hat{w}_t\|^2).
 \end{aligned}$$

The last inequality is from the strongly convexity of regularizer R . By Lemma 4, we next bound term C

$$\text{term C} = \langle \nabla \tilde{f}_t(w_t), \hat{w}_{t+1} - u_t \rangle \leq \frac{1}{\eta} (D_{\mathcal{R}}(u_t, \hat{w}_t) - D_{\mathcal{R}}(u_t, \hat{w}_{t+1}) - D_{\mathcal{R}}(\hat{w}_{t+1}, \hat{w}_t)).$$

From the Algorithm 1, we know that only when the pseudolabel makes mistake and the active agent does not query the label, $\tilde{y} \neq y$. Similar to term A, we have the bound for term D

$$\begin{aligned}
 \text{term D} &= \langle \nabla f_t(w_t) - \nabla \tilde{f}_t(w_t), \hat{w}_{t+1} - u_t \rangle \\
 &= \langle M_t(1 - Z_t)(\nabla f_t(w_t) - \nabla \tilde{f}_t(w_t)), \hat{w}_{t+1} - u_t \rangle \\
 &\leq M_t(1 - Z_t) \|\nabla f_t(w_t) - \nabla \tilde{f}_t(w_t)\|_* \|\hat{w}_{t+1} - u_t\| \\
 &\leq 2M_t(1 - Z_t)GD \\
 &\leq 2M_tGD.
 \end{aligned}$$

The first inequality is from the Hölder inequality, and second inequality is from the Assumption 4 and Assumption 5. Also from the Hölder inequality, the last term can be bounded as $2(LD + G)\|w_t - \hat{w}_t\| \leq 2(LD + G)^2\eta + \frac{1}{2\eta} \|w_t - \hat{w}_t\|^2$.

Finally, we have the path-length version of dynamic regret bound

$$\begin{aligned}
 & \mathbb{E}\left[\sum_{t=1}^T l_t(w_t)\right] - \sum_{t=1}^T l_t(u_t) \\
 & \leq \mathbb{E}\left[\sum_{t=1}^T \text{term A} + \text{term B} + \text{term C} + \text{term D} + 2(LD + G)^2\eta + \frac{1}{2\eta}\|w_t - \hat{w}_t\|^2\right] \\
 & \leq \mathbb{E}\left[\sum_{t=1}^T 2\eta M_t G^2 + \sum_{t=1}^T 2M_t G D + \sum_{t=1}^T \frac{1}{\eta}(D_{\mathcal{R}}(u_t, \hat{w}_t) - D_{\mathcal{R}}(u_t, \hat{w}_{t+1}))\right] + 2(LD + G)^2\eta T \\
 & = (2\eta G^2 + 2GD)\mathbb{E}\left[\sum_{t=1}^T M_t\right] + \mathbb{E}\left[\sum_{t=1}^T \frac{1}{\eta}(D_{\mathcal{R}}(u_t, \hat{w}_t) - D_{\mathcal{R}}(u_t, \hat{w}_{t+1}))\right] + 2(LD + G)^2\eta T.
 \end{aligned}$$

By the condition $D_{\mathcal{R}}(x, z) - D_{\mathcal{R}}(y, z) \leq \gamma\|x - y\|, \forall x, y, z \in \mathcal{K}$, we can get

$$\begin{aligned}
 & \mathbb{E}\left[\sum_{t=1}^T \frac{1}{\eta}(D_{\mathcal{R}}(u_t, \hat{w}_t) - D_{\mathcal{R}}(u_t, \hat{w}_{t+1}))\right] \\
 & \leq \mathbb{E}\left[\sum_{t=1}^{T-1} \frac{1}{\eta}(D_{\mathcal{R}}(u_{t+1}, \hat{w}_{t+1}) - D_{\mathcal{R}}(u_t, \hat{w}_{t+1}))\right] + \frac{1}{\eta}D_{\mathcal{R}}(u_1, \hat{w}_1) \\
 & \leq \sum_{t=1}^{T-1} \frac{1}{\eta}\gamma\|u_{t+1} - u_t\| + \frac{1}{\eta}D_{\mathcal{R}}(u_1, \hat{w}_1).
 \end{aligned}$$

From the above, we thus have

$$\mathbb{E}\left[\sum_{t=1}^T l_t(w_t)\right] - \sum_{t=1}^T l_t(u_t) \leq (2\eta G^2 + 2GD)\mathbb{E}\left[\sum_{t=1}^T M_t\right] + 2(LD + G)^2\eta T + \frac{\epsilon_w}{\eta} + \sum_{t=1}^{T-1} \frac{1}{\eta}\gamma\|u_{t+1} - u_t\|.$$

■

Next, we give a general version of our regret bound analysis, concluding both the separable case and the general case.

Theorem 5. *Take the same assumptions as Lemma 1, Algorithm 1 has the following bound for $\eta \leq \frac{1}{4(LD+G)}$*

$$\text{D-Regret}(\{P_t\}, T) \leq (2\eta G^2 + 2GD)\mathbb{E}\left[\sum_{t=1}^T M_t\right] + 2(LD + G)^2\eta + \frac{\epsilon_w + \gamma D}{\eta}T + 4\sqrt{\frac{\gamma D T F V_T}{\eta}}.$$

Proof Sketch. This proof shares the same technique with Zhang et al. (2020b), which introduces detailed proof for converting path-length bound to temporal variability bound. The key of this converting is to specify a sequence of $\{u_1, \dots, u_T\}$ in the following way.

$$\{u_1, \dots, u_T\} = \left\{ w_1^*, \underbrace{w_{\mathcal{I}_2}^*, \dots, w_{\mathcal{I}_2}^*}_{\Delta \text{ times}}, \underbrace{w_{\mathcal{I}_3}^*, \dots, w_{\mathcal{I}_3}^*}_{\Delta \text{ times}}, \dots, \underbrace{w_{\mathcal{I}_{\lceil T-\Delta \rceil+1}}^*, \dots, w_{\mathcal{I}_{\lceil T-\Delta \rceil+1}}^*}_{\Delta \text{ times}} \right\}.$$

This piece-wise stationary sequence starts with w_1^* and next changes every $\Delta \in [T]$ iterations. We specify u_t as the best fixed decision $w_{\mathcal{I}_i}^* = \arg \min_{w \in \mathcal{K}} \sum_{t \in \mathcal{I}_i} l_t(w)$ of the corresponding interval \mathcal{I}_i . The rest is same as the proof of Lemma 2 in (Zhang et al., 2020b). ■

Within Theorem 5, it is simple to bound both the separable and the general (non-separable) cases.

Proof of Theorem 1. By the result of Lemma 1 we know that

$$\mathbb{E}\left[\sum_{t=1}^T M_t\right] \leq \frac{2G^2}{\sigma^2}(\gamma C_T + \epsilon_v).$$

Plugging in Theorem 5, we then have

$$\begin{aligned} \text{D-Regret}(\{P_t\}, T) &\leq (2\eta G^2 + 2GD)\mathbb{E}\left[\sum_{t=1}^T M_t\right] + 2(LD + G)^2\eta T + \frac{\epsilon_w + \gamma D}{\eta} + 4\sqrt{\frac{\gamma DTFV_T}{\eta}} \\ &\leq \frac{4(\eta G^4 + G^3D)}{\sigma^2}(\gamma C_T + \epsilon_v) + 2(LD + G)^2\eta T + \frac{\epsilon_w + \gamma D}{\eta} + 4\sqrt{\frac{\gamma DTFV_T}{\eta}}. \end{aligned}$$

■

Similarly, we can generate it to the separable case.

Proof of Theorem 4. By the result of Lemma 3, we know that

$$\mathbb{E}\left[\sum_{t=1}^T M_t\right] \leq \frac{2G^2}{\sigma^2}(\gamma C_T + \epsilon_v + \frac{\sigma}{G^2}T\alpha^*).$$

Plugging in Theorem 5, we then have

$$\begin{aligned} \text{D-Regret}(\{P_t\}, T) &\leq (2\eta G^2 + 2GD)\mathbb{E}\left[\sum_{t=1}^T M_t\right] + 2(LD + G)^2\eta T + \frac{\epsilon_w + \gamma D}{\eta} + 4\sqrt{\frac{\gamma DTFV_T}{\eta}} \\ &\leq \frac{4(\eta G^4 + G^3D)}{\sigma^2}(\gamma C_T + \epsilon_v + \frac{\sigma}{G^2}T\alpha^*) + 2(LD + G)^2\eta T + \frac{\epsilon_w + \gamma D}{\eta} + 4\sqrt{\frac{\gamma DTFV_T}{\eta}}. \end{aligned}$$

■

D.3 Lower bound

Here, we show the lower bound for Theorem 2.

Proof of Theorem 2. We here create an example for the worst case that satisfies our assumptions.

Assume we have two data points $(-1, 0)$ and $(1, 0)$ with the same probability $1/2$ to be sampled. Then we let $(-1, 0)$ to be class 1 and $(1, 0)$ to be class -1 when $t = 1, \dots, \frac{T}{2}$, and let $(-1, 0)$ to be class -1 and $(1, 0)$ to be class 1 when $t = \frac{T}{2} + 1, \dots, T$. We use the hinge loss $l_t = \max\{0, 1 - y_t w_t x_t\}$, and the decision space is $\{w \mid \|w\|_2 \leq 1\}$. It is easy to verify that this setting satisfies our assumptions where $V_T \leq 2, C_T \leq 2$.

For any unsupervised self-training algorithm that begins with a good initial $w = (-1, 0)$, it is impossible to get the information for the label change in hindsight. Then the learner takes no update when $t = \frac{T}{2} + 1, \dots, T$, and therefore suffers from $T/2$ regret, which is an order of T .

■

E Extension to Multiclass

In this section, we extend the results to the multiclass case.

Algorithm 2 Multiclass Online Self Adaptive Mirror Descent (MOSAMD)

Input: Active probability controller σ , aggressive step size τ_t , conservative step size η , initial data.
Initial: Learn from initial data, get aggressive model θ_1 and conservative model \hat{w}_1 .
for $t = 1, \dots, T$ **do**
 observe data sample x_t
 pseudolabel:
 give the pseudolabel provided by the aggressive model $\hat{y}_t = \arg \max_{s_t \in \mathcal{Y}} H_t^{s_t}(\theta_t)$
 self-adaptation:
 adapt the conservative model $w_t = \arg \min_{w \in \mathcal{K}} \eta f(w; x_t, \hat{y}_t) + D_{\mathcal{R}}(w, \hat{w}_t)$ before making the decision
 active query:
 compute the confidence score $p_t = H_t^{\hat{y}_t}(\theta_t) - \max_{s_t \neq \hat{y}_t, s_t \in \mathcal{Y}} H_t^{s_t}(\theta_t)$
 draw a Bernoulli random variable with probability $Z_t \sim \text{Bernoulli}(\sigma / (\sigma + p_t))$
 if $Z_t = 1$ **then**
 query label y_t , compute the margin $\Psi_t(\theta_t) = H_t^{y_t}(\theta_t) - \max_{s_t \neq y_t, s_t \in \mathcal{Y}} H_t^{s_t}(\theta_t)$, and let $\tilde{y}_t = y_t$
 update the aggressive model $\theta_{t+1} = \arg \min_{\theta \in \mathcal{K}} -\tau_t \langle \nabla \Psi_t(\theta_t), \theta \rangle + D_{\mathcal{R}}(\theta, \theta_t)$
 else
 let $\theta_{t+1} = \theta_t$ and $\tilde{y}_t = \hat{y}_t$
 end if
 update the conservative model $\hat{w}_{t+1} = \arg \min_{w \in \mathcal{K}} \eta \langle \nabla f(w_t; x_t, \tilde{y}_t), w \rangle + D_{\mathcal{R}}(w, \hat{w}_t)$
end for

E.1 Multiclass setting

Denote \mathcal{Y} to be the class set. The multiclass setting is slightly different from the binary setting, and we present the important formulations and assumptions of multiclass case as follows.

Denote the soft prediction over instance x as $H(\theta; x)$, which outputs $|\mathcal{Y}|$ prediction scores:

$$H^s(\theta; x), s \in \mathcal{Y}.$$

Denote $H(\theta; x_t) = H_t(\theta)$ for simplicity. In each round t , the margin function is defined as

$$\Psi_t(\theta) := H_t^{y_t}(\theta) - \max_{s_t \neq y_t, s_t \in \mathcal{Y}} H_t^{s_t}(\theta),$$

which is the gap between the prediction score of the real class and the irrelevant class with the highest score. Assumption 2 is modified as

Assumption 6 (Multiclass Separation). *There exists $\alpha^* > 0$ such that for each time step $t = 1, \dots, T$, the data distribution P_t can be classified almost correctly with a margin R , i.e., there exists $v_t \in \mathcal{K}$ and a constant α^* such that*

$$\mathbb{E}_{(x_t, y_t) \sim P_t} [\max\{0, R - \Psi_t(v_t)\}] \leq \alpha^*.$$

We further assume that there exists a constant C_T such that

$$\sum_{t=1}^{T-1} \|v_t - v_{t+1}\| \leq C_T,$$

i.e., the classifiers with margin R change continually.

We further assume the convexity and bounded gradient on the margin function.

Assumption 7 (Margin Function). *We assume that $-\Psi_t(\cdot)$ is convex, and $\|\nabla \Psi_t(\theta_t)\|_* \leq G$.*

Others are the same as the binary class case.

E.2 Multiclass OSAMD

We here present the Multiclass OSAMD (MOSAMD) as Algorithm 2. Specifically, we modify three areas in the binary OSAMD:

1. Pseudolabel is given by the class with the largest prediction scores $\hat{y}_t = \max_{s_t \in \mathcal{Y}} H_t^{s_t}(\theta_t)$;

2. The uncertainty is measured by the difference between the largest and second largest prediction scores $p_t = H_t^{y_t}(\theta_t) - \max_{s_t \neq \hat{y}_t, s_t \in \mathcal{Y}} H_t^{s_t}(\theta_t)$, which is designed to compute the query rate;
3. The margin is defined by the gap between the prediction score of the real class and the irrelevant class with the highest score $\Psi_t(\theta) = H_t^{y_t}(\theta) - \max_{s_t \neq y_t, s_t \in \mathcal{Y}} H_t^{s_t}(\theta)$, based on which the pseudolabel (aggressive) model updates.

E.3 Analysis

In this subsection, we analyze the theoretical performance of MOSAMD in the general case, which can be reduced to the separable case by setting $C = \infty$. We first begin with the pseudolabel errors bound, and then present the dynamic regret bound.

E.3.1 Pseudolabel Errors

Here, we present the theoretical bound of pseudolabel errors for the multiclass case.

Lemma 12 (Pseudolabel Errors). *Let regularizer $\mathcal{R} : \mathcal{K} \mapsto \mathbb{R}$ be a 1-strongly convex function on \mathcal{K} with respect to a norm $\|\cdot\|$. Assume that $D_{\mathcal{R}}(\cdot, \cdot)$ satisfies $D_{\mathcal{R}}(x, z) - D_{\mathcal{R}}(y, z) \leq \gamma\|x - y\|, \forall x, y, z \in \mathcal{K}$. Set $\tau_t = \min\{C, \frac{\max\{0, \sigma - \Psi_t(\theta_t)\}}{\|\nabla \Psi_t(\theta_t)\|_*^2}\}$, $\sigma \leq R$. The expected number of pseudolabel errors made by Algorithm 2 is bounded by*

$$\mathbb{E}[\sum_{t=1}^T M_t] \leq \frac{2G^2}{\sigma^2} (\gamma C_T + \epsilon_v + \frac{\sigma}{G^2} T \alpha^*).$$

where $M_t = \mathbf{1}_{\hat{y}_t \neq y_t}$ is the instantaneous mistake indicator.

The proof shares the same idea with the binary case. Before proving the theorem, we shall begin with important lemmas.

Denote $f_t^r(\theta) = \max\{0, r - \Psi_t(\theta)\}$. We first give the recursive of the multiclass case.

Lemma 13. *Take the same assumptions as Lemma 12. Then for algorithm 2, the following inequality holds*

$$\tau_t r - \tau_t \Psi_t(\theta_t) - \frac{\tau_t^2}{2} \|\nabla \Psi_t(\theta_t)\|_*^2 \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t).$$

for $r > 0$.

Proof. First, by the definition of f_t^r , we have

$$\begin{aligned} r - f_t^r(v_t) &= r - \max\{0, r - \Psi_t(v_t)\} \leq \Psi_t(v_t) \\ &= \Psi_t(v_t) - \Psi_t(\theta_t) + \Psi_t(\theta_t). \end{aligned}$$

By the convexity of $-\Psi_t(\cdot)$, we have

$$\begin{aligned} \Psi_t(v_t) - \Psi_t(\theta_t) &= -\Psi_t(\theta_t) - (-\Psi_t(v_t)) \\ &\leq \langle -\nabla \Psi_t(\theta_t), \theta_t - v_t \rangle \\ &\leq \langle -\nabla \Psi_t(\theta_t), \theta_{t+1} - v_t \rangle + \langle -\Psi_t(\theta_t), \theta_t - \theta_{t+1} \rangle. \end{aligned}$$

By the update rule of θ and Lemma 4, the first term can be bounded that

$$\langle -\nabla \Psi_t(\theta_t), \theta_{t+1} - v_t \rangle \leq \frac{1}{\tau_t} (D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) - D_{\mathcal{R}}(\theta_{t+1}, \theta_t)).$$

Due to Hölder inequality and the fact that $ab \leq \frac{\eta}{2} a^2 + \frac{1}{2\eta} G^2$ for $\eta > 0$, we obtain for the second term

$$\begin{aligned} \langle -\nabla \Psi_t(\theta_t), \theta_t - \theta_{t+1} \rangle &\leq \|\nabla \Psi_t(\theta_t)\|_* \|\theta_{t+1} - \theta_t\| \\ &\leq \frac{\tau_t}{2} \|\nabla \Psi_t(\theta_t)\|_*^2 + \frac{1}{2\tau_t} \|\theta_{t+1} - \theta_t\|^2. \end{aligned}$$

Due to the strong convexity of regularizer \mathcal{R} , we have $D_{\mathcal{R}}(x, y) \geq \frac{1}{2}\|x - y\|^2$ for any $x, y \in \mathcal{X}$ (Mohri et al., 2018). Therefore, by plugging the above term, we obtain that

$$\begin{aligned} r - f_t^r(v_t) &\leq \frac{1}{\tau_t}(D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) - \frac{1}{2}\|\theta_{t+1} - \theta_t\|^2) \\ &\quad + \frac{\tau_t}{2}\|\nabla\Psi_t(\theta_t)\|_*^2 + \frac{1}{2\tau_t}\|\theta_{t+1} - \theta_t\|^2 + \Psi_t(\theta_t). \end{aligned}$$

By rearranging, we have

$$\tau_t r - \tau_t \Psi_t(\theta_t) - \frac{\tau_t^2}{2}\|\nabla\Psi_t(\theta_t)\|_*^2 \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t).$$

■

Denote the instantaneous mistake by $M_t(w) = \mathbf{1}_{\hat{y}_t \neq y_t}$, and let $L_t(w) = \mathbf{1}_{\hat{y}_t = y_t, \Psi_t(w) \leq r}$ to be the indicator of the right decision but in the margin r , where $\mathbf{1}_{(\cdot)}$ is the indicator function. We then have the following relationship

Lemma 14. *Take the same assumptions as Lemma 12. For Algorithm 2, let $\tau_t = 0$ if $f_t^r(\theta_t) = 0$, then the following inequality holds for every t*

$$\begin{aligned} M_t Z_t \tau_t (r + |\Psi_t(\theta_t)| - \frac{\tau_t}{2}\|\nabla\Psi_t(\theta_t)\|_*^2) + L_t Z_t \tau_t (r - |\Psi_t(\theta_t)| - \frac{\tau_t}{2}\|\nabla\Psi_t(\theta_t)\|_*^2) \\ \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t) \end{aligned}$$

for $r > 0$.

Proof. From Lemma 13, we know that

$$\tau_t r - \tau_t \Psi_t(\theta_t) - \frac{\tau_t^2}{2}\|\nabla\Psi_t(\theta_t)\|_*^2 \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t).$$

Therefore, we can obtain

$$\begin{aligned} M_t Z_t \tau_t (r + |\Psi_t(\theta_t)| - \frac{\tau_t}{2}\|\nabla\Psi_t(\theta_t)\|_*^2) + L_t Z_t \tau_t (r - |\Psi_t(\theta_t)| - \frac{\tau_t}{2}\|\nabla\Psi_t(\theta_t)\|_*^2) \\ \leq M_t Z_t (D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t)) + L_t Z_t (D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t)) \\ = (M_t + L_t) Z_t (D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t)). \end{aligned}$$

From Algorithm 2, we know that if $Z_t = 0$, then $\tau_t = 0, \theta_t = \theta_{t+1}$. And if $M_t + L_t = 0$, we get $\Psi_t(\theta_t) \geq r$, then $\tau_t = 0, \theta_t = \theta_{t+1}$. Therefore, we have

$$(M_t + L_t) Z_t (D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t)) = D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t).$$

We finally get

$$\begin{aligned} M_t Z_t \tau_t (r + |\Psi_t(\theta_t)| - \frac{\tau_t}{2}\|\nabla\Psi_t(\theta_t)\|_*^2) + L_t Z_t \tau_t (r - |\Psi_t(\theta_t)| - \frac{\tau_t}{2}\|\nabla\Psi_t(\theta_t)\|_*^2) \\ \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t). \end{aligned}$$

■

Next, we give a similar result as Lemma 8 of binary case.

Lemma 15. *Take the same assumptions as Lemma 12. Then for Algorithm 2, let $\tau_t = \min\{C, \frac{f_t^r(\theta_t)}{\|\nabla\Psi_t(\theta_t)\|_*^2}\}$. then the following inequality holds*

$$\min\{C, \frac{r}{G^2}\} \frac{1}{2} M_t Z_t (r + p_t) \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t),$$

for $r \leq R$.

Proof. First, according to Lemma 14, we have

$$\begin{aligned} & M_t Z_t \tau_t (r + |\Psi_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla \Psi_t(\theta_t)\|_*^2) + L_t Z_t \tau_t (r - |\Psi_t(\theta_t)| - \frac{\tau_t}{2} \|\nabla \Psi_t(\theta_t)\|_*^2) \\ & \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t). \end{aligned}$$

Since we take

$$\tau_t = \min\left\{C, \frac{f_t^r(\theta_t)}{\|\nabla \Psi_t(\theta_t)\|_*^2}\right\} \leq f_t^r(\theta_t) / \|\nabla \Psi_t(\theta_t)\|_*^2.$$

Similar to Lemma 7, we have

$$\frac{\tau_t}{2} M_t Z_t (r + |\Psi_t(\theta_t)|) \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \tau_t f_t^r(v_t).$$

Since we know that

$$M_t \tau_t = M_t \min\left\{C, \frac{f_t^r(\theta_t)}{\|\nabla \Psi_t(\theta_t)\|_*^2}\right\} \leq M_t \min\left\{C, \frac{r}{G^2}\right\}.$$

Since $\tau_t \leq C$. We therefore have

$$\min\left\{C, \frac{r}{G^2}\right\} \frac{1}{2} M_t Z_t (r + |\Psi_t(\theta_t)|) \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + C f_t^r(v_t).$$

By the definition, we could infer that $p_t \leq |\Psi_t(\theta_t)|$. Because if $\hat{y} = y$ then $p_t = |\Psi_t(\theta_t)|$, and if $\hat{y} \neq y$ then $H_t^y(\theta_t) \leq H_t^{s_t}(\theta_t)$, where $s_t = \max_{s_t \neq \hat{y}_t, s_t \in \mathcal{Y}} H_t^{s_t}(\theta_t)$, which leads to

$$p_t = H_t^{\hat{y}}(\theta_t) - H_t^{s_t}(\theta_t) \leq H_t^{\hat{y}}(\theta_t) - H_t^y(\theta_t) = |\Psi_t(\theta_t)|.$$

Thus we obtain

$$\begin{aligned} \min\left\{C, \frac{r}{G^2}\right\} \frac{1}{2} M_t Z_t (r + p_t) & \leq \min\left\{C, \frac{r}{G^2}\right\} \frac{1}{2} M_t Z_t (r + |\Psi_t(\theta_t)|) \\ & \leq D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + C f_t^r(v_t). \end{aligned}$$

■

Within the above lemmas, we are now ready to prove the Theorem 12.

Proof of Lemma 12. First by the proof of Lemma 1, we have

$$\sum_{t=1}^T D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) \leq \epsilon_v + \gamma C_T.$$

Then

$$\begin{aligned} \mathbb{E}\left[\sum_{t=1}^T M_t\right] & = \frac{1}{r} \mathbb{E}\left[\sum_{t=1}^T M_t Z_t (r + p_t)\right] \\ & = \frac{2}{r^2} \max\left\{\frac{r}{C}, G^2\right\} \mathbb{E}\left[\sum_{t=1}^T \min\left\{C, \frac{r}{G^2}\right\} \frac{1}{2} M_t Z_t (r + p_t)\right] \\ & \leq \frac{2}{r^2} \max\left\{\frac{r}{C}, G^2\right\} \mathbb{E}\left[\sum_{t=1}^T D_{\mathcal{R}}(v_t, \theta_t) - D_{\mathcal{R}}(v_t, \theta_{t+1}) + \sum_{t=1}^T C f_t^r(v_t)\right] \\ & \leq \frac{2}{r^2} \max\left\{\frac{r}{C}, G^2\right\} (\epsilon_v + \gamma C_T + C \sum_{t=1}^T l_t^r(v_t)) \\ & = \frac{2G^2}{\sigma^2} (\gamma C_T + \epsilon_v + \frac{\sigma}{G^2} T \alpha^*), \end{aligned}$$

where $r = \sigma$, $C = \sigma/G^2$. The second inequality comes from $l_t^r(v_t) = \mathbb{E}[f_t^r(v_t)]$, and the last inequality comes from $l_t^r(v_t) \leq l_t^R(v_t) \leq \alpha^*$. We thus end the proof. ■

E.3.2 Regret Bound

The regret bound analysis is actually the same as the binary case, since the Proposition 1 and Theorem 5 do not depend on the number of class. For contentedness, we present the result as follows.

Theorem 6 (Regret Bound). *Under the same conditions and parameters in Lemma 12. Algorithm 2 achieves the following regret bound*

$$\text{D-Regret}^{OSAMD}(\{P_t\}, T) \leq \frac{4(\eta G^4 + G^3 D)}{\sigma^2} (\gamma C_T + \epsilon_v + \frac{\sigma}{G^2} T \alpha^*) + 2(LD + G)^2 \eta T + \frac{\epsilon_w + \gamma D}{\eta} + 4\sqrt{\frac{\gamma D T F V_T}{\eta}}.$$

The proof is also the same as Theorem 4. From this, we know that our result still works in the multiclass case.

F Experimental Details

In this section, we provide experimental details in Section 6.

Datasets We provide detailed descriptions of our datasets as follows:

1. *Rotating Gaussian*: We simulate a non-stationary environment with continual domain shift. We use two Gaussian distributions with center points (5, 0) and (15, 0), and covariance matrix $3I$ (I denotes identity matrix), to represent class 1 and -1 . We let the center points averagely rotate from 0° to 180° counterclockwise in 2000 time steps, and in each time, we sample one data instance. Therefore, every data sample comes from a different domain. All the time, we keep $P(Y = 1) = P(Y = -1) = 1/2$.
2. *Rotating MNIST*: We randomly select and shuffle 35000 images from the original MNIST dataset, using the first 10,000 images with no rotation as the source dataset. We averagely rotate the next 25,000 images from 0° to 90° counterclockwise to be the target dataset with a continually changing domain.
3. *Portraits*: It is a realistic dataset, which contains 37,921 photos of high school seniors labeled by gender across many years. This real dataset suffers from a natural continual domain shift, including covariate shift and label shift, as shown in previous works (Ginosar et al., 2015; Kumar et al., 2020). We downsample all the images to 32×32 , and do no other preprocessing. We take the first 2000 images as the source domain. We use the next 16000 images as target data with a continually changing domain, and test the online adaptation.
4. *Cover-Type*: It is a realistic dataset from the UCI repository. This goal is to predict the forest Cover-Type at a particular location given 54 features (Blackard and Dean, 1999). The original Cover-Type dataset contains 581012 samples and has 7 type classes to be predicted. In our experiment, we leave the examples in the first two classes (which compose the majority of the dataset, have 500k samples in total) and sort the examples by increasing horizontal distance to the water body. Then we split the data into a source domain (first 50K examples), an intermediate domain (next 400K examples), and a target domain (final 50K examples). This dataset setting follows the setting in Kumar et al. (2020).

Baselines We provide a detailed introduction about the baselines we compare. Since we are the first to study the OACA setting, no specific baseline is suitable for this setting. To demonstrate the efficacy of our design, we compare with the following baselines:

1. *Passive-aggressive active (PAA) learning*: The design of our aggressive model is similar to online active learning (OAL) algorithms. To demonstrate the advantage of the online teacher-student structure, we compare with one typical algorithm PAA. Although OAL is a well-studied topic of statistical learning, most recently proposed methods are not suitable for implementation in deep learning settings. For instance, the second-order OAL algorithm (Hao et al., 2017) is designed only for the linear model, and requires additional computation cost of the second-order matrix. More recently, Zhang et al. (2018, 2019) studied OAL with class imbalance, but provided no additional improvement in the balance cases. Therefore, comparing with PAA is sufficient to show the advantage against online active learning.
2. *Online mirror descent with all labels (OMD (all))*: Our theory shows that the regret of OSAMD is aligned with the lower bound for online learning with full labels. To verify the theoretical results, we compare with online mirror descent with all labels, which has been shown to attain the lower bound (Besbes et al., 2015; Jadbabaie et al., 2015). By observing whether the regret or accumulated loss of OSAMD is aligned with OMD (all), we could empirically verify the theory.

3. *Online mirror descent with uniform sampled labels (OMD (partial))*: Online mirror descent with uniform sampled labels is the naive way to deal with the OACA problem. By comparing it with OMD (all), we could know whether the naive method can solve this problem. By comparing it with OSAMD, we can demonstrate the advantage of our sophisticated design.

Next, we introduce the baselines for the ablation study. Recall the online teacher-student structure consists of self-adaptation and active query, we then compare with the following baselines to show the efficacy of each component:

1. *OSAMD without Self-adaptation*: To evaluate the efficacy of self-adaptation, we run a OMD with the same active queries as OSAMD. By comparing it with OSAMD, we could empirically observe the efficacy of the design of self-adaptation.
2. *OSAMD without Active-query*: To evaluate the efficacy of self-adaptation, we replace the active queries with uniform sampled labels in OSAMD. By comparing it with OSAMD, we could empirically observe the efficacy of the design of active-query.

Model and Parameters setting We provide our model and parameters setting as follows:

1. *Models*:

- For Rotating Gaussian, we set objective function to be the svm loss $f(w; x, y) = \max\{0, 1 - yw^T x\} + C\|w\|_2^2$ with penalty parameter 0.2, the soft prediction is $H(w) = w^T x$.
- For Rotating MNIST and Portraits, we design the same neural network feature extractor with two conv layers. We use filter size of 5×5 , stride of 2×2 , 64 output channels, and relu activation for each layer. After the final convolution layer, we add a dropout layer with a probability of 0.5 and a batchnorm layer after dropout. The extracted features are then flattened and fed into fully connected layers with 2 and 10 outputs, respectively, for Portraits and Rotating MNIST. Each of the output neurons is matched with a specific prediction class.
- For Cover-Type, we used a two hidden layer feedforward. Each linear hidden layer contains 30 neurons. Dropout layer with a probability of 0.5 is added before the final fully connected layer. The final output is activated by softmax.

2. *Parameters*:

- For Rotating Gaussian, the step size η is set to be 0.01 for both OMD and OSAMD. We set active controller $\sigma = 0.35$, and aggressive step size $\tau_t = \min\{1, \max\{0, 1 - y_t \theta_t^T x_t\} / \|x_t\|_2^2\}$ for OSAMD. We use l_2 norm as the regularizer \mathcal{R} , and initialize all the models with $[0.4, 0, -4]$. For the implicit gradient update of self adaptation, we run 20 inner gradient descent loops to approximate the optimal.
- For Rotating MNIST, the step size η is set to be 0.000005 for both OMD and OSAMD (abbreviation for MOSAMD). We set active controller $\sigma = 0.2$, and aggressive step size $\tau_t = \min\{0.006, 0.0027 * \max\{0, 1 - y_t \Psi_t(\theta_t)\}\}$ for OSAMD, where $\Psi_t(\theta_t)$ is the margin function of the deep learning model. l_2 norm is used as the regularizer \mathcal{R} and all the models are initialized with a model pre-trained with the source data, i.e., the first 10000 images. For the implicit gradient update of self adaptation, we run 10 inner gradient descent loops to approximate the optimal.
- For Portraits, the step size η is set to be 0.000001 for both OMD and OSAMD. We set active controller $\sigma = 0.15$, and aggressive step size $\tau_t = \min\{0.0025, 0.0012 * \max\{0, 1 - y_t H_t(\theta_t)\}\}$ for OSAMD, where $H_t(\theta_t)$ is the output of the deep learning model. l_2 norm is used as the regularizer \mathcal{R} and all the models are initialized with a model pre-trained with the source data, i.e., the first 2000 images. For the implicit gradient update of self adaptation, we run 20 inner gradient descent loops to approximate the optimal.
- For Cover-Type, the step size η is set to be 0.0000015 for both OMD and OSAMD. We set active controller $\sigma = 0.005$, and aggressive step size $\tau_t = \min\{0.02, 0.01 * \max\{0, 1 - y_t H_t(\theta_t)\}\}$ for OSAMD, where $H_t(\theta_t)$ is the output of the deep learning model. l_2 norm is used as the regularizer \mathcal{R} and all the models are initialized with a model pre-trained with the source data, i.e., the first 50K examples. For the implicit gradient update of self adaptation, we run 5 inner gradient descent loops to approximate the optimal.

Implementation:

1. *Set Up*: The training and evaluation of models are realized with PyTorch (<https://pytorch.org>). We repeat every experiment over 10 times, and report the mean performance across independent runs. We also present the confidence intervals to eschew the experimental randomness.
2. *Computation Resources*: We have run the simulation on a single Intel(R) Xeon(R) E5-2650 CPU, and the

deep learning experiments on a single 16GB GeForce GTX 1080 Ti GPU.

G Social Impact

For the social impact, as a study on a general learning problem, our work will not incur ethical issues by itself. However, ethical issues may arise if our learning method is improperly applied to some application fields - just as any other general learning method if it is misused.