Online Learning with Optimism and Delay

Genevieve Flaspohler [1] [2] Francesco Orabona [3] Judah Cohen [4] Soukayna Mouatadid [5] Miruna Oprescu [6]
Paulo Orenstein [7] Lester Mackey [6]

Abstract

Inspired by the demands of real-time climate and weather forecasting, we develop optimistic online learning algorithms that require no parameter tuning and have optimal regret guarantees under delayed feedback. Our algorithms—DORM, DORM+, and AdaHedgeD—arise from a novel reduction of delayed online learning to optimistic online learning that reveals how optimistic hints can mitigate the regret penalty caused by delay. We pair this delay-as-optimism perspective with a new analysis of optimistic learning that exposes its robustness to hinting errors and a new meta-algorithm for learning effective hinting strategies in the presence of delay. We conclude by benchmarking our algorithms on four subseasonal climate forecasting tasks, demonstrating low regret relative to state-of-the-art forecasting models.

1. Introduction

Online learning is a sequential decision-making paradigm in which a learner is pitted against a potentially adversarial environment (Shalev-Shwartz, 2007; Orabona, 2019). At time t, the learner must select a play $\mathbf{w}_t$ from some set of possible plays $\mathbf{W}$. The environment then reveals the loss function ℓ_t and the learner pays the cost $\ell_t(\mathbf{w}_t)$. The learner uses information collected in previous rounds to improve its plays in subsequent rounds. *Optimistic* online learners additionally make use of side-information or "hints" about expected future losses to improve their plays. Over a period of length T, the goal of the learner is to minimize *regret*, an objective that quantifies the performance gap between the learner and the best possible constant play in retrospect in some competitor set $\mathbf{U}$: $\text{Regret}_T = \sup_{u \in \mathbf{U}} \sum_{t=1}^{T} \ell_t(\mathbf{w}_t) - \ell_t(\mathbf{u})$. Adversarial online learning algorithms provide robust performance in many complex real-world online prediction problems such as climate or weather forecasting.

In traditional online learning paradigms, the loss for round t is revealed to the learner immediately at the end of round t. However, many real-world applications produce delayed feedback, i.e., the loss for round t is not available until round $t + D$ for some delay period D.[1] Existing delayed online learning algorithms achieve optimal worst-case regret rates against adversarial loss sequences, but each has drawbacks when deployed for real applications with short horizons T. Some use only a small fraction of the data to train each learner (Weinberger & Ordentlich, 2002; Joulani et al., 2013); others tune their parameters using uniform bounds on future gradients that are often challenging to obtain or overly conservative in applications (McMahan & Streeter, 2014; Quanrud & Khashabi, 2015; Joulani et al., 2016; Korotin et al., 2020; Hsieh et al., 2020). Only the concurrent work of Hsieh et al. (2020, Thm. 13) can make use of optimistic hints and only for the special case of unconstrained online gradient descent.

In this work, we aim to develop robust and practical algorithms for real-world delayed online learning. To this end, we introduce three novel algorithms—DORM, DORM+, and AdaHedgeD—that use every observation to train the learner, have no parameters to tune, exhibit optimal worst-case regret rates under delay, *and* enjoy improved performance when accurate hints for unobserved losses are available. We begin by formulating delayed online learning as a special case of optimistic online learning and use this "delay-as-optimism" perspective to develop:

1. A formal reduction of delayed online learning to optimistic online learning (Lems. 1 and 2),
2. The first optimistic tuning-free and self-tuning algorithms with optimal regret guarantees under delay (DORM, DORM+, and AdaHedgeD),
3. A tightening of standard optimistic online learning regret bounds that reveals the robustness of optimistic algorithms to inaccurate hints (Thms. 3 and 4),

[1] Dept. of EECS, Massachusetts Institute of Technology [2] Dept. of AOSE, Woods Hole Oceanographic Institution [3] Dept. of ECE, Boston University [4] Atmospheric and Environmental Research [5] Dept. of CS, University of Toronto [6] Microsoft Research New England [7] Instituto de Matemática Pura e Aplicada. Correspondence to: Genevieve Flaspohler <geflaspo@mit.edu>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021. Copyright 2021 by the author(s).

[1] Our initial presentation will assume constant delay D, but we provide extensions to variable and unbounded delays in App. O.

4. The first general analysis of follow-the-regularized-leader (Thms. 5 and 10) and online mirror descent algorithms (Thm. 6) with optimism and delay, and
5. The first meta-algorithm for learning a low-regret optimism strategy under delay (Thm. 13).

We validate our algorithms on the problem of subseasonal forecasting in Sec. 7. Subseasonal forecasting—predicting precipitation and temperature 2-6 weeks in advance—is a crucial task for allocating water resources and preparing for weather extremes (White et al., 2017). Subseasonal forecasting presents several challenges for online learning algorithms. First, real-time subseasonal forecasting suffers from delayed feedback: multiple forecasts are issued before receiving feedback on the first. Second, the regret horizons are short: a common evaluation period for semimonthly forecasting is one year, resulting in 26 total forecasts. Third, forecasters cannot have difficult-to-tune parameters in real-time, practical deployments. We demonstrate that our algorithms DORM, DORM+, and AdaHedgeD sucessfully overcome these challenges and achieve consistently low regret compared to the best forecasting models.

Our Python library for Optimistic Online Learning under Delay (PoolD) and experiment code are available at https://github.com/geflaspohler/poold.

Notation For integers a, b, we use the shorthand $[b] \triangleq \{1, \dots, b\}$ and $\mathbf{g}_{a:b} \triangleq \sum_{i=a}^{b} \mathbf{g}_i$. We say a function f is *proper* if it is somewhere finite and never $-\infty$. We let $\partial f(\mathbf{w}) = \{\mathbf{g} \in \mathbb{R}^d : f(\mathbf{u}) \geq f(\mathbf{w}) + \langle \mathbf{g}, \mathbf{u} - \mathbf{w} \rangle, \ \forall \mathbf{u} \in \mathbb{R}^d\}$ denote the set of *subgradients* of f at $\mathbf{w} \in \mathbb{R}^d$ and say f is *μ-strongly convex* over a convex set $\mathbf{W} \subseteq \operatorname{int} \operatorname{dom} f$ with respect to $\|\cdot\|$ with dual norm $\|\cdot\|_*$ if $\forall \mathbf{w}, \mathbf{u} \in \mathbf{W}$ and $\mathbf{g} \in \partial f(\mathbf{w})$, we have $f(\mathbf{u}) \geq f(\mathbf{w}) + \langle \mathbf{g}, \mathbf{u} - \mathbf{w} \rangle + \frac{\mu}{2}\|\mathbf{w} - \mathbf{u}\|^2$. For differentiable ψ, we define the Bregman divergence $\mathcal{B}_\psi(\mathbf{w}, \mathbf{u}) \triangleq \psi(\mathbf{w}) - \psi(\mathbf{u}) - \langle \nabla\psi(\mathbf{u}), \mathbf{w} - \mathbf{u} \rangle$. We define $\operatorname{diam}(\mathbf{W}) = \inf_{\mathbf{w}, \mathbf{w}' \in \mathbf{W}} \|\mathbf{w} - \mathbf{w}'\|$, $(r)_+ \triangleq \max(r, 0)$, and $\min(r, s)_+ \triangleq (\min(r, s))_+$.

2. Preliminaries: Optimistic Online Learning

Standard online learning algorithms, such as follow the regularized leader (FTRL) and online mirror descent (OMD) achieve optimal worst-case regret against adversarial loss sequences (Orabona, 2019). However, many loss sequences encountered in applications are not truly adversarial. *Optimistic* online learning algorithms aim to improve performance when loss sequences are partially predictable, while remaining robust to adversarial sequences (see, e.g., Azoury & Warmuth, 2001; Chiang et al., 2012; Rakhlin & Sridharan, 2013b; Steinhardt & Liang, 2014). In optimistic online learning, the learner is provided with a "hint" in the form of a pseudo-loss $\tilde{\ell}_t$ at the start of round t that represents a guess for the true unknown loss. The online learner can incorporate this hint before making play $\mathbf{w}_t$.

In standard formulations of optimistic online learning, the convex pseudo-loss $\tilde{\ell}_t(\mathbf{w}_t)$ is added to the standard FTRL or OMD regularized objective function and leads to optimistic variants of these algorithms: optimistic FTRL (OFTRL, Rakhlin & Sridharan, 2013a) and single-step optimistic OMD (SOOMD, Joulani et al., 2017, Sec. 7.2). Let $\tilde{\mathbf{g}}_t \in \partial\tilde{\ell}_t(\mathbf{w}_{t-1})$ and $\mathbf{g}_t \in \partial\ell_t(\mathbf{w}_t)$ denote subgradients of the pseudo-loss and true loss respectively. The inclusion of an optimistic hint leads to the following linearized update rules for play $\mathbf{w}_{t+1}$:

$$\mathbf{w}_{t+1} = \operatorname*{argmin}_{\mathbf{w} \in \mathbf{W}} \langle \mathbf{g}_{1:t} + \tilde{\mathbf{g}}_{t+1}, \mathbf{w} \rangle + \lambda\psi(\mathbf{w}), \quad \text{(OFTRL)}$$

$$\mathbf{w}_{t+1} = \operatorname*{argmin}_{\mathbf{w} \in \mathbf{W}} \langle \mathbf{g}_t + \tilde{\mathbf{g}}_{t+1} - \tilde{\mathbf{g}}_t, \mathbf{w} \rangle + \mathcal{B}_{\lambda\psi}(\mathbf{w}, \mathbf{w}_t)$$
$$\text{with} \quad \tilde{\mathbf{g}}_0 = \mathbf{0} \quad \text{and arbitrary} \quad \mathbf{w}_0 \quad \text{(SOOMD)}$$

where $\tilde{\mathbf{g}}_{t+1} \in \mathbb{R}^d$ is the hint subgradient, $\lambda \geq 0$ is a regularization parameter, and ψ is proper regularization function that is 1-strongly convex with respect to a norm $\|\cdot\|$. The optimistic learner enjoys reduced regret whenever the hinting error $\|\mathbf{g}_{t+1} - \tilde{\mathbf{g}}_{t+1}\|_*$ is small (Rakhlin & Sridharan, 2013a; Joulani et al., 2017). Common choices of optimistic hints include the last observed subgradient or average of previously observed subgradients (Rakhlin & Sridharan, 2013a). We note that the standard FTRL and OMD updates can be recovered by setting the optimistic hints to zero.

3. Online Learning with Optimism and Delay

In the delayed feedback setting with constant delay of length D, the learner only observes $(\ell_i)_{i=1}^{t-D}$ before making play $\mathbf{w}_{t+1}$. In this setting, we propose counterparts of the OFTRL and SOOMD online learning algorithms, which we call *optimistic delayed FTRL (*ODFTRL*)* and *delayed optimistic online mirror descent (*DOOMD*)* respectively:

$$\mathbf{w}_{t+1} = \operatorname*{argmin}_{\mathbf{w} \in \mathbf{W}} \langle \mathbf{g}_{1:t-D} + \mathbf{h}_{t+1}, \mathbf{w} \rangle + \lambda\psi(\mathbf{w}) \quad \text{(ODFTRL)}$$

$$\mathbf{w}_{t+1} = \operatorname*{argmin}_{\mathbf{w} \in \mathbf{W}} \langle \mathbf{g}_{t-D} + \mathbf{h}_{t+1} - \mathbf{h}_t, \mathbf{w} \rangle + \mathcal{B}_{\lambda\psi}(\mathbf{w}, \mathbf{w}_t)$$
$$\text{with} \quad \mathbf{h}_0 \triangleq \mathbf{0} \quad \text{and arbitrary} \quad \mathbf{w}_0, \quad \text{(DOOMD)}$$

for hint vector $\mathbf{h}_{t+1}$. Our use of the notation $\mathbf{h}_{t+1}$ instead of $\tilde{\mathbf{g}}_{t+1}$ for the optimistic hint here is suggestive. Our regret analysis in Thms. 5 and 6 reveals that, instead of hinting only for the "future" missing loss $\mathbf{g}_{t+1}$, delayed online learners should uses hints $\mathbf{h}_t$ that guess at the summed subgradients of all delayed and future losses: $\mathbf{h}_t = \sum_{s=t-D}^{t} \tilde{\mathbf{g}}_s$.

3.1. Delay as Optimism

To analyze the regret of the ODFTRL and DOOMD algorithms, we make use of the first key insight of this paper:

> *Learning with delay is a special case of learning with optimism.*

In particular, ODFTRL and DOOMD are instances of OFTRL and SOOMD respectively with a particularly "bad" choice of optimistic hint $\tilde{\mathbf{g}}_{t+1}$ that deletes the unobserved loss subgradients $\mathbf{g}_{t-D+1:t}$.

Lemma 1 (ODFTRL is OFTRL with a bad hint)**.** *ODFTRL is* OFTRL *with* $\tilde{\mathbf{g}}_{t+1} = \mathbf{h}_{t+1} - \sum_{s=t-D+1}^{t} \mathbf{g}_s$.

Lemma 2 (DOOMD is SOOMD with a bad hint)**.** *DOOMD is* SOOMD *with* $\tilde{\mathbf{g}}_{t+1} = \tilde{\mathbf{g}}_t + \mathbf{g}_{t-D} - \mathbf{g}_t + \mathbf{h}_{t+1} - \mathbf{h}_t = \mathbf{h}_{t+1} - \sum_{s=t-D+1}^{t} \mathbf{g}_s$.

The implication of this reduction of delayed online learning to optimistic online learning is that *any* regret bound shown for undelayed OFTRL or SOOMD immediately yields a regret bound for ODFTRL and DOOMD under delay. As we demonstrate in the remainder of the paper, this novel connection between delayed and optimistic online learning allows us to bound the regret of optimistic, self-tuning, and tuning-free algorithms for the first time under delay.

Finally, it is worth reflecting on the key property of OFTRL and SOOMD that enables the delay-to-optimism reduction: each algorithm depends on $\mathbf{g}_t$ and $\tilde{\mathbf{g}}_{t+1}$ only through the sum $\mathbf{g}_{1:t} + \tilde{\mathbf{g}}_{t+1}$.[2] For the "bad" hints of Lems. 1 and 2, these sums are observable even though $\mathbf{g}_t$ and $\tilde{\mathbf{g}}_{t+1}$ are not separately observable at time t due to delay. A number of alternatives to SOOMD have been proposed for optimistic OMD (Chiang et al., 2012; Rakhlin & Sridharan, 2013a;b; Kamalaruban, 2016). Unlike SOOMD, these procedures all incorporate optimism in two steps, as in the updates

$$\mathbf{w}_{t+1/2} = \operatorname{argmin}_{\mathbf{w}\in\mathbf{W}} \langle \mathbf{g}_t, \mathbf{w}\rangle + \mathcal{B}_{\lambda\psi}(\mathbf{w}, \mathbf{w}_{t-1/2}) \quad \text{and}$$
$$\mathbf{w}_{t+1} = \operatorname{argmin}_{\mathbf{w}\in\mathbf{W}} \langle \tilde{\mathbf{g}}_{t+1}, \mathbf{w}\rangle + \mathcal{B}_{\lambda\psi}(\mathbf{w}, \mathbf{w}_{t+1/2}) \qquad (1)$$

described in Rakhlin & Sridharan (2013a, Sec. 2.2). It is unclear how to reduce delayed OMD to an instance of one of these two-step procedures, as knowledge of the unobserved $\mathbf{g}_t$ is needed to carry out the first step.

3.2. Delayed and Optimistc Regret Bounds

To demonstrate the utility of our delay-as-optimism perspective, we first present the following new regret bounds for OFTRL and SOOMD, proved in Apps. B and C respectively.

Theorem 3 (OFTRL regret)**.** *If ψ is nonnegative, then, for all $\mathbf{u} \in \mathbf{W}$, the* OFTRL *iterates $\mathbf{w}_t$ satisfy*

$$\text{Regret}_T(\mathbf{u}) \le \lambda\psi(\mathbf{u}) + \tfrac{1}{\lambda}\sum_{t=1}^{T} \text{huber}(\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*, \|\mathbf{g}_t\|_*).$$

Theorem 4 (SOOMD regret)**.** *If ψ is differentiable and $\tilde{\mathbf{g}}_{T+1} \triangleq \mathbf{0}$, then, $\forall \mathbf{u} \in \mathbf{W}$, the* SOOMD *iterates $\mathbf{w}_t$ satisfy*

$$\text{Regret}_T(\mathbf{u}) \le \mathcal{B}_{\lambda\psi}(\mathbf{u}, \mathbf{w}_0) + \tfrac{1}{\lambda}\sum_{t=1}^{T} \text{huber}(\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*, \|\mathbf{g}_t + \tilde{\mathbf{g}}_{t+1} - \tilde{\mathbf{g}}_t\|_*).$$

Both results feature the robust Huber penalty (Huber, 1964)

$$\text{huber}(x, y) \triangleq \tfrac{1}{2}x^2 - \tfrac{1}{2}(|x| - |y|)_+^2 \le \min(\tfrac{1}{2}x^2, |y||x|)$$

in place of the more common squared error term $\frac{1}{2}\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*^2$. As a result, Thms. 3 and 4 strictly improve the rate-optimal OFTRL and SOOMD regret bounds of Rakhlin & Sridharan (2013a); Mohri & Yang (2016); Orabona (2019, Thm. 7.28) and Joulani et al. (2017, Sec. 7.2) by revealing a previously undocumented robustness to inaccurate hints $\tilde{\mathbf{g}}_t$. We will use this robustness to large hint error $\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*$ to establish optimal regret bounds under delay.

As an immediate consequence of this regret analysis and our delay-as-optimism perspective, we obtain the first general analyses of FTRL and OMD with optimism and delay.

Theorem 5 (ODFTRL regret)**.** *If ψ is nonnegative, then, for all $\mathbf{u} \in \mathbf{W}$, the* ODFTRL *iterates $\mathbf{w}_t$ satisfy*

$$\text{Regret}_T(\mathbf{u}) \le \lambda\psi(\mathbf{u}) + \tfrac{1}{\lambda}\sum_{t=1}^{T} \mathbf{b}_{t,F} \quad \textit{for}$$
$$\mathbf{b}_{t,F} \triangleq \text{huber}(\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^{t} \mathbf{g}_s\|_*, \|\mathbf{g}_t\|_*).$$

Theorem 6 (DOOMD regret)**.** *If ψ is differentiable and $\mathbf{h}_{T+1} \triangleq \mathbf{g}_{T-D+1:T}$, then, for all $\mathbf{u} \in \mathbf{W}$, the* DOOMD *iterates $\mathbf{w}_t$ satisfy*

$$\text{Regret}_T(\mathbf{u}) \le \mathcal{B}_{\lambda\psi}(\mathbf{u}, \mathbf{w}_0) + \tfrac{1}{\lambda}\sum_{t=1}^{T} \mathbf{b}_{t,O} \quad \textit{for}$$
$$\mathbf{b}_{t,O} \triangleq \text{huber}(\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^{t} \mathbf{g}_s\|_*, \|\mathbf{g}_{t-D} + \mathbf{h}_{t+1} - \mathbf{h}_t\|_*).$$

Our results show a compounding of regret due to delay: the $\mathbf{b}_{t,F}$ term of Thm. 5 is of size $\mathcal{O}(D+1)$ whenever $\|\mathbf{h}_t\|_* = \mathcal{O}(D+1)$, and the same holds for $\mathbf{b}_{t,O}$ of Thm. 6 if $\|\mathbf{h}_{t+1} - \mathbf{h}_t\|_* = \mathcal{O}(1)$. An optimal setting of λ therefore delivers $\mathcal{O}(\sqrt{(D+1)T})$ regret, yielding the minimax optimal rate for adversarial learning under delay (Weinberger & Ordentlich, 2002). Thms. 5 and 6 also reveal the heightened value of optimism in the presence of delay: in addition to providing an effective guess of the future subgradient $\mathbf{g}_t$, an optimistic hint can approximate the missing delayed feedback ($\sum_{s=t-D}^{t-1} \mathbf{g}_s$) and thereby significantly reduce the penalty of delay. If, on the other hand, the hints are a poor proxy for the missing loss subgradients, the novel huber term ensures that we still only pay the minimax optimal $\sqrt{D+1}$ penalty for delayed feedback.

Related work A classical approach to delayed feedback in online learning is the so-called "replication" strategy in which $D+1$ distinct learners take turns observing and responding to feedback (Weinberger & Ordentlich, 2002;

[2] For SOOMD, $\mathbf{g}_t + \tilde{\mathbf{g}}_{t+1} - \tilde{\mathbf{g}}_t = \mathbf{g}_{1:t} + \tilde{\mathbf{g}}_{t+1} - (\mathbf{g}_{1:t-1} + \tilde{\mathbf{g}}_t)$.

Joulani et al., 2013; Agarwal & Duchi, 2011; Mesterharm, 2005). While minimax optimal in adversarial settings, this strategy has the disadvantage that each learner only sees $\frac{T}{D+1}$ losses and is completely isolated from the other replicates, exacerbating the problem of short prediction horizons. In contrast, we develop and analyze non-replicated delayed online learning strategies that use a combination of optimistic hinting and self-tuned regularization to mitigate the effects of delay while retaining optimal worst-case behavior.

To our knowledge, Thm. 5 and its adaptive generalization Thm. 10 provide the first general analysis of delayed FTRL with optimism, apart from the concurrent work of Hsieh et al. (2020, Thm. 1). Hsieh et al. (2020, Thm. 13) and Quanrud & Khashabi (2015, Thm. 2.1) focus only on delayed gradient descent, Korotin et al. (2020) study General Hedging, and Joulani et al. (2016, Thm. 4) and Quanrud & Khashabi (2015, Thm. A.5) study non-optimistic OMD under delay. Thms. 5, 6, and 10 strengthen these results from the literature which feature a sum of subgradient norms ($\sum_{s=t-D}^{t-1} \|\mathbf{g}_s\|_*$ or $D\|\mathbf{g}_t\|_*$) in place of $\|\mathbf{h}_t - \sum_{s=t-D}^{t-1} \mathbf{g}_s\|_*$. Even in the absence of optimism, the latter can be significantly smaller: e.g., if the gradients $\mathbf{g}_s$ are i.i.d. mean-zero vectors, the former has size $\Omega(D)$ while the latter has expectation $\mathcal{O}(\sqrt{D})$. In the absence of optimism, McMahan & Streeter (2014) obtain a bound comparable to Thm. 5 for the special case of one-dimensional unconstrained online gradient descent.

In the absence of delay, Cutkosky (2019) introduces meta-algorithms for imbuing learning procedures with optimism while remaining robust to inaccurate hints; however, unlike OFTRL and SOOMD, the procedures of Cutkosky require separate observation of $\tilde{\mathbf{g}}_{t+1}$ and each $\mathbf{g}_t$, making them unsuitable for our delay-to-optimism reduction.

3.3. Tuning Regularizers with Optimism and Delay

The online learning algorithms introduced so far all include a regularization parameter λ. In theory and in practice, these algorithms only achieve low regret if the regularization parameter λ is chosen appropriately. In standard FTRL, for example, one such setting that achieves optimal regret is $\lambda = \sqrt{\frac{\sum_{t=1}^{T} \|\mathbf{g}_t\|_*^2}{\sup_{\mathbf{u}\in\mathbf{U}} \psi(\mathbf{u})}}$. This choice, however, cannot be used in practice as it relies on knowledge of all future unobserved loss subgradients. To make use of online learning algorithms, the tuning parameter λ is often set using coarse upper bounds on, e.g., the maximum possible subgradient norm. However, these bounds are often very conservative and lead to poor real-world performance.

In the following sections, we introduce two strategies for tuning regularization with optimism and delay. Sec. 4 introduces the DORM and DORM+ algorithms, variants of ODFTRL and DOOMD that are *entirely tuning-free*. Sec. 5 introduces the AdaHedgeD algorithm, an adaptive variant of ODFTRL that is *self-tuning*; a sequence of regularization parameters λ_t are set automatically using new, tighter bounds on algorithm regret. All three algorithms achieve the minimax optimal regret rate under delay, support optimism, and have strong real-world performance as shown in Sec. 7.

4. Tuning-free Learning with Optimism and Delay

Regret matching (RM) (Blackwell, 1956; Hart & Mas-Colell, 2000) and regret matching+ (RM+) (Tammelin et al., 2015) are online learning algorithms that have strong empirical performance. RM was developed to find correlated equilibria in two-player games and is commonly used to minimize regret over the simplex. RM+ is a modification of RM designed to accelerate convergence and used to effectively solve the game of Heads-up Limit Texas Hold'em poker (Bowling et al., 2015). RM and RM+ support neither optimistic hints nor delayed feedback, and known regret bounds have a suboptimal scaling with respect to the problem dimension d (Cesa-Bianchi & Lugosi, 2006; Orabona & Pál, 2015). To extend these algorithms to the delayed and optimistic setting and recover the optimal regret rate, we introduce our generalizations, *delayed optimistic regret matching* (DORM)

$$\begin{aligned} &\mathbf{w}_{t+1} = \tilde{\mathbf{w}}_{t+1}/\langle \mathbf{1}, \tilde{\mathbf{w}}_{t+1}\rangle \quad \text{for} \qquad\qquad \text{(DORM)} \\ &\tilde{\mathbf{w}}_{t+1} \triangleq \max(\mathbf{0}, (\mathbf{r}_{1:t-D} + \mathbf{h}_{t+1})/\lambda)^{q-1} \end{aligned}$$

and *delayed optimistic regret matching*+ (DORM+)

$$\begin{aligned} &\mathbf{w}_{t+1} = \tilde{\mathbf{w}}_{t+1}/\langle \mathbf{1}, \tilde{\mathbf{w}}_{t+1}\rangle \text{ for } \mathbf{h}_0 = \tilde{\mathbf{w}}_0 \triangleq \mathbf{0}, \quad \text{(DORM+)} \\ &\tilde{\mathbf{w}}_{t+1} \triangleq \max\left(\mathbf{0}, \tilde{\mathbf{w}}_t^{p-1} + (\mathbf{r}_{t-D} + \mathbf{h}_{t+1} - \mathbf{h}_t)/\lambda\right)^{q-1}, \end{aligned}$$

Each algorithm makes use of an instantaneous regret vector $\mathbf{r}_t \triangleq \mathbf{1}\langle \mathbf{g}_t, \mathbf{w}_t\rangle - \mathbf{g}_t$ that quantifies the relative performance of each expert with respect to the play $\mathbf{w}_t$ and the linearized loss subgradient $\mathbf{g}_t$. The updates also include a parameter $q \geq 2$ and its conjugate exponent $p = q/(q-1)$ that is set to recover the minimax optimal scaling of regret with the number of experts (see Cor. 9). We note that DORM and DORM+ recover the standard RM and RM+ algorithms when $D = 0$, $\lambda = 1$, $q = 2$, and $\mathbf{h}_t = \mathbf{0}$, $\forall t$.

4.1. Tuning-free Regret Bounds

To bound the regret of the DORM and DORM+ plays, we prove that DORM is an instance of ODFTRL and DORM+ is an instance of DOOMD. This connection enables us to immediately provide regret guarantees for these regret-matching algorithms under delayed feedback and with optimism. We first highlight a remarkable property of DORM and DORM+ that is the basis of their tuning-free nature. Under mild conditions:

> The normalized DORM and DORM+ iterates $\mathbf{w}_t$ are *independent* of the choice of regularization parameter λ.

Lemma 7 (DORM and DORM+ are independent of λ)**.** *If the subgradient $\mathbf{g}_t$ and hint $\mathbf{h}_{t+1}$ only depend on λ through $(\mathbf{w}_s, \lambda^{q-1}\tilde{\mathbf{w}}_s, \mathbf{g}_{s-1}, \mathbf{h}_s)_{s\leq t}$ and $(\mathbf{w}_s, \lambda^{q-1}\tilde{\mathbf{w}}_s, \mathbf{g}_s, \mathbf{h}_s)_{s\leq t}$ respectively, then the* DORM *and* DORM+ *iterates $(\mathbf{w}_t)_{t\geq 1}$ are independent of the choice of $\lambda > 0$.*

Lem. 7, proved in App. E, implies that DORM and DORM+ are *automatically* optimally tuned with respect to λ, even when run with a default value of $\lambda = 1$. Hence, these algorithms are tuning-free, a very appealing property for real-world deployments of online learning.

To show that DORM and DORM+ also achieve optimal regret scaling under delay, we connect them to ODFTRL and DOOMD operating on the nonnegative orthant with a special surrogate loss $\hat{\ell}_t$ (see App. D for our proof):

Lemma 8 (DORM is ODFTRL and DORM+ is DOOMD)**.** *The* DORM *and* DORM+ *iterates are proportional to* ODFTRL *and* DOOMD *iterates respectively with $\mathbf{W} \triangleq \mathbb{R}^d_+$, $\psi(\tilde{\mathbf{w}}) = \frac{1}{2}\|\tilde{\mathbf{w}}\|_p^2$, and loss $\hat{\ell}_t(\tilde{\mathbf{w}}) = \langle \tilde{\mathbf{w}}, -\mathbf{r}_t\rangle$.*

Lem. 8 enables the following optimally-tuned regret bounds for DORM and DORM+ run with any choice of λ:

Corollary 9 (DORM and DORM+ regret)**.** *Under the assumptions of Lem. 7, for all $\mathbf{u} \in \triangle_{d-1}$ and any choice of $\lambda > 0$, the* DORM *and* DORM+ *iterates $\mathbf{w}_t$ satisfy*

$$\begin{aligned}\text{Regret}_T(\mathbf{u}) &\leq \inf_{\lambda>0} \tfrac{\lambda}{2}\|\mathbf{u}\|_p^2 + \tfrac{1}{\lambda(p-1)} \textstyle\sum_{t=1}^T \mathbf{b}_{t,q} \\ &= \sqrt{\tfrac{\|\mathbf{u}\|_p^2}{2(p-1)} \textstyle\sum_{t=1}^T \mathbf{b}_{t,q}} \leq \sqrt{\tfrac{d^{2/q}(q-1)}{2} \textstyle\sum_{t=1}^T \mathbf{b}_{t,\infty}}\end{aligned}$$

where $\mathbf{h}_{T+1} \triangleq \mathbf{r}_{T-D+1:T}$ and, for each $c \in [2, \infty]$,

$$\begin{aligned}\mathbf{b}_{t,c} &\overset{\text{(DORM)}}{=} \text{huber}(\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^t \mathbf{r}_s\|_c, \|\mathbf{r}_t\|_c) \quad \textit{and} \\ \mathbf{b}_{t,c} &\overset{\text{(DORM+)}}{=} \text{huber}(\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^t \mathbf{r}_s\|_c^2, \\ &\qquad\qquad \|\mathbf{r}_{t-D} + \mathbf{h}_{t+1} - \mathbf{h}_t\|_c).\end{aligned}$$

If, in addition, $q = \text{argmin}_{q'\geq 2} d^{2/q'}(q' - 1)$, then $\text{Regret}_T(\mathbf{u}) \leq \sqrt{(2\log_2(d) - 1)\sum_{t=1}^T \mathbf{b}_{t,\infty}}$.

Cor. 9, proved in App. F, suggests a natural hinting strategy for reducing the regret of DORM and DORM+: predict the sum of unobserved instantaneous regrets $\sum_{s=t-D}^t \mathbf{r}_s$. We explore this strategy empirically in Sec. 7. Cor. 9 also highlights the value of the q parameter in DORM and DORM+: using the easily computed value $q = \text{argmin}_{q'\geq 2} d^{2/q'}(q' - 1)$ yields the minimax optimal $\sqrt{\log_2(d)}$ dependence of regret on dimension (Cesa-Bianchi & Lugosi, 2006; Orabona & Pál, 2015). By Lem. 8, setting q in this way is equivalent to selecting a robust $\frac{1}{2}\|\cdot\|_p^2$ regularizer (Gentile, 2003) for the underlying ODFTRL and DOOMD problems.

Related work Without delay, Farina et al. (2021) independently developed optimistic versions of RM and RM+ by reducing them to OFTRL and a two-step variant of optimistic OMD (1). Unlike SOOMD, this two-step optimistic OMD requires separate observation of $\tilde{\mathbf{g}}_{t+1}$ and $\mathbf{g}_t$, making it unsuitable for our delay-as-optimism reduction and resulting in a different algorithm from DORM+ even when $D = 0$. In addition, their regret bounds and prior bounds for RM and RM+ (special cases of DORM and DORM+ with $q = 2$) have suboptimal regret when the dimension d is large (Bowling et al., 2015; Zinkevich et al., 2007).

5. Self-tuned Learning with Optimism and Delay

In this section, we analyze an adaptive version of ODFTRL with time-varying regularization $\lambda_t\psi$ and develop strategies for setting λ_t appropriately in the presence of optimism and delay. We begin with a new general regret analysis of optimistic delayed *adaptive* FTRL (ODAFTRL)

$$\mathbf{w}_{t+1} = \operatorname*{argmin}_{\mathbf{w}\in\mathbf{W}} \langle \mathbf{g}_{1:t-D} + \mathbf{h}_{t+1}, \mathbf{w}\rangle + \lambda_{t+1}\psi(\mathbf{w}) \tag{ODAFTRL}$$

where $\mathbf{h}_{t+1} \in \mathbb{R}^d$ is an arbitrary hint vector revealed before $\mathbf{w}_{t+1}$ is generated, ψ is 1-strongly convex with respect to a norm $\|\cdot\|$, and $\lambda_t \geq 0$ is a regularization parameter.

Theorem 10 (ODAFTRL regret)**.** *If ψ is nonnegative and λ_t is non-decreasing in t, then, $\forall\mathbf{u} \in \mathbf{W}$, the* ODAFTRL *iterates $\mathbf{w}_t$ satisfy*

$$\begin{aligned}&\text{Regret}_T(\mathbf{u}) \leq \lambda_T\psi(\mathbf{u}) + \textstyle\sum_{t=1}^T \min(\frac{\mathbf{b}_{t,F}}{\lambda_t}, \mathbf{a}_{t,F}) \quad \textit{with} \\ &\quad \mathbf{b}_{t,F} \triangleq \text{huber}(\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^t \mathbf{g}_s\|_*, \|\mathbf{g}_t\|_*) \quad \textit{and} \\ &\quad \mathbf{a}_{t,F} \triangleq \text{diam}(\mathbf{W}) \min\left(\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^t \mathbf{g}_s\|_*, \|\mathbf{g}_t\|_*\right).\end{aligned} \tag{2}$$

The proof of this result in App. G builds on a new regret bound for undelayed optimistic adaptive FTRL (OAFTRL). In the absence of delay ($D = 0$), Thm. 10 strictly improves existing regret bounds (Rakhlin & Sridharan, 2013a; Mohri & Yang, 2016; Joulani et al., 2017) for OAFTRL by providing tighter guarantees whenever the hinting error $\|\mathbf{h}_t - \sum_{s=t-D}^t \mathbf{g}_t\|_*$ is larger than the subgradient magnitude $\|\mathbf{g}_t\|_*$. In the presence of delay, Thm. 10 benefits both from robustness to hinting error in the worst case and the ability to exploit accurate hints in the best case. The bounded-domain factors $\mathbf{a}_{t,F}$ strengthen both standard OAFTRL regret bounds and the concurrent bound of Hsieh et al. (2020, Thm. 1) when $\text{diam}(\mathbf{W})$ is small and will enable us to design practical λ_t-tuning strategies under delay without any prior knowledge of unobserved subgradients. We now turn to these self-tuning protocols.

5.1. Conservative Tuning with Delayed Upper Bound

Setting aside the $\mathbf{a}_{t,F}$ bounded-domain factors in Thm. 10 for now, the adaptive sequence $\lambda_t = \sqrt{\frac{\sum_{s=1}^{t} \mathbf{b}_{s,F}}{\sup_{\mathbf{u}\in\mathbf{U}} \psi(\mathbf{u})}}$ is known to be a near-optimal minimizer of the ODAFTRL regret bound (McMahan, 2017, Lemma 1). However, this value is unobservable at time t. A common strategy is to play the conservative value $\lambda_t = \sqrt{\frac{(D+1)B_0+\sum_{s=1}^{t-D-1} \mathbf{b}_{s,F}}{\sup_{\mathbf{u}\in\mathbf{U}} \psi(\mathbf{u})}}$, where B_0 is a uniform upper bound on the unobserved $\mathbf{b}_{s,F}$ terms (Joulani et al., 2016; McMahan & Streeter, 2014). In practice, this requires computing an *a priori* upper bound on any subgradient norm that could possibly arise and often leads to extreme over-regularization (see Sec. 7).

As a preliminary step towards fully adaptive settings of λ_t, we analyze in App. H a new *delayed upper bound* (DUB) tuning strategy which relies only on observed $\mathbf{b}_{s,F}$ terms and does not require upper bounds for future losses.

Theorem 11 (DUB regret)**.** *Fix $\alpha > 0$, and, for $\mathbf{a}_{t,F}, \mathbf{b}_{t,F}$ as in* (2)*, consider the* delayed upper bound *(*DUB*) sequence*

$$\lambda_{t+1} = \tfrac{2}{\alpha} \max_{j\le t-D-1} \mathbf{a}_{j-D+1:j,F} + \tfrac{1}{\alpha}\sqrt{\textstyle\sum_{i=1}^{t-D} \mathbf{a}_{i,F}^2 + 2\alpha \mathbf{b}_{i,F}}. \qquad \text{(DUB)}$$

If ψ is nonnegative, then, for all $\mathbf{u} \in \mathbf{W}$, the ODAFTRL *iterates $\mathbf{w}_t$ satisfy*

$$\text{Regret}_T(\mathbf{u}) \le \big(\tfrac{\psi(\mathbf{u})}{\alpha} + 1\big) \big(2 \max_{t\in[T]} \mathbf{a}_{t-D:t-1,F} + \sqrt{\textstyle\sum_{t=1}^{T} \mathbf{a}_{t,F}^2 + 2\alpha \mathbf{b}_{t,F}}\big).$$

As desired, the DUB setting of λ_t depends only on previously observed $\mathbf{a}_{t,F}$ and $\mathbf{b}_{t,F}$ terms and achieves optimal regret scaling with the delay period D. However, the terms $\mathbf{a}_{t,F}$, $\mathbf{b}_{t,F}$ are themselves potentially loose upper bounds for the instantaneous regret at time t. In the following section, we show how the DUB regularization setting can be refined further to produce AdaHedgeD adaptive regularization.

5.2. Refined Tuning with AdaHedgeD

As noted by Erven et al. (2011); de Rooij et al. (2014); Orabona (2019), the effectiveness of an adaptive regularization setting λ_t that uses an upper bound on regret (such as $\mathbf{b}_{t,F}$) relies heavily on the tightness of that bound. In practice, we want to set λ_t using as tight a bound as possible. Our next result introduces a new tuning sequence that can be used with delayed feedback and is inspired by the popular AdaHedge algorithm (Erven et al., 2011). It makes use of the tightened regret analysis underlying Thm. 10 to enable tighter settings of λ_t compared to DUB, while still controlling algorithm regret (see proof in App. I).

Theorem 12 (AdaHedgeD regret)**.** *Fix $\alpha > 0$, and consider the* delayed AdaHedge-style *(*AdaHedgeD*) sequence*

$$\lambda_{t+1} = \tfrac{1}{\alpha} \textstyle\sum_{s=1}^{t-D} \delta_s \qquad \textit{for} \qquad \text{(AdaHedgeD)}$$

$$\delta_t \triangleq \min(F_{t+1}(\mathbf{w}_t, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t),\ \langle \mathbf{g}_t, \mathbf{w}_t - \bar{\mathbf{w}}_t\rangle, F_{t+1}(\hat{\mathbf{w}}_t, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t) + \langle \mathbf{g}_t, \mathbf{w}_t - \hat{\mathbf{w}}_t\rangle)_+$$

$$\textit{with} \quad \bar{\mathbf{w}}_t \triangleq \text{argmin}_{\mathbf{w}\in\mathbf{W}} F_{t+1}(\mathbf{w}, \lambda_t), \qquad (3)$$

$$\hat{\mathbf{w}}_t \triangleq \text{argmin}_{\mathbf{w}\in\mathbf{W}} F_{t+1}(\mathbf{w}, \lambda_t) + \min\big(\tfrac{\|\mathbf{g}_t\|_*}{\|\mathbf{h}_t - \mathbf{g}_{t-D:t}\|_*}, 1\big)\langle \mathbf{h}_t - \mathbf{g}_{t-D:t}, \mathbf{w}\rangle,$$

$$\textit{and} \quad F_{t+1}(\mathbf{w}, \lambda_t) \triangleq \lambda_t \psi(\mathbf{w}) + \langle \mathbf{g}_{1:t}, \mathbf{w}\rangle.$$

If ψ is nonnegative, then, for all $\mathbf{u} \in \mathbf{W}$, the ODAFTRL *iterates satisfy*

$$\text{Regret}_T(\mathbf{u}) \le \big(\tfrac{\psi(\mathbf{u})}{\alpha} + 1\big) \big(2 \max_{t\in[T]} \mathbf{a}_{t-D:t-1,F} + \sqrt{\textstyle\sum_{t=1}^{T} \mathbf{a}_{t,F}^2 + 2\alpha \mathbf{b}_{t,F}}\big).$$

Remarkably, Thm. 12 yields a minimax optimal $\mathcal{O}(\sqrt{(D+1)T} + D)$ dependence on the delay parameter and nearly matches the Thm. 5 regret of the optimal constant λ tuning. Although this regret bound is identical to that in Thm. 11, in practice the λ_t values produced by AdaHedgeD can be orders of magnitude smaller than those of DUB, granting additional adaptivity. We evaluate the practical implications of these λ_t settings in Sec. 7.

As a final note, when ψ is bounded on $\mathbf{U}$, we recommend choosing $\alpha = \sup_{\mathbf{u}\in\mathbf{U}} \psi(\mathbf{u})$ so that $\frac{\psi(\mathbf{u})}{\alpha} \le 1$. For negative entropy regularization $\psi(\mathbf{u}) = \sum_{j=1}^{d} \mathbf{u}_j \ln(\mathbf{u}_j) + \ln(d)$ on the simplex $\mathbf{U} = \mathbf{W} = \triangle_{d-1}$, this yields $\alpha = \ln(d)$ and a regret bound with minimax optimal $\sqrt{\ln(d)}$ dependence on d (Cesa-Bianchi & Lugosi, 2006; Orabona & Pál, 2015).

Related work Our AdaHedgeD δ_t terms differ from standard AdaHedge increments (see, e.g., Orabona, 2019, Sec. 7.6) due to the accommodation of delay, the incorporation of optimism, and the inclusion of the final two terms in the $\min$. These non-standard terms are central to reducing the impact of delay on our regret bounds. Prior and concurrent approaches to adaptive tuning under delay do not incorporate optimism and require an explicit upper bound on all future subgradient norms, a quantity which is often difficult to obtain or very loose (McMahan & Streeter, 2014; Joulani et al., 2016; Hsieh et al., 2020). Our optimistic algorithms, DUB and AdaHedgeD, admit comparable regret guarantees (Thms. 11 and 12) but require no prior knowledge of future subgradients.

6. Learning to Hint with Delay

As we have seen, optimistic hints play an important role in online learning under delay: effective hinting can counteract the increase in regret under delay. In this section, we consider the problem of choosing amongst several competing hinting strategies. We show that this problem can again be treated as a delayed online learning problem. In the following, we will call the original online learning problem the "base problem" and the learning-to-hint problem the "hinting problem."

Suppose that, at time t, we observe the hints $\tilde{\mathbf{g}}_t$ of m different hinters arranged into a $d \times m$ matrix H_t. Each column of H_t is one hinter's best estimate of the sum of missing loss subgradients $\mathbf{g}_{t-D:t}$. Our aim is to output a sequence of combined hints $\mathbf{h}_t(\omega_t) \triangleq H_t\omega_t$ with low regret relative to the best constant combination strategy $\omega \in \Omega \triangleq \triangle_{m-1}$ in hindsight. To achieve this using delayed online learning, we make use of a convex loss function $l_t(\omega)$ for the hint learner that upper bounds the base learner regret.

Assumption 1 (Convex regret bound). *For any hint sequence $(\mathbf{h}_t)_{t=1}^T$ and $\mathbf{u} \in \Omega$, the base problem admits the regret bound* $\text{Regret}_T(\mathbf{u}) \le C_0(\mathbf{u}) + C_1(\mathbf{u})\sqrt{\sum_{t=1}^T f_t(\mathbf{h}_t)}$ *for $C_1(\mathbf{u}) \ge 0$ and convex functions f_t independent of* $\mathbf{u}$.

As we detail in App. K, Assump. 1 holds for all of the learning algorithms introduced in this paper. For example, by Cor. 9, if the base learner is DORM, we may choose $C_0(\mathbf{u}) = 0$, $C_1(\mathbf{u}) = \sqrt{\frac{\|\mathbf{u}\|_p^2}{2(p-1)}}$, and the $\mathcal{O}(D)$ convex function $f_t(\mathbf{h}_t) = \|\mathbf{r}_t\|_q \|\mathbf{h}_t - \sum_{s=t-D}^t \mathbf{r}_s\|_q \ge \mathbf{b}_{t,q}$.[3]

For any base learner satisfying Assump. 1, we choose $l_t(\omega) = f_t(H_t\omega)$ as our hinting loss, use the tuning-free DORM+ algorithm to output the combination weights ω_t on each round, and provide the hint $\mathbf{h}_t(\omega_t) = H_t\omega_t$ to the base learner. The following result, proved in App. J, shows that this learning to hint strategy performs nearly as well as the best constant hint combination strategy in restrospect.

Theorem 13 (Learning to hint regret). *Suppose the base problem satisfies Assump. 1 and the hinting problem is solved with* DORM+ *hint iterates ω_t, hinting losses $l_t(\omega) = f_t(H_t\omega)$, no meta-hints for the hinting problem, and $q =$* $\operatorname{argmin}_{q' \ge 2} m^{2/q'}(q'-1)$. *Then the base problem with hints $\mathbf{h}_t(\omega_t) = H_t\omega_t$ satisfies*

$$\text{Regret}_T(\mathbf{u}) \le C_0(\mathbf{u}) + C_1(\mathbf{u})\sqrt{\inf_{\omega\in\Omega} \sum_{t=1}^T f_t(\mathbf{h}_t(\omega))}$$
$$+ C_1(\mathbf{u})\big((2\log_2(m) - 1)(\tfrac{1}{2}\xi_T + \sum_{t=1}^{T-1} \text{huber}(\xi_t, \zeta_t))\big)^{1/4}$$
$$\textit{for} \quad \xi_t \triangleq 4(D+1)\sum_{s=t-D}^t \|\gamma_s\|_\infty^2, \quad \gamma_t \in \partial l_t(\omega_t),$$
$$\textit{and} \quad \zeta_t \triangleq 4\|\gamma_{t-D}\|_\infty \sum_{s=t-D}^t \|\gamma_s\|_\infty.$$

To quantify the size of this regret bound, consider again the DORM base learner with $f_t(\mathbf{h}_t) = \|\mathbf{r}_t\|_q \|\mathbf{h}_t - \sum_{s=t-D}^t \mathbf{r}_s\|_q$. By Lem. 26 in App. K, $\|\gamma_t\|_\infty \le d^{1/q}\|H_t\|_\infty \|\mathbf{r}_t\|_q$ for $\|H_t\|_\infty$ the maximum absolute entry of H_t. Each column of H_t is a sum $D+1$ subgradient hints, so $\|H_t\|_\infty$ is $\mathcal{O}(D+1)$. Thus, for this choice of hinter loss, the huber(ξ_t, ζ_t) term is $\mathcal{O}((D+1)^3)$, and the hint learner suffers only $\mathcal{O}(T^{1/4}(D+1)^{3/4})$ additional regret from learning to hint. Notably, this additive regret penalty is $\mathcal{O}(\sqrt{(D+1)T})$ if $D = \mathcal{O}(T)$ (and $o(\sqrt{(D+1)T})$ when $D = o(T)$), so the learning to hint strategy of Thm. 13 preserves minimax optimal regret rates.

Related work Rakhlin & Sridharan (2013a, Sec. 4.1) propose and analyze a method to learn optimism strategies for a two-step OMD base learner. Unlike Thm. 13, the approach does not accommodate delay, and the analyzed regret is only with respect to single hinting strategies $\omega \in \{\mathbf{e}_j\}_{j\in[m]}$ rather than combination strategies, $\omega \in \triangle_{m-1}$.

7. Experiments

We now apply the online learning techniques developed in this paper to the problem of adaptive ensembling for subseasonal forecasting. Our experiments are based on the subseasonal forecasting data of Flaspohler et al. (2021) that provides the forecasts of $d = 6$ machine learning and physics-based models for both temperature and precipitation at two forecast horizons: 3-4 weeks and 5-6 weeks. In operational subseasonal forecasting, feedback is delayed; models make $D = 2$ or 3 forecasts (depending on the forecast horizon) before receiving feedback. We use delayed, optimistic online learning to play a time-varying convex combination of input models and compete with the best input model over a year-long prediction period ($T = 26$ semimonthly dates). The loss function is the geographic root-mean squared error (RMSE) across 514 locations in the Western United States.

We evaluate the relative merits of the delayed online learning techniques presented by computing yearly regret and mean RMSE for the ensemble plays made by the online leaner in each year from 2011-2020. Unless otherwise specified, all online learning algorithms use the `recent_g` hint $\tilde{\mathbf{g}}_s$, which approximates each unobserved subgradient at time t with the most recent observed subgradient $\mathbf{g}_{t-D-1}$. See App. L for full experimental details, App. N for algorithmic details, and App. M for extended experimental results.

Competing with the best input model The primary benefit of online learning in this setting is its ability to achieve small average regret, i.e., to perform nearly as well as the best input model in the competitor set $\mathbf{U}$ without knowing which is best in advance. We run our three delayed online learners—DORM, DORM+, and AdaHedgeD—on all four subseasonal prediction tasks and measure their average loss.

The average yearly RMSE for the three online learning algorithms and the six input models is shown in Table 1. The DORM+ algorithm tracks the performance of the best input model for all tasks except Temp. 5-6w. All online learning

[3] The alternative choice $f_t(\mathbf{h}_t) = \frac{1}{2}\|\mathbf{h}_t - \sum_{s=t-D}^t \mathbf{g}_s\|_q^2$ also bounds regret but may have size $\Theta(D^2)$ rather than $\mathcal{O}(D)$.

Table 1: **Average RMSE of the 2011-2020 semimonthly forecasts**: The average RMSE for online learning algorithms (left) and input models (right) over a 10-year evaluation period with the top-performing learners and input models bolded and blue. In each task, the online learners compare favorably with the best input model and learn to downweight the lower-performing candidates, like the worst models italicized in red.

	AdaHedgeD	DORM	DORM+	Model1	Model2	Model3	Model4	Model5	Model6
Precip. 3-4w	21.726	21.731	**21.675**	**21.973**	22.431	22.357	21.978	21.986	*23.344*
Precip. 5-6w	21.868	21.957	**21.838**	22.030	22.570	22.383	22.004	**21.993**	*23.257*
Temp. 3-4w	2.273	2.259	**2.247**	**2.253**	2.352	2.394	2.277	2.319	*2.508*
Temp. 5-6w	2.316	2.316	**2.303**	**2.270**	2.368	2.459	2.278	2.317	*2.569*

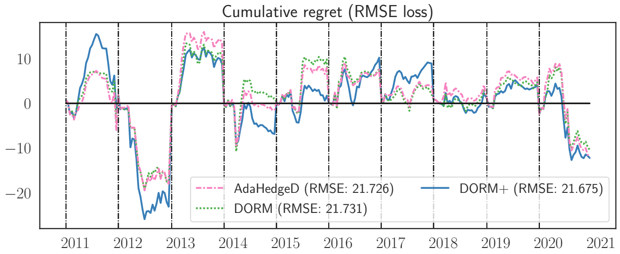

Figure 1: **Overall performance**: Yearly cumulative regret under RMSE loss for the the Precip. 3-4w task. The zero line corresponds to the performance of the best input model in a given year.

algorithms achieve negative regret for both precipitation tasks. Fig. 1 shows the yearly cumulative regret (in terms of the RMSE loss) of the online learning algorithms over the 10-year evaluation period. There are several years (e.g., 2012, 2014, 2020) in which all online learning algorithms substantially outperform the best input forecasting model. The consistently low regret year-to-year of DORM+ compared to DORM and AdaHedgeD makes it a promising candidate for real-world delayed subseasonal forecasting. Notably, RM+ (a special case of DORM+) is known to have small *tracking regret*, i.e., it competes well even with strategies that switch between input models a bounded number of times (Tammelin et al., 2015, Thm. 2). We suspect that this is one source of DORM+'s superior performance. We also note that the self-tuned AdaHedgeD performs comparably to the the optimally-tuned DORM, demonstrating the effectiveness of our self-tuning strategy.

Impact of regularization We evaluate the impact of the three regularization strategies developed in this paper: 1) the upper bound DUB strategy, 2) the tighter AdaHedgeD strategy, and 3) the DORM+ algorithm that is tuning-free. This tuning-free property has evident practical benefits, as this section demonstrates.

Fig. 2 shows the yearly regret of the DUB, AdaHedgeD, and DORM+ algorithms. A consistent pattern appears in the yearly regret: DUB has moderate positive regret, AdaHedgeD has both the largest positive and negative regret values, and DORM+ sits between these two extremes. If we examine the weights played by each algorithm (Fig. 3), the weights of DUB and AdaHedgeD appear respectively over- and under-regularized compared to DORM+ (the top model for this task). DUB's use of the upper bound $\mathbf{b}_{t,F}$ results in a very large regularization setting ($\lambda_T = 142.881$) and a virtually uniform weight setting. AdaHedgeD's tighter bound δ_t produces a value for $\lambda_T = 3.005$ that is two orders of magnitude smaller. However, in this short-horizon forecasting setting, AdaHedgeD's aggressive plays result in higher average RMSE. By nature of it's λ_t-free updates, DORM+ produces more moderately regularized plays $\mathbf{w}_t$ and negative regret.

To replicate or not to replicate In this section, we compare the performance of replicated and non-replicated variants of our DORM+ algorithm. Both algorithms perform well (see App. M.3), but in all tasks, DORM+ outperforms replicated DORM+ (in which $D+1$ independent copies of DORM+ make staggered predictions). Fig. 4 provides an example of the weight plots produced by the replication strategy in the Temp. 5-6w task with $D = 3$. The separate nature of the replicated learner's plays is evident in the weight plots and leads to an average RMSE of 2.315, versus 2.303 for DORM+ in the Temp. 5-6w task.

Learning to hint Finally, we examine the effect of optimism on the DORM+ algorithms and the ability of our "learning to hint" strategy to recover the performance of the best optimism strategy in retrospect. Following the hint construction protocol in App. N.2, we run the DORM+ base algorithm with $m = 4$ subgradient hinting strategies: $\tilde{\mathbf{g}}_s = \mathbf{g}_{t-D-1}$ (`recent_g`), $\tilde{\mathbf{g}}_s = \mathbf{g}_{s-D-1}$ (`prev_g`),

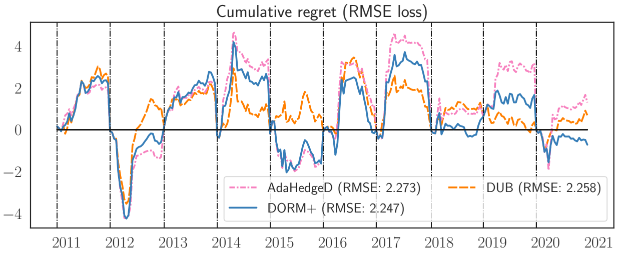

Figure 2: **Regret of regularizers**: Yearly cumulative regret (in terms of the RMSE loss) for the three regularization strategies for the Temp. 3-4w task.

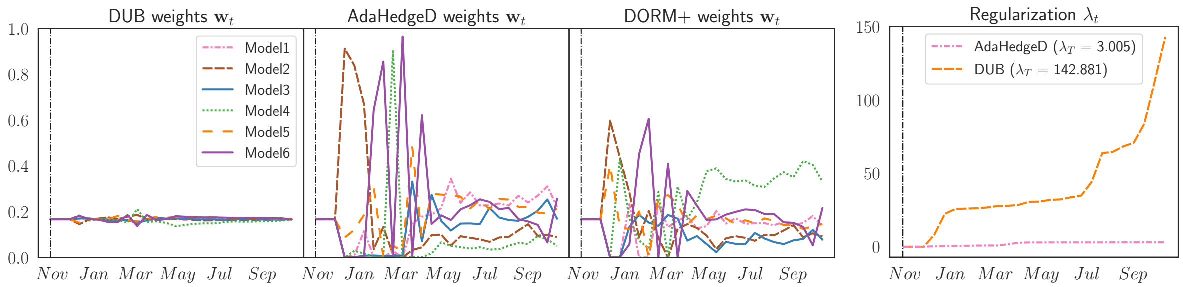

Figure 3: **Impact of regularization:** The plays $\mathbf{w}_t$ of online learning algorithms used to combine the input models for the Temp. 3-4w task in the 2020 evaluation year. The weights of DUB and AdaHedgeD appear respectively over and under regularized compared to DORM+ (the top model for this task) due to their selection of regularization strength λ_t (right).

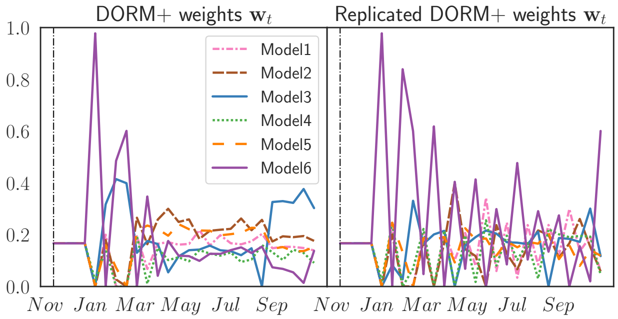

Figure 4: **To replicate or not to replicate**: The plays $\mathbf{w}_t$ of standard DORM+ and replicated DORM+ algorithms for the Temp. 5-6w task in the final evaluation year.

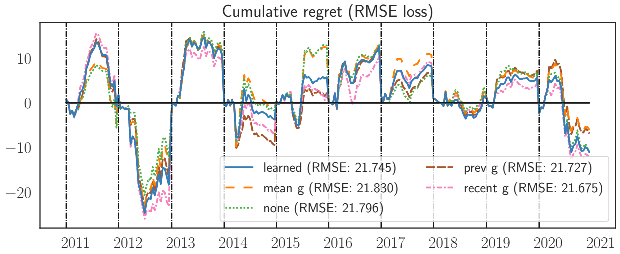

Figure 5: **Learning to hint**: Yearly cumulative regret (in terms of the RMSE loss) for the adaptive hinting and four constant hinting strategies for the Precip. 3-4w task.

$\tilde{\mathbf{g}}_s = \frac{D+1}{t-D-1}\mathbf{g}_{1:t-D-1}$ (mean_g), or $\tilde{\mathbf{g}}_s = \mathbf{0}$ (none). We also use DORM+ as the meta-algorithm for hint learning to produce the learned optimism strategy that plays a convex combination of the four hinters. In Fig. 5, we first note that several optimism strategies outperform the none hinter, confirming the value of optimism in reducing regret. The learned variant of DORM+ avoids the worst-case performance of the individual hinters in any given year (e.g., 2015), while staying competitive with the best strategy (although it does not outperform the dominant recent_g strategy overall). We believe the performance of the online hinter could be further improved by developing tighter convex bounds on the regret of the base problem in the spirit of Assump. 1.

8. Conclusion

In this work, we confronted the challenges of delayed feedback and short regret horizons in online learning with optimism, developing practical non-replicated, self-tuned and tuning-free algorithms with optimal regret guarantees. Our "delay as optimism" reduction and our refined analysis of optimistic learning produced novel regret bounds for both optimistic and delayed online learning and elucidated the connections between these two problems. Within the subseasonal forecasting domain, we demonstrated that delayed online learning methods can produce zero-regret forecast ensembles that perform robustly from year-to-year. Our results highlighted DORM+ as a particularly promising candidate due to its tuning-free nature and small tracking regret.

In future work, we are excited to further develop optimism strategies under delay by 1) employing tighter convex loss bounds on the regret of the base algorithm to improve the learning to hint algorithm, 2) exploring the relative impact of hinting for "past" ($\mathbf{g}_{t-D:t-1}$) versus "future" ($\mathbf{g}_t$) missing subgradients (see App. M.5 for an initial exploration), and 3) developing adaptive self-tuning variants of the DOOMD algorithm. Within the subseasonal domain, we plan to leverage the flexibility of our optimism formulation to explore hinting strategies that use meteorological expertise to improve beyond the generic mean and past subgradient hints and to deploy our open-source subseasonal forecasting algorithms operationally.

Acknowledgements

This work was supported by Microsoft AI for Earth, an NSF GRFP, and the NSF grants no. 1925930 "Collaborative Research: TRIPODS Institute for Optimization and Learning", no. 1908111 "AF: Small: Collaborative Research: New Representations for Learning Algorithms and Secure Computation", and no. 2022446 "Foundations of Data Science Institute". FO also thanks Nicolò Cesa-Bianchi and Christian Kroer for discussions on RM and RM+.

References

Agarwal, A. and Duchi, J. C. Distributed delayed stochastic optimization. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

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

Blackwell, D. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956.

Bowling, M., Burch, N., Johanson, M., and Tammelin, O. Heads-up limit hold'em poker is solved. *Science*, 347(6218):145–149, 2015. ISSN 0036-8075. doi: 10.1126/science.1259433.

Cesa-Bianchi, N. and Lugosi, G. *Prediction, learning, and games*. Cambridge university press, 2006.

Chiang, C.-K., Yang, T., Lee, C.-J., Mahdavi, M., Lu, C.-J., Jin, R., and Zhu, S. Online optimization with gradual variations. In Mannor, S., Srebro, N., and Williamson, R. C. (eds.), *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23, pp. 6.1–6.20, Edinburgh, Scotland, 25–27 Jun 2012.

Cutkosky, A. Combining online learning guarantees. In Beygelzimer, A. and Hsu, D. (eds.), *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 895–913, Phoenix, USA, 25–28 Jun 2019. PMLR.

Danskin, J. M. *The theory of max-min and its application to weapons allocation problems*, volume 5. Springer Science & Business Media, 2012.

de Rooij, S., van Erven, T., Grünwald, P. D., and Koolen, W. M. Follow the leader if you can, hedge if you must. *Journal of Machine Learning Research*, 15(37):1281–1316, 2014.

Erven, T., Koolen, W. M., Rooij, S., and Grünwald, P. Adaptive hedge. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, volume 24, pp. 1656–1664. Curran Associates, Inc., 2011.

Farina, G., Kroer, C., and Sandholm, T. Faster game solving via predictive blackwell approachability: Connecting regret matching and mirror descent. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(6):5363–5371, May 2021.

Flaspohler, G., Orabona, F., Cohen, J., Mouatadid, S., Oprescu, M., Orenstein, P., and Mackey, L. Replication Data for: Online Learning with Optimism and Delay, 2021. URL https://doi.org/10.7910/DVN/IOCFCY.

Gentile, C. The robustness of the p-norm algorithms. *Machine Learning*, 53(3):265–299, 2003.

Hart, S. and Mas-Colell, A. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

Hsieh, Y.-G., Iutzeler, F., Malick, J., and Mertikopoulos, P. Multi-agent online optimization with delays: Asynchronicity, adaptivity, and optimism. *arXiv preprint arXiv:2012.11579*, 2020.

Huber, P. J. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101, 1964. doi: 10.1214/aoms/1177703732. URL https://doi.org/10.1214/aoms/1177703732.

Hwang, J., Orenstein, P., Cohen, J., Pfeiffer, K., and Mackey, L. Improving subseasonal forecasting in the western U.S. with machine learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2325–2335, 2019.

Joulani, P., Gyorgy, A., and Szepesvári, C. Online learning under delayed feedback. In *International Conference on Machine Learning*, pp. 1453–1461, 2013.

Joulani, P., Gyorgy, A., and Szepesvári, C. Delay-tolerant online convex optimization: Unified analysis and adaptive-gradient algorithms. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

Joulani, P., György, A., and Szepesvári, C. A modular analysis of adaptive (non-) convex optimization: Optimism, composite objectives, and variational bounds. In *International Conference on Algorithmic Learning Theory*, pp. 681–720. PMLR, 2017.

Kamalaruban, P. Improved optimistic mirror descent for sparsity and curvature. *arXiv preprint arXiv:1609.02383*, 2016.

Koolen, W., Van Erven, T., and Grunwald, P. Learning the learning rate for prediction with expert advice. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, volume 27, pp. 2294–2302. Curran Associates, Inc., 2014.

Korotin, A., V'yugin, V., and Burnaev, E. Adaptive hedging under delayed feedback. *Neurocomputing*, 397:356–368, 2020.

Liu, J. and Wright, S. J. Asynchronous stochastic coordinate descent: Parallelism and convergence properties. *SIAM Journal on Optimization*, 25(1):351–376, 2015.

Liu, J., Wright, S., Ré, C., Bittorf, V., and Sridhar, S. An asynchronous parallel stochastic coordinate descent algorithm. In *International Conference on Machine Learning*, pp. 469–477. PMLR, 2014.

McMahan, B. and Streeter, M. Delay-tolerant algorithms for asynchronous distributed online learning. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, volume 27, pp. 2915–2923. Curran Associates, Inc., 2014.

McMahan, H. B. A survey of algorithms and analysis for adaptive online learning. *The Journal of Machine Learning Research*, 18 (1):3117–3166, 2017.

McQuade, S. and Monteleoni, C. Global climate model tracking using geospatial neighborhoods. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26, 2012.

Mesterharm, C. On-line learning with delayed label feedback. In *International Conference on Algorithmic Learning Theory*, pp. 399–413. Springer, 2005.

Mohri, M. and Yang, S. Accelerating online convex optimization via adaptive prediction. In *Artificial Intelligence and Statistics*, pp. 848–856. PMLR, 2016.

Monteleoni, C. and Jaakkola, T. Online learning of non-stationary sequences. In Thrun, S., Saul, L., and Schölkopf, B. (eds.), *Advances in Neural Information Processing Systems*, volume 16, pp. 1093–1100. MIT Press, 2004.

Monteleoni, C., Schmidt, G. A., Saroha, S., and Asplund, E. Tracking climate models. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 4(4):372–392, 2011.

Nesterov, Y. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization*, 22(2):341–362, 2012.

Nowak, K., Beardsley, J., Brekke, L. D., Ferguson, I., and Raff, D. Subseasonal prediction for water management: Reclamation forecast rodeo I and II. In *100th American Meteorological Society Annual Meeting*. AMS, 2020.

Orabona, F. A modern introduction to online learning. *ArXiv*, abs/1912.13213, 2019.

Orabona, F. and Pál, D. Scale-free algorithms for online linear optimization. In *International Conference on Algorithmic Learning Theory*, pp. 287–301. Springer, 2015.

Orabona, F. and Pál, D. Optimal non-asymptotic lower bound on the minimax regret of learning with expert advice. *arXiv preprint arXiv:1511.02176*, 2015.

Quanrud, K. and Khashabi, D. Online learning with adversarial delays. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 28, pp. 1270–1278, 2015.

Rakhlin, A. and Sridharan, K. Online learning with predictable sequences. In Shalev-Shwartz, S. and Steinwart, I. (eds.), *Proceedings of the 26th Annual Conference on Learning Theory*, pp. 993–1019. PMLR, 2013a.

Rakhlin, S. and Sridharan, K. Optimization, learning, and games with predictable sequences. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, pp. 3066–3074. Curran Associates, Inc., 2013b.

Recht, B., Re, C., Wright, S., and Niu, F. Hogwild!: A lock-free approach to parallelizing stochastic gradient descent. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, volume 24, pp. 693–701. Curran Associates, Inc., 2011.

Rockafellar, R. T. *Convex analysis*, volume 36. Princeton university press, 1970.

Shalev-Shwartz, S. *Online learning: Theory, algorithms, and applications*. PhD thesis, The Hebrew University, 2007.

Shalev-Shwartz, S. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2): 107–194, 2012.

Sra, S., Yu, A. W., Li, M., and Smola, A. AdaDelay: Delay adaptive distributed stochastic optimization. In Gretton, A. and Robert, C. C. (eds.), *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51, pp. 957–965. PMLR, 2016.

Steinhardt, J. and Liang, P. Adaptivity and optimism: An improved exponentiated gradient algorithm. In *International Conference on Machine Learning*, pp. 1593–1601, 2014.

Syrgkanis, V., Agarwal, A., Luo, H., and Schapire, R. E. Fast convergence of regularized learning in games. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

Tammelin, O., Burch, N., Johanson, M., and Bowling, M. Solving heads-up limit texas hold'em. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

Weinberger, M. J. and Ordentlich, E. On delayed prediction of individual sequences. *IEEE Transactions on Information Theory*, 48(7):1959–1976, 2002.

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J., Lazo, J. K., Kumar, A., Vitart, F., Coughlan de Perez, E., Ray, A. J., Murray, V., et al. Potential applications of subseasonal-to-seasonal (s2s) predictions. *Meteorological applications*, 24(3):315–325, 2017.

Zinkevich, M., Johanson, M., Bowling, M. H., and Piccione, C. Regret minimization in games with incomplete information. In Platt, J., Koller, D., Singer, Y., and Roweis, S. (eds.), *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007.