Online Learning with Optimism and Delay

Anonymous Authors[1]

Abstract

Inspired by the demands of real-time time-series forecasting, we develop and analyze optimistic online learning algorithms under delayed feedback. We present a novel "delay as optimism" analysis that reduces online learning under delay to optimistic online learning. This reduction enables optimal regret bounds for delayed online learning and exposes how side-information or optimistic "hints" can be used to combat the effects of delay. We use these theoretical tools to develop the first optimistic online learning algorithms that require no parameter tuning and have optimal regret guarantees under delay. These algorithms — DORM, DORM+, and AdaHedgeD— are robust and practical choices for real-world time-series forecasting. 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 classical sequential decision-making paradigm in which a learner is pitted against a potentially adversarial environment (Orabona, 2019; Shalev-Shwartz, 2007). 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 objective 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 can provide robust performance in many complex real-world online prediction problems such as climate or financial 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. Several delayed algorithms are known to achieve optimal worst-case regret rates against adversarial loss sequences (Weinberger & Ordentlich, 2002; Joulani et al., 2013; McMahan & Streeter, 2014; Joulani et al., 2017), but each has its 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 rely on uniform upper bounds on future loss gradients to set their tuning parameters (McMahan & Streeter, 2014; Joulani et al., 2017). None leverage optimistic hints to improve performance when the delayed losses are partially predictable. The concurrent work of Hsieh et al. (2020) analyzes optimistic gradient descent under delay but relies on uniform bounds on future gradients that are often challenging to obtain and overly conservative in applications.

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 past and future losses are available. We begin by viewing 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 of App. B).
4. The first general analysis of follow-the-regularized leader algorithms under delay (Thm. 5 of App. B and

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

Thm. 18 of App. I).

5. The first analysis of delayed online mirror descent algorithms with optimistic hints (Thm. 6 of App. B).

We validate our algorithms on the problem of subseasonal forecasting in Sec. 5. Subseasonal forecasting — predicting precipitation and temperature 2-6 weeks in advance — is a crucial task for allocating water resources, managing wildfires, and preparing for other weather extremes (White et al., 2017). Several challenges emerge when applying existing online learning methods to subseasonal forecasting that our algorithms are equipped to manage. 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, self-tuned or tuning-free algorithms are essential for real-time, practical deployment. We demonstrate that our algorithms DORM, DORM+, and AdaHedgeD all produce strong performance and that DORM+ particularly achieves consistently low regret compared to the best forecasting models.

Open-source Python code implementing DORM, DORM+ and AdaHedgeD and recreating our subseasonal forecasting experiments is available at `redacted`.

Notation For integers a, b, we use the shorthand $\mathbf{g}_{a:b} \triangleq \sum_{i=a}^{b} \mathbf{g}_i$. 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 $\|\cdot\|$ 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 $\operatorname{diam}(\mathbf{W}) = \inf_{\mathbf{w}, \mathbf{w}' \in \mathbf{W}} \|\mathbf{w} - \mathbf{w}'\|$ and $(r)_+ \triangleq \max(r, 0)$.

2. Online Learning with Optimism and Delay

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 achieve improved performance when loss sequences are partially predictable, while maintaining robustness to adversarial sequences (see, e.g., Rakhlin & Sridharan, 2013b; Steinhardt & Liang, 2014; Kamalaruban, 2016; Chiang et al., 2012). In many formulations of optimistic online learning, the learner is provided with a pseudo-loss $\tilde{\ell}_t$ at the start of round t that represents a guess for the true, unknown loss at time t. The online learner can incorporate this hint into its learning process before making play $\mathbf{w}_t$. When loss feedback is delayed by D time steps, the learner observes the losses $\{\ell_s\}_{s=1}^{t-D-1}$ and the optimistic pseudolosses $\{\tilde{\ell}_s\}_{s=1}^{t}$ before playing $\mathbf{w}_t$.

In the remainder of the text, we use the following notation for the subdifferential of the online learning loss and optimistic pseudoloss respectively: $\mathbf{g}_t \in \partial \ell_t(\mathbf{w}_t)$, $\tilde{\mathbf{g}}_t \in \partial \tilde{\ell}_t(\mathbf{w}_{t-1})$.

In the delayed and optimistic setting, we propose counterparts of standard FTRL and OMD online learning algorithms, which we call *optimistic delayed FTRL (*ODFTRL*)* and *delayed optimistic online mirror descent (*DOOMD*)* respectively. These algorithms produce iterates $\mathbf{w}_t$ satisfying,

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

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

for constant delay period D, regularization parameter λ, and optimistic hint vector $\mathbf{h}_t = \sum_{s=t-D}^{t} \tilde{\mathbf{g}}_s$, representing our best guess of the summed gradients of missing delayed and future losses.

2.1. Delay as Optimism

A first key insight of this paper is that, for ODFTRL and DOOMD,

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

In particular, ODFTRL and DOOMD are instances of optimistic FTRL (OFTRL) and single-step optimistic OMD (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* $\tilde{\mathbf{g}}_{t+1} = \tilde{\mathbf{g}}_t + \mathbf{g}_{t-D} - \mathbf{g}_t + \mathbf{h}_{t+1} - \mathbf{h}_t = \mathbf{h}_{t+1} - \sum_{s=t-D+1}^{t} \mathbf{g}_s$.

In App. B, we demonstrate that, as an immediate consequence of our delay-as-optimism perspective, we can provide new regret bounds for ODFTRL and DOOMD. The form of these delayed regret bounds reveals the heightened value of optimism in the presence of delay: in addition to providing an effective guess of a 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, we still only pay the minimax optimal $\sqrt{D+1}$ penalty for delayed feedback. It remains to choose the regularization parameter λ to achieve this minimax optimal regret rate.

3. Tuning-free Learning with Optimism and Delay

As an application of our ODFTRL and DOOMD analysis, we introduce and analyze delayed and optimistic versions of two popular tuning-free online learning algorithms: regret matching (RM) (Blackwell, 1956; Hart & Mas-Colell, 2000) and regret matching+ (RM+) (Tammelin et al., 2015). 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 solve the game of Heads-up Limit Texas Hold'em poker (Bowling et al., 2015).

Our generalizations, *delayed optimistic regret matching* (DORM)

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

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

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

allow for delay D, regularization parameter λ, optimistic hints $\mathbf{h}_t$, and a parameter $q \geq 2$ and its conjugate exponent $p = q/(q-1)$. We refer to $\mathbf{r}_t$ as the *instantaneous regret* of each expert with respect to the play $\mathbf{w}_t$ and the linearized loss vector $\mathbf{g}_t$, and note that DORM and DORM+ recover the standard RM and RM+ algorithms when $D = 0$, $\lambda = 1$, $q = 2$, and $\mathbf{h}_t = \mathbf{0}$, $\forall t$.

While these updates may look unfamiliar, we show in App. E that they are special cases of the ODFTRL and DOOMD algorithms. Specifically, we connect DORM to ODFTRL and DORM+ to DOOMD, which enables us to extend previous regret bounds to DORM and DORM+. Additionally, under mild conditions detailed in App. E, we highlight a remarkable property:

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

This result, shown in App. E, implies that DORM and DORM+ are *automatically* optimally tuned with respect to λ, even when run with a default value of $\lambda = 1$.

4. Self-tuned Learning with Optimism and Delay

We now analyze an adaptive version of ODFTRL with time-varying regularization $\lambda_t\psi$ and develop strategies for automatically tuning λ_t in the presence of optimism and delay. Our objective is to achieve the minimax optimal regret rate *and* to find a setting of λ_t that performs well in pratical applications. 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 relies heavily on the tightness of that bound. Our next result introduces analyzes a new tuning strategy inspired by the popular AdaHedge algorithm (Erven et al., 2011) and based on a new tighter bound on ODFTRL regret:

Fix $\alpha > 0$, and consider the *delayed AdaHedge-style* (AdaHedgeD) regularizer sequence within an ODFTRL update:

$$\begin{aligned}
&\lambda_{t+1} = \tfrac{1}{\alpha}\textstyle\sum_{s=1}^{t-D}\delta_t \qquad \text{for} \qquad \text{(AdaHedgeD)}\\
&\delta_t \triangleq \min\left(F_{t+1}(\mathbf{w}_t, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t), \langle \mathbf{g}_t, \mathbf{w}_t - \bar{\mathbf{w}}_t\rangle\right)_+\\
&\quad \text{with} \quad \bar{\mathbf{w}}_t = \operatorname{argmin}_{\mathbf{w}\in\mathbf{W}} F_{t+1}(\mathbf{w}, \lambda_t) \qquad (1)\\
&\quad \text{and} \quad F_{t+1}(\mathbf{w}, \lambda_t) \triangleq \lambda_t\psi(\mathbf{w}) + \langle \mathbf{g}_{1:t}, \mathbf{w}\rangle.
\end{aligned}$$

Remarkably, as we show in App. I, this setting of adaptive regularization yields a minimax optimal $\mathcal{O}(\sqrt{(D+1)T} + D)$ dependence on the delay parameter and nearly matches the regret of the optimal constant λ tuning in hindsight.

5. Experiments

We apply the online learning techniques developed in this paper to the problem of adaptive ensembling for subseasonal forecasting. Our experiments are based on the public subseasonal forecasting codebase of Orenstein (2021) that uses $d = 6$ physics-based numerical models and machine learning models (CFSv2++, Climatology++, LocalBoosting , MultiLLR , Persistence++, and Salient++) to predict temperature and precipitation 2-6 weeks ahead. In this mid-range climate forecasting task, forecast feedback is delayed; the 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 the d input models, such that $\mathbf{w}_t \in \triangle_{d-1}$. Our objective is to compete with the best input model over a year-long prediction period ($T = 26$ semimonthly dates). The loss function for each forecast date is the geographic root-mean squared error (RMSE) across 514 locations in the Western United States.

We consider four subseasonal prediction tasks – predicting temperature and precipitation at two horizons, weeks 3-4 and weeks 5-6 – and evaluate yearly regret and mean RMSE for each year from 2011-2020. Unless otherwise specified, all online learning algorithms use the previous gradient

Table 1: **Average RMSE of the 2011-2020 semimonthly forecasts**: The average RMSE for online learning algorithms (left) and individual models (right) over a 10-year evaluation period. The top performing model for each task is bolded and shown in green.

	AdaHedgeD	DORM	DORM+	CFSv2++	Clim.++	LocalBoosting	MultiLLR	Persist.++	Salient++
Precip. 34w	21.837	21.737	**21.675**	21.978	21.986	22.357	22.431	21.973	23.344
Precip. 56w	21.987	21.957	**21.838**	22.004	21.993	22.383	22.570	22.030	23.257
Temp. 34w	2.287	2.259	**2.247**	2.277	2.319	2.394	2.352	2.253	2.508
Temp. 56	2.321	2.318	2.304	**2.278**	2.317	2.440	2.368	2.284	2.569

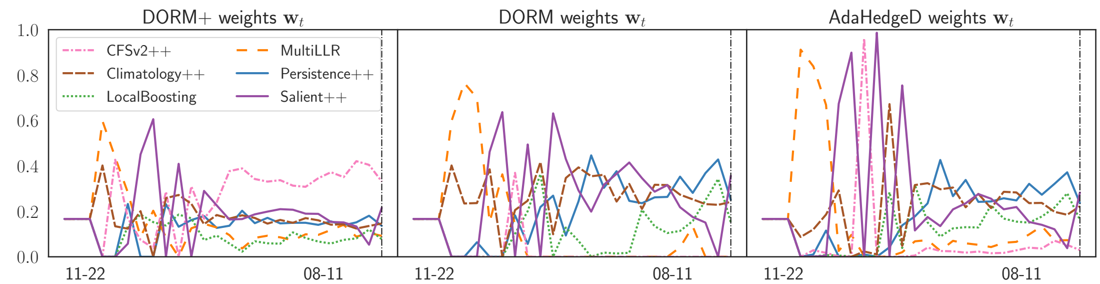

Figure 1: **Impact of regularization:** The plays $\mathbf{w}_t$ of online learning algorithms used to combine the input models for the Temp. 34w task in the 2020 evaluation year. DORM and AdaHedgeD are both FTRL-based algorithms and have similar plays; AdaHedgeD appears to be less regularized. DORM+, on the other hand, is an OMD-based algorithm and was designed to handle applications where the "best" expert model changes frequently. DORM+ is the top learning algorithm for this subseasonal forecasting task, indicating the importance of this adaptivity property.

optimism strategy $\mathbf{h}_t = (D+1)\mathbf{g}_{t-D-1}$. See App. M for full experimental details and App. N for algorithmic details.

5.1. Competing with the best input model

Our primary objective in online learning is to achieve zero average regret, i.e., to perform as well as the best input model in the competitor set $\mathbf{U}$. To evaluate model regret, we run our three delayed online learning algorithms — DORM, DORM+, and AdaHedgeD— on all four subseasonal prediction tasks and measure their average RMSE loss.

The average yearly RMSE for the three online learning algorithms and the six input models is shown in Table 1. The DORM+ algorithm outperforms the best input model for all tasks except Temp. 56w. All online learning algorithms achieve negative regret for both precipitation tasks. Fig. 1 shows an example of the weights played by the three algorithms. Fig. 2 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 achieve negative regret, outperforming the best input forecasting model. The consistently low regret year-to-year of DORM+ makes it a promising candidate for real-world delayed subseasonal forecasting.

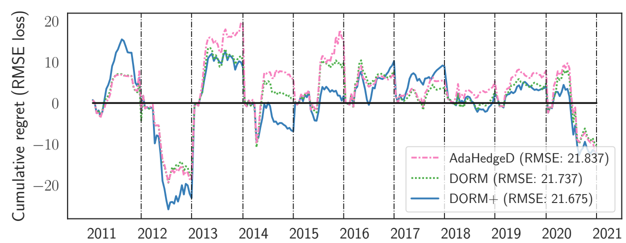

Figure 2: **Overall performance**: Yearly cumulative regret under RMSE loss for the three delayed online learning algorithms presented, over the 10-year evaluation period for the Precip. 34w task. The zero line corresponds to the performance of the best input model in a given year. Negative values indicate that the online learner outperformed the best input model in a given year.

6. Conclusion

In this work, we overcame the challenges of delayed feedback and short regret horizons in online learning with optimism. We developed three practical non-replicated, self-tuned and tuning-free algorithms with optimal regret guarantees — DORM, DORM+, and AdaHedgeD. Our "delay as optimism" reduction and 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 state-of-the art forecasting ensembles robustly from year-to-year. Our results highlighted DORM+ as a particularly promising candidate for subseasonal forecasting due to its tuning-free nature and adaptivity when the best input model changes frequently. Through theoretical and experimental validation, we have presented DORM, DORM+, and AdaHedgeD as practical and robust algorithms for delayed time-series forecasting that can be applied in a variety of application domains to improve the quality of sequential decision-making.

References

Agarwal, A. and Duchi, J. C. Distributed delayed stochastic optimization. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pp. 5451–5452. IEEE, 2012.

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.

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. *Advances in Neural Information Processing Systems*, 24: 1656–1664, 2011.

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.

Hwang, J., Orenstein, P., Cohen, J., Pfeiffer, K., and Mackey, L. Improving subseasonal forecasting in the western us 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. *arXiv preprint arXiv:1709.02726*, 2017.

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

McMahan, B. and Streeter, M. Delay-tolerant algorithms for asynchronous distributed online learning. In *Advances in Neural Information Processing Systems*, pp. 2915–2923, 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.

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.

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

Orenstein, P. Subseasonal forecasting models, 2021. URL http://w3.impa.br/~pauloo/subseasonal_forecasting/code.zip.

Rakhlin, A. and Sridharan, K. Online learning with predictable sequences. In *Conference on Learning Theory*, pp. 993–1019, 2013a.

Rakhlin, S. and Sridharan, K. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, pp. 3066–3074, 2013b.

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

A. Extended Literature Review

We review here additional prior work not detailed in the main paper.

A.1. General online learning

We recommend the monographs of Shalev-Shwartz et al. (2012); Orabona (2019) and the textbook of Cesa-Bianchi & Lugosi (2006) for surveys of the field of online learning and Joulani et al. (2017); McMahan (2017) for widely applicable and modular analyses of online learning algorithms.

A.2. Online learning with optimism but without delay

Syrgkanis et al. (2015) analyzed optimistic FTRL and two-step variant of optimistic MD without delay. The work focuses on a particular form of optimism (using the last observed gradient as a hint) and shows improved rates of convergence to correlated equilibria in multiplayer games. In the absence of delay, Steinhardt & Liang (2014) combined optimism and adaptivity to obtain improvements over standard optimistic regret bounds.

A.3. Online learning with delay but without optimism

Overview Joulani et al. (2013; 2016); McMahan & Streeter (2014) provide broad reviews of progress on delayed online learning.

Delayed stochastic optimization Recht et al. (2011); Agarwal & Duchi (2012); Nesterov (2012); Liu et al. (2014); Liu & Wright (2015); Sra et al. (2016) studied the effects of delay on stochastic optimization but do not treat the adversarial setting studied here.

Non-replicated algorithms Korotin et al. (2020); Quanrud & Khashabi (2015) developed online learning algorithms that do not rely on replication to perform online learning under delay.

FTRL-Prox vs. FTRL Joulani et al. (2016) analyzed the delayed feedback regret of the *FTRL-Prox* algorithm, which regularizes toward the last played iterate as in online mirror descent, but did not study the standard FTRL algorithms (sometimes called *FTRL-Centered*) analyzed in this work.

A.4. Self-tuning online learning without delay or optimism

In the absence of optimism and delay, de Rooij et al. (2014); Orabona & Pál (2015); Koolen et al. (2014) developed alternative variants of FTRL algorithms that self-tune their learning rates.

A.5. Online learning without delay for climate forecasting

Monteleoni et al. (2011) applied the Learn-α online learning algorithm of Monteleoni & Jaakkola (2004) to the task of ensembling climate models. The authors considered historical temperature data from 20 climate models and tracked the changing sequence of which model predicts best at any given time. In this context, the algorithm used was based on a set of generalized Hidden Markov Models, in which the identity of the current best model is the hidden variable and the updates are derived as Bayesian updates. This work was extended to take into account the influence of regional neighboring locations when performing updates (McQuade & Monteleoni, 2012). These initial results demonstrated the promise of applying online learning to climate model ensembling, but both methods rely on receiving feedback without delay.

B. Online Learning with Optimism and Delay

The inclusion of $\tilde{\ell}_t$ within an standard FTRL or OMD online learning algorithm leads to the following linearized update rules:

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

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

where $\tilde{\mathbf{g}}_{t+1} \in \mathbb{R}^d$ is an arbitrary hint vector revealed before $\mathbf{w}_{t+1}$ is played, $\lambda \geq 0$ is a regularization parameter, and ψ is proper and 1-strongly convex with respect to a norm $\|\cdot\|$. If the hints are a good approximation of the missing losses, the online learner can make use of the hint to improve it's regret; if they are a poor approximation, the learner suffers only a small constant regret penalty.

B.1. Delay as Optimism

It is worth reflecting on the uniqueness of Lem. 2. A number of alternative optimistic OMD algorithms have been proposed in the literature (Chiang et al., 2012; Rakhlin & Sridharan, 2013a;b; Kamalaruban, 2016). Unlike SOOMD, these procedures all incorporate optimism in two steps, as in the updates

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

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.

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. For example, consider the following new regret bounds for OFTRL and SOOMD, proved in Apps. C and D respectively.

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

$$\text{Regret}_T(\mathbf{u}) \leq \lambda\psi(\mathbf{u}) +$$
$$\tfrac{1}{\lambda}\textstyle\sum_{t=1}^T \min(\tfrac{1}{2}\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*^2, \|\mathbf{g}_t\|_*\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*).$$

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*

$$\text{Regret}_T(\mathbf{u}) \leq \mathcal{B}_{\lambda\psi}(\mathbf{u}, \mathbf{w}_0) +$$
$$\tfrac{1}{\lambda}\textstyle\sum_{t=1}^T \min(\tfrac{1}{2}\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*^2, \|\mathbf{g}_t + \tilde{\mathbf{g}}_{t+1} - \tilde{\mathbf{g}}_t\|_*\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*).$$

The first argument of the $\min$ in Thms. 3 and 4 recovers standard OFTRL regret bounds (see, e.g., Rakhlin & Sridharan, 2013a; Mohri & Yang, 2016; Orabona, 2019, Thm. 7.28) and the SOOMD bound sketched in Joulani et al. (2017, Sec. 7.2). We will show that the second argument of the $\min$ provides an important strengthening that is critical for obtaining optimal regret bounds under delayed feedback.

As an immediate consequence of our delay-as-optimism perspective, we can provide the following pair of new regret bounds for ODFTRL and DOOMD.

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

$$\text{Regret}_T(\mathbf{u}) \leq \lambda\psi(\mathbf{u}) + \tfrac{1}{\lambda}\textstyle\sum_{t=1}^T \mathbf{b}_{t,F}^2,$$
$$\textit{for} \quad \mathbf{b}_{t,F}^2 \triangleq \min\big(\tfrac{1}{2}\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^t \mathbf{g}_s\|_*^2,$$
$$\|\mathbf{g}_t\|_*\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^t \mathbf{g}_s\|_*\big).$$

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*

$$\begin{aligned}
&\text{Regret}_T(\mathbf{u}) \leq \mathcal{B}_{\lambda\psi}(\mathbf{u}, \mathbf{w}_0) + \tfrac{1}{\lambda}\textstyle\sum_{t=1}^T \mathbf{b}_{t,O}^2, \\
&\textit{for} \quad \mathbf{b}_{t,O}^2 \triangleq \min(\tfrac{1}{2}\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^t \mathbf{g}_s\|_*^2, \\
&\qquad \|\mathbf{g}_{t-D} + \mathbf{h}_{t+1} - \mathbf{h}_t\|_* \|\mathbf{h}_t - \textstyle\sum_{s=t-D}^t \mathbf{g}_s\|_*).
\end{aligned}$$

The $\mathbf{b}_{t,*}^2$ terms in Thms. 5 and 6 are of size $\mathcal{O}((D+1)G^2)$, where $G \geq \|\mathbf{g}_t\|_* \, \forall t \in [T]$. They reveal a compounding of regret due to delay that results in an extra $\sqrt{D+1}$ factor in the worst case relative to the standard undelayed setting. This $\sqrt{D+1}$ scaling of regret is known to be minimax optimal in the adversarial setting (Weinberger & Ordentlich, 2002).

The form of these delayed regret bounds also reveals the heightened value of optimism in the presence of delay: in addition to providing an effective guess of a 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 second term in the $\min$ ensures that we still only pay the minimax optimal $\sqrt{D+1}$ penalty for delayed feedback.

Related work Weinberger & Ordentlich (2002); Joulani et al. (2013); Agarwal & Duchi (2012); Mesterharm (2005) analyze replicated learning algorithms under delay and observe an explicit multiplicative $\sqrt{D+1}$ penalty due to delay. While our bounds also feature this penalty in the worst case, the penalty can be significantly reduced by optimism. To our knowledge, Thm. 5 is the first general analysis of delayed non-replicated FTRL. In the absence of optimism, McMahan & Streeter (2014) obtain a comparable bound for the special case of unconstrained online gradient descent. Joulani et al. (2016) studies non-optimistic OMD under delay and presents the iterate drift bound used in our DOOMD analysis. Thm. 6 strengthens their bounds, which feature a sum of gradient norms $\sum_{s=t-D}^t \|\mathbf{g}_s\|_*$ in place of $\|\mathbf{h}_t - \sum_{s=t-D}^t \mathbf{g}_s\|_*$. Even in the absence of optimism, the latter can be significantly smaller due to cancellations in the gradient sum.

C. Proof of Thm. 3: OFTRL regret

We will prove the following more general result for optimistic adaptive FTRL (OAFTRL)

$$\mathbf{w}_{t+1} = \text{argmin}_{\mathbf{w} \in \mathbf{W}} \langle \mathbf{g}_{1:t} + \tilde{\mathbf{g}}_{t+1}, \mathbf{w} \rangle + \lambda_{t+1}\psi(\mathbf{w}), \tag{OAFTRL}$$

from which Thm. 3 will follow with the choice $\lambda_t = \lambda$ for all $t \geq 1$.

Theorem 7 (OAFTRL regret)**.** *If ψ is nonnegative and $(\lambda_t)_{t\geq 1}$ is non-decreasing, then, $\forall \mathbf{u} \in \mathbf{W}$, the* OAFTRL *iterates $\mathbf{w}_t$ satisfy,*

$$\begin{aligned}
&\text{Regret}_T(\mathbf{u}) \\
&\quad \leq \lambda_T \psi(\mathbf{u}) + \textstyle\sum_{t=1}^T \delta_t \\
&\quad \leq \lambda_T \psi(\mathbf{u}) + \textstyle\sum_{t=1}^T \min\left(\tfrac{1}{2\lambda_t}\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*^2, \tfrac{1}{\lambda_t}\|\mathbf{g}_t\|_*\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*, \text{diam}(\mathbf{W})\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*, \text{diam}(\mathbf{W})\|\mathbf{g}_t\|_*\right),
\end{aligned}$$

for

$$\begin{aligned}
&\delta_t \triangleq \min\left(F_{t+1}(\mathbf{w}_t, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t), \langle \mathbf{g}_t, \mathbf{w}_t - \bar{\mathbf{w}}_t \rangle\right)_+ \quad \textit{with} \\
&\bar{\mathbf{w}}_t \triangleq \text{argmin}_{\mathbf{w} \in \mathbf{W}} F_{t+1}(\mathbf{w}, \lambda_t) \quad \textit{and} \quad F_{t+1}(\mathbf{w}, \lambda_t) \triangleq \lambda_t \psi(\mathbf{w}) + \langle \mathbf{g}_{1:t}, \mathbf{w} \rangle.
\end{aligned}$$

Proof. Consider a sequence of arbitrary auxiliary gradient hints $\tilde{\mathbf{g}}_1^*, \ldots, \tilde{\mathbf{g}}_T^* \in \mathbb{R}^d$ and the auxiliary OAFTRL sequence

$$\mathbf{w}_{t+1}^* = \text{argmin}_{\mathbf{w}^* \in \mathbf{W}} \langle \mathbf{g}_{1:t} + \tilde{\mathbf{g}}_{t+1}^*, \mathbf{w}^* \rangle + \lambda_{t+1}\psi(\mathbf{w}^*) \quad \text{for} \quad 0 \leq t \leq T \quad \text{with} \quad \tilde{\mathbf{g}}_{T+1}^* \triangleq \mathbf{0} \quad \text{and} \quad \lambda_{T+1} = \lambda_T. \tag{2}$$

Generalizing the forward regret decomposition of Joulani et al. (2017) and the prediction drift decomposition of Joulani et al. (2016), we will decompose the regret of our original $(\mathbf{w}_t)_{t=1}^T$ sequence into the regret of the auxiliary sequence $(\mathbf{w}_t^*)_{t=1}^T$ and the drift between $(\mathbf{w}_t)_{t=1}^T$ and $(\mathbf{w}_t^*)_{t=1}^T$.

For each time t, define the auxiliary optimistic objective function $\tilde{F}_t^*(\mathbf{w}) = F_t(\mathbf{w}) + \langle \tilde{\mathbf{g}}_t^*, \mathbf{w} \rangle$. Fixing any $\mathbf{u} \in \mathbf{W}$, we have the regret bound

$$\begin{aligned}\text{Regret}_T(\mathbf{u}) &= \textstyle\sum_{t=1}^T \ell_t(\mathbf{w}_t) - \ell_t(\mathbf{u}) \leq \sum_{t=1}^T \langle \mathbf{g}_t, \mathbf{w}_t - \mathbf{u} \rangle \quad \text{(since each } \ell_t \text{ is convex with } \mathbf{g}_t \in \partial \ell_t(\mathbf{w}_t)\text{)} \\ &= \underbrace{\textstyle\sum_{t=1}^T \langle \mathbf{g}_t, \mathbf{w}_t - \mathbf{w}_t^* \rangle}_{\text{drift}} + \underbrace{\textstyle\sum_{t=1}^T \langle \mathbf{g}_t, \mathbf{w}_t^* - \mathbf{u} \rangle}_{\text{auxiliary regret}}.\end{aligned}$$

To control the drift term we employ the following lemma, proved in App. C.1, which bounds the difference between two OAFTRL optimizers with different losses but common regularizers.

Lemma 8 (OAFTRL difference bound)**.** *The* OAFTRL *and auxiliary* OAFTRL *iterates* (2)*,* $\mathbf{w}_t$ *and* $\mathbf{w}_t^*$*, satisfy*

$$\|\mathbf{w}_t - \mathbf{w}_t^*\| \leq \min(\tfrac{1}{\lambda_t} \|\tilde{\mathbf{g}}_t - \tilde{\mathbf{g}}_t^*\|_*, \text{diam}(\mathbf{W})).$$

Letting $a = \text{diam}(\mathbf{W}) \in \mathbb{R} \cup \{\infty\}$, we now bound each drift term summand using the Fenchel-Young inequality for dual norms and Lem. 8:

$$\langle \mathbf{g}_t, \mathbf{w}_t - \mathbf{w}_t^* \rangle \leq (\langle \mathbf{g}_t, \mathbf{w}_t - \mathbf{w}_t^* \rangle)_+ \leq \|\mathbf{g}_t\|_* \|\mathbf{w}_t - \mathbf{w}_t^*\| \leq \min\left(\tfrac{1}{\lambda_t} \|\mathbf{g}_t\|_* \|\tilde{\mathbf{g}}_t - \tilde{\mathbf{g}}_t^*\|_*, a\|\mathbf{g}_t\|_*\right).$$

To control the auxiliary regret, we begin by invoking the OAFTRL regret bound of Orabona (2019, proof of Thm. 7.27), the nonnegativity of ψ, and the assumption that $(\lambda_t)_{t \geq 1}$ is non-decreasing:

$$\begin{aligned}\textstyle\sum_{t=1}^T \langle \mathbf{g}_t, \mathbf{w}_t^* - \mathbf{u} \rangle &\leq \lambda_{T+1}\psi(\mathbf{u}) - \lambda_1 \psi(\mathbf{w}_1^*) + \textstyle\sum_{t=1}^T F_{t+1}(\mathbf{w}_t^*, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t) + (\lambda_t - \lambda_{t+1})\psi(\mathbf{w}_{t+1}^*) \\ &\leq \lambda_{T+1}\psi(\mathbf{u}) - \lambda_1 \psi(\mathbf{w}_1^*) + \textstyle\sum_{t=1}^T F_{t+1}(\mathbf{w}_t^*, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t)\end{aligned}$$

We next bound the summands in this expression in two ways. Since $\mathbf{w}_t^*$ is the minimizer of $\tilde{F}_t^*$, we may apply the Fenchel-Young inequality for dual norms to conclude that

$$\begin{aligned}F_{t+1}(\mathbf{w}_t^*, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t) &= \tilde{F}_t^*(\mathbf{w}_t^*) + \langle \mathbf{w}_t^*, \mathbf{g}_t - \tilde{\mathbf{g}}_t^* \rangle - (\tilde{F}_t^*(\bar{\mathbf{w}}) + \langle \bar{\mathbf{w}}, \mathbf{g}_t - \tilde{\mathbf{g}}_t^* \rangle) \\ &\leq \langle \mathbf{w}_t^* - \bar{\mathbf{w}}, \mathbf{g}_t - \tilde{\mathbf{g}}_t^* \rangle \leq \|\mathbf{w}_t^* - \bar{\mathbf{w}}\| \|\mathbf{g}_t - \tilde{\mathbf{g}}_t^*\|_* \leq a\|\mathbf{g}_t - \tilde{\mathbf{g}}_t^*\|_*.\end{aligned}$$

Moreover, by Orabona (2019, proof of Thm. 7.27) and the fact that $\bar{\mathbf{w}}_t$ minimizes $F_{t+1}(\cdot, \lambda_t)$ over $\mathbf{W}$,

$$F_{t+1}(\mathbf{w}_t^*, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t) \leq \tfrac{\|\mathbf{g}_t - \tilde{\mathbf{g}}_t^*\|_*^2}{2\lambda_t}.$$

Now, introduce the shorthand

$$\delta_t(\tilde{\mathbf{g}}_t^*) = F_{t+1}(\mathbf{w}_t^*, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t) + \langle \mathbf{g}_t, \mathbf{w}_t - \mathbf{w}_t^* \rangle.$$

By our established bounds and the value $\delta_t = \min(\delta_t(\tilde{\mathbf{g}}_t), \delta_t(\mathbf{g}_t))_+$ satisfies:

$$\delta_t = \min(\delta_t(\tilde{\mathbf{g}}_t), \delta_t(\mathbf{g}_t))_+ \leq \min\Big(\min\left(\tfrac{1}{2\lambda_t}\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*^2, a\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*\right), \min\left(\tfrac{1}{\lambda_t}\|\mathbf{g}_t\|_*\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*, a\|\mathbf{g}_t\|_*\right)\Big), \tag{3}$$

Since $\tilde{\mathbf{g}}_t^*$ is arbitrary, we obtain the advertised regret bounds,

$$\begin{aligned}\text{Regret}_T(\mathbf{u}) &\leq \inf_{\tilde{\mathbf{g}}_1^*, \ldots, \tilde{\mathbf{g}}_T^* \in \mathbb{R}^d} \lambda_{T+1}\psi(\mathbf{u}) + \textstyle\sum_{t=1}^T \delta_t(\tilde{\mathbf{g}}_t^*) \\ &= \lambda_{T+1}\psi(\mathbf{u}) + \textstyle\sum_{t=1}^T \inf_{\tilde{\mathbf{g}}_t^* \in \mathbb{R}^d} \delta_t(\tilde{\mathbf{g}}_t^*) \\ &\leq \lambda_{T+1}\psi(\mathbf{u}) + \textstyle\sum_{t=1}^T \min(\delta_t(\tilde{\mathbf{g}}_t), \delta_t(\mathbf{g}_t))_+ \\ &\leq \lambda_{T+1}\psi(\mathbf{u}) + \textstyle\sum_{t=1}^T \min\Big(\min\left(\tfrac{1}{2\lambda_t}\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*^2, a\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*\right), \min\left(\tfrac{1}{\lambda_t}\|\mathbf{g}_t\|_*\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*, a\|\mathbf{g}_t\|_*\right)\Big).\end{aligned}$$

□

C.1. Proof of Lem. 8: OAFTRL difference bound

Fix any time t, and define the optimistic objective function $\tilde{F}_t(\mathbf{w}) = \lambda_t \psi(\mathbf{w}) + \sum_{i=1}^{t-1} \langle \mathbf{g}_i, \mathbf{w} \rangle + \langle \tilde{\mathbf{g}}_t, \mathbf{w} \rangle$ and the auxiliary optimistic objective function $\tilde{F}_t^*(\mathbf{w}) = \lambda_t \psi(\mathbf{w}) + \sum_{i=1}^{t-1} \langle \mathbf{g}_i, \mathbf{w} \rangle + \langle \tilde{\mathbf{g}}_t^*, \mathbf{w} \rangle$ so that $\mathbf{w}_t \in \operatorname{argmin}_{\mathbf{w} \in \mathbf{W}} \tilde{F}_t(\mathbf{w})$ and $\mathbf{w}_t^* \in \operatorname{argmin}_{\mathbf{w} \in \mathbf{W}} \tilde{F}_t^*(\mathbf{w})$. We have

$$\begin{aligned} \tilde{F}_t^*(\mathbf{w}_t) - \tilde{F}_t^*(\mathbf{w}_t^*) &\geq \tfrac{\lambda_t}{2} \|\mathbf{w}_t - \mathbf{w}_t^*\|^2 \quad \text{by the strong convexity of } \tilde{F}_t^* \text{ and} \\ \tilde{F}_t(\mathbf{w}_t^*) - \tilde{F}_t(\mathbf{w}_t) &\geq \tfrac{\lambda_t}{2} \|\mathbf{w}_t - \mathbf{w}_t^*\|^2 \quad \text{by the strong convexity of } \tilde{F}_t. \end{aligned}$$

Summing the above inequalities and applying the Fenchel-Young inequality for dual norms, we obtain

$$\lambda_t \|\mathbf{w}_t - \mathbf{w}_t^*\|^2 \leq \langle \tilde{\mathbf{g}}_t^* - \tilde{\mathbf{g}}_t, \mathbf{w}_t - \mathbf{w}_t^* \rangle \leq \|\tilde{\mathbf{g}}_t - \tilde{\mathbf{g}}_t^*\|_* \|\mathbf{w}_t - \mathbf{w}_t^*\|,$$

which yields the first half of our target bound after rearrangement. The second half follows from the definition of diameter, as $\|\mathbf{w}_t - \mathbf{w}_t^*\| \leq \operatorname{diam}(\mathbf{W})$.

D. Proof of Thm. 4: SOOMD regret

We will prove the following more general result for adaptive SOOMD (ASOOMD)

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

from which Thm. 4 will follow with the choice $\lambda_t = \lambda$ for all $t \geq 1$.

Theorem 9 (ASOOMD regret)**.** *Fix any $\lambda_{T+1} \geq 0$. If each $(\lambda_{t+1} - \lambda_t)\psi$ is proper and differentiable, $\lambda_0 \triangleq 0$, and $\tilde{\mathbf{g}}_{T+1} \triangleq \mathbf{0}$, then, for all $\mathbf{u} \in \mathbf{W}$, the* ASOOMD *iterates $\mathbf{w}_t$ satisfy*

$$\begin{aligned} \text{Regret}_T(\mathbf{u}) \leq & \textstyle\sum_{t=0}^T (\lambda_{t+1} - \lambda_t) \mathcal{B}_\psi(\mathbf{u}, \mathbf{w}_t) + \\ & \textstyle\sum_{t=1}^T \min(\frac{1}{2\lambda_{t+1}} \|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*^2, \frac{1}{\lambda_{t+1}} \|\mathbf{g}_t + \tilde{\mathbf{g}}_{t+1} - \tilde{\mathbf{g}}_t\|_* \|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*, \operatorname{diam}(\mathbf{W}) \|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*). \end{aligned}$$

Proof. Fix any $\mathbf{u} \in \mathbf{W}$, instantiate the notation of Joulani et al. (2017, Sec. 7.2), and consider the choices

- $r_1 = \lambda_2 \psi$, $r_t = (\lambda_{t+1} - \lambda_t)\psi$ for $t \geq 2$, so that $r_{1:t} = \lambda_{t+1}\psi$ for $t \geq 1$,
- $q_t = \tilde{q}_t + \langle \tilde{\mathbf{g}}_{t+1} - \tilde{\mathbf{g}}_t, \cdot \rangle$ for $t \geq 0$,
- $\tilde{q}_0(\mathbf{w}) = \lambda_1 \mathcal{B}_\psi(\mathbf{w}, \mathbf{w}_0)$ and $\tilde{q}_t \equiv 0$ for all $t \geq 1$,
- $p_1 \triangleq r_1 - q_0 = r_1 - \tilde{q}_0 - \langle \tilde{\mathbf{g}}_1 - \tilde{\mathbf{g}}_0, \cdot \rangle = \lambda_2 \psi - \lambda_1 \mathcal{B}_\psi(\cdot, \mathbf{w}_0) - \langle \tilde{\mathbf{g}}_1 - \tilde{\mathbf{g}}_0, \cdot \rangle$,
- $p_t \triangleq r_t - q_{t-1} = r_t - \tilde{q}_{t-1} - \langle \tilde{\mathbf{g}}_t - \tilde{\mathbf{g}}_{t-1}, \cdot \rangle = (\lambda_{t+1} - \lambda_t)\psi - \langle \tilde{\mathbf{g}}_t - \tilde{\mathbf{g}}_{t-1}, \cdot \rangle$ for all $t \geq 2$.

Since, for each t, $\delta_t = 0$ and ℓ_t is convex, the ADA-MD regret inequality of Joulani et al. (2017, Eq. (24)) and the choice

$\tilde{\mathbf{g}}_{T+1} = 0$ imply that

$$\begin{aligned}
\mathrm{Regret}_T(\mathbf{u}) &= \sum_{t=1}^{T} \ell_t(\mathbf{w}_t) - \sum_{t=1}^{T} \ell_t(\mathbf{u}) \\
&\leq -\sum_{t=1}^{T} \mathcal{B}_{\ell_t}(\mathbf{u}, \mathbf{w}_t) + \sum_{t=0}^{T} q_t(\mathbf{u}) - q_t(\mathbf{w}_{t+1}) + \sum_{t=1}^{T} \mathcal{B}_{p_t}(\mathbf{u}, \mathbf{w}_t) \\
&\quad - \sum_{t=1}^{T} \mathcal{B}_{r_{1:t}}(\mathbf{w}_{t+1}, \mathbf{w}_t) + \sum_{t=1}^{T} \langle \mathbf{g}_t, \mathbf{w}_t - \mathbf{w}_{t+1} \rangle + \sum_{t=1}^{T} \delta_t \\
&\leq \lambda_1(\mathcal{B}_\psi(\mathbf{u}, \mathbf{w}_0) - \mathcal{B}_\psi(\mathbf{w}_1, \mathbf{w}_0)) + \sum_{t=0}^{T} \langle \tilde{\mathbf{g}}_{t+1} - \tilde{\mathbf{g}}_t, \mathbf{u} - \mathbf{w}_{t+1} \rangle \\
&\quad + \sum_{t=1}^{T} (\lambda_{t+1} - \lambda_t) \mathcal{B}_\psi(\mathbf{u}, \mathbf{w}_t) + \sum_{t=1}^{T} \langle \mathbf{g}_t, \mathbf{w}_t - \mathbf{w}_{t+1} \rangle - \lambda_{t+1} \mathcal{B}_\psi(\mathbf{w}_{t+1}, \mathbf{w}_t) \\
&= \sum_{t=0}^{T} (\lambda_{t+1} - \lambda_t) \mathcal{B}_\psi(\mathbf{u}, \mathbf{w}_t) + \sum_{t=0}^{T} \langle \mathbf{g}_t - \tilde{\mathbf{g}}_t, \mathbf{w}_t - \mathbf{w}_{t+1} \rangle - \lambda_{t+1} \mathcal{B}_\psi(\mathbf{w}_{t+1}, \mathbf{w}_t). \qquad (4)
\end{aligned}$$

To obtain our advertised bound, we begin with the expression (4) and invoke the 1-strong convexity of ψ and the nonnegativity of $\mathcal{B}_{\lambda\psi}(\mathbf{w}_1, \mathbf{w}_0)$ to find

$$\begin{aligned}
\mathrm{Regret}_T(\mathbf{u}) &\leq \textstyle\sum_{t=0}^{T} (\lambda_{t+1} - \lambda_t) \mathcal{B}_\psi(\mathbf{u}, \mathbf{w}_t) + \sum_{t=0}^{T} \langle \mathbf{g}_t - \tilde{\mathbf{g}}_t, \mathbf{w}_t - \mathbf{w}_{t+1} \rangle - \lambda_{t+1} \mathcal{B}_\psi(\mathbf{w}_{t+1}, \mathbf{w}_t) \\
&\leq \textstyle\sum_{t=0}^{T} (\lambda_{t+1} - \lambda_t) \mathcal{B}_\psi(\mathbf{u}, \mathbf{w}_t) + \sum_{t=1}^{T} \langle \mathbf{g}_t - \tilde{\mathbf{g}}_t, \mathbf{w}_t - \mathbf{w}_{t+1} \rangle - \frac{\lambda_{t+1}}{2} \|\mathbf{w}_t - \mathbf{w}_{t+1}\|^2. \qquad (5)
\end{aligned}$$

We will bound the final sum in this expression using two lemmas. The first is a bound on the difference between subsequent ASOOMD iterates distilled from Joulani et al. (2016, proof of Prop. 2).

Lemma 10 (ASOOMD iterate bound (Joulani et al., 2016, proof of Prop. 2))**.** *If ψ is differentiable and 1-strongly convex with respect to $\|\cdot\|$, then the* ASOOMD *iterates satisfy*

$$\|\mathbf{w}_t - \mathbf{w}_{t+1}\| \leq \tfrac{1}{\lambda_{t+1}} \|\mathbf{g}_t + \tilde{\mathbf{g}}_{t+1} - \tilde{\mathbf{g}}_t\|_*.$$

The second, proved in App. D.1, is a general bound on $\langle \mathbf{g}, \mathbf{v} \rangle - \frac{\lambda}{2}\|\mathbf{v}\|^2$ under a norm constraint on $\mathbf{v}$.

Lemma 11 (Norm-constrained conjugate)**.** *For any $\mathbf{g} \in \mathbb{R}^d$ and $\lambda, c > 0$,*

$$\begin{aligned}
\sup_{\mathbf{v} \in \mathbb{R}^d : \|\mathbf{v}\| \leq \frac{c}{\lambda}} \langle \mathbf{g}, \mathbf{v} \rangle - \tfrac{\lambda}{2}\|\mathbf{v}\|^2 &= \tfrac{1}{\lambda} \min(\|\mathbf{g}\|_*, c)(\|\mathbf{g}\|_* - \tfrac{1}{2}\min(\|\mathbf{g}\|_*, c)) \\
&\leq \tfrac{1}{\lambda} \min(\tfrac{1}{2}\|\mathbf{g}\|_*^2, c\|\mathbf{g}\|_*).
\end{aligned}$$

By Lems. 10 and 11 and the definition of $a \triangleq \mathrm{diam}(\mathbf{W})$, each summand in our regret bound (5) satisfies

$$\begin{aligned}
\langle \mathbf{g}_t - \tilde{\mathbf{g}}_t, \mathbf{w}_t - \mathbf{w}_{t+1} \rangle - \tfrac{\lambda_{t+1}}{2}\|\mathbf{w}_t - \mathbf{w}_{t+1}\|^2 &\leq \sup_{\mathbf{v} \in \mathbb{R}^d : \|\mathbf{v}\| \leq \min(\frac{1}{\lambda_{t+1}}\|\mathbf{g}_t + \tilde{\mathbf{g}}_{t+1} - \tilde{\mathbf{g}}_t\|_*, a)} \langle \mathbf{g}_t - \tilde{\mathbf{g}}_t, \mathbf{v} \rangle - \tfrac{\lambda_{t+1}}{2}\|\mathbf{v}\|^2 \\
&\leq \min(\tfrac{1}{2\lambda_{t+1}}\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*^2, \tfrac{1}{\lambda_{t+1}}\|\mathbf{g}_t + \tilde{\mathbf{g}}_{t+1} - \tilde{\mathbf{g}}_t\|_* \|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*, a\|\mathbf{g}_t - \tilde{\mathbf{g}}_t\|_*)
\end{aligned}$$

yielding the advertised result. □

D.1. Proof of Lem. 11: Norm-constrained conjugate

By the definition of the dual norm,

$$\begin{aligned}
\sup_{\mathbf{v} \in \mathbb{R}^d : \|\mathbf{v}\| \leq \frac{c}{\lambda}} \langle \mathbf{g}, \mathbf{v} \rangle - \tfrac{\lambda}{2}\|\mathbf{v}\|^2 &= \sup_{a \leq \frac{c}{\lambda}} \sup_{\mathbf{v} \in \mathbb{R}^d : \|\mathbf{v}\| \leq a} \langle \mathbf{g}, \mathbf{v} \rangle - \tfrac{\lambda}{2}a^2 = \sup_{a \leq \frac{c}{\lambda}} a\|\mathbf{g}\|_* - \tfrac{\lambda}{2}a^2 \\
&= \tfrac{1}{\lambda} \min(\|\mathbf{g}\|_*, c)(\|\mathbf{g}\|_* - \tfrac{1}{2}\min(\|\mathbf{g}\|_*, c)) \leq \tfrac{1}{\lambda} c\|\mathbf{g}\|_*.
\end{aligned}$$

We compare to the value of the unconstrained optimization problem to obtain the final inequality:

$$\sup_{a \leq \frac{c}{\lambda}} a\|\mathbf{g}\|_* - \frac{\lambda}{2}a^2 \leq \sup_{a>0} a\|\mathbf{g}\|_* - \frac{\lambda}{2}a^2 = \frac{1}{\lambda}\frac{1}{2}\|\mathbf{g}\|_*^2.$$

E. Tuning-free Learning with Optimism and Delay

We connect DORM to ODFTRL and DORM+ to DOOMD through the following lemma, which recovers the unnormalized iterates $\tilde{\mathbf{w}}_t$ by running ODFTRL or DOOMD on the positive orthant with an appropriate regularizer and surrogate loss.

Lemma 12 (DORM is ODFTRL and DORM+ is DOOMD)**.** *The scaled* DORM *and* DORM+ *iterates* $\tilde{\mathbf{w}}_{t+1}\|\tilde{\mathbf{w}}_{t+1}\|_p^{p-2}$ *are respectively* ODFTRL *and* DOOMD *iterates with* $\mathbf{W} \triangleq \mathbb{R}_+^d$, $\psi(\tilde{\mathbf{w}}) = \frac{1}{2}\|\tilde{\mathbf{w}}\|_p^2$, *and* $\mathbf{g}_{t-D}^{\text{ODFTRL}} = \mathbf{g}_{t-D}^{\text{DOOMD}} = -\mathbf{r}_{t-D}$.

We will use Lem. 12, proved in App. F, to bound the regret of DORM and DORM+, but we first highlight a second remarkable property:

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

Lemma 13 (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+ *plays* $(\mathbf{w}_t)_{t\geq 1}$ *are independent of the choice of* $\lambda > 0$.

Lem. 13, proved in App. G, implies that DORM and DORM+ are *automatically* optimally tuned with respect to λ, even when run with a default value of $\lambda = 1$. Indeed, in App. H we prove the following regret guarantee for DORM and DORM+ with any setting of λ.

Corollary 14 (DORM and DORM+ regret)**.** *Under the assumptions of Lem. 13, for all* $\mathbf{u} \in \triangle_{d-1}$, *the* DORM *and* DORM+ *iterates* $\mathbf{w}_t$ *satisfy*

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

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

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

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

Cor. 14 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 will explore this strategy empirically in Sec. 5.

Cor. 14 also highlights the value of the q parameter in DORM and DORM+. Known regret bounds for standard 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). However, running DORM and DORM+ with 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).

F. Proof of Lem. 12: DORM is ODFTRL and DORM + is DOOMD

Our derivations will make use of several facts about ℓ^p norms, summarized in the next lemma.

Lemma 15 (ℓ^p norm facts)**.** *For $p \in (1, \infty)$, $\psi(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|_p^2$, and any vectors $\mathbf{w}, \mathbf{v} \in \mathbb{R}^d$ and $\tilde{\mathbf{w}}_0 \in \mathbb{R}_+^d$,*

$$\nabla\psi(\mathbf{w}) = \nabla\tfrac{1}{2}\|\mathbf{w}\|_p^2 = \mathrm{sign}(\mathbf{w})|\mathbf{w}|^{p-1}/\|\mathbf{w}\|_p^{p-2} \tag{6}$$

$$\langle \mathbf{w}, \nabla\psi(\mathbf{w})\rangle = \|\mathbf{w}\|_p^2 = 2\psi(\mathbf{w})$$

$$\psi^*(\mathbf{v}) = \sup_{\mathbf{w}\in\mathbb{R}^d} \langle \mathbf{w}, \mathbf{v}\rangle - \psi(\mathbf{w}) = \tfrac{1}{2}\|\mathbf{v}\|_q^2 \quad \textit{for} \quad 1/q = 1 - 1/p \tag{7}$$

$$\nabla\psi^*(\mathbf{v}) = \mathrm{sign}(\mathbf{v})|\mathbf{v}|^{q-1}/\|\mathbf{v}\|_q^{q-2}$$

$$\psi_+^*(\mathbf{v}) = \sup_{\mathbf{w}\in\mathbb{R}_+^d} \langle \mathbf{w}, \mathbf{v}\rangle - \psi(\mathbf{w}) = \sup_{\mathbf{w}\in\mathbb{R}^d} \langle \mathbf{w}, (\mathbf{v})_+\rangle - \psi(\mathbf{w}) = \tfrac{1}{2}\|(\mathbf{v})_+\|_q^2$$

$$\nabla\psi_+^*(\mathbf{v}) = \operatorname*{argmax}_{\mathbf{w}\in\mathbb{R}_+^d} \langle \mathbf{w}, \mathbf{v}\rangle - \psi(\mathbf{w}) = \operatorname*{argmin}_{\mathbf{w}\in\mathbb{R}_+^d} \psi(\mathbf{w}) - \langle \mathbf{w}, \mathbf{v}\rangle = (\mathbf{v})_+^{q-1}/\|(\mathbf{v})_+\|_q^{q-2} \tag{8}$$

$$\begin{aligned}
\min_{\tilde{\mathbf{w}}\in\mathbb{R}_+^d} \mathcal{B}_{\lambda\psi}(\tilde{\mathbf{w}}, \tilde{\mathbf{w}}_0) - \langle \mathbf{v}, \tilde{\mathbf{w}}\rangle &= \lambda(\langle \tilde{\mathbf{w}}_0, \nabla\psi(\tilde{\mathbf{w}}_0)\rangle - \psi(\tilde{\mathbf{w}}_0) - \sup_{\tilde{\mathbf{w}}\in\mathbb{R}_+^d} \langle \tilde{\mathbf{w}}, \nabla\psi(\tilde{\mathbf{w}}_0) + \mathbf{v}/\lambda\rangle - \psi(\tilde{\mathbf{w}})) \\
&= \lambda(\langle \tilde{\mathbf{w}}_0, \nabla\psi(\tilde{\mathbf{w}}_0)\rangle - \psi(\tilde{\mathbf{w}}_0) - \psi_+^*(\nabla\psi(\tilde{\mathbf{w}}_0) + \mathbf{v}/\lambda)) \\
&= \lambda(\psi(\tilde{\mathbf{w}}_0) - \psi_+^*(\nabla\psi(\tilde{\mathbf{w}}_0) + \mathbf{v}/\lambda)) \\
&= \lambda(\psi(\tilde{\mathbf{w}}_0) - \tfrac{1}{2}\|(\nabla\psi(\tilde{\mathbf{w}}_0) + \mathbf{v}/\lambda)_+\|_q^2) \\
&= \lambda(\tfrac{1}{2}\|\tilde{\mathbf{w}}_0\|_p^2 - \tfrac{1}{2}\|(\tilde{\mathbf{w}}_0^{p-1}/\|\tilde{\mathbf{w}}_0\|_p^{p-2} + \mathbf{v}/\lambda)_+\|_q^2).
\end{aligned}$$

Proof. The fact (6) follows from the chain rule as

$$\begin{aligned}
\nabla_j \tfrac{1}{2}\|\mathbf{w}\|_p^2 = \tfrac{1}{2}\nabla_j(\|\mathbf{w}\|_p^p)^{2/p} = \tfrac{1}{p}(\|\mathbf{w}\|_p^p)^{(2/p)-1}\nabla_j\|\mathbf{w}\|_p^p = \tfrac{1}{p}\|\mathbf{w}\|_p^{2-p}\nabla_j \textstyle\sum_{j'=1}^d |\mathbf{w}_{j'}|^p \\
= \tfrac{1}{p}\|\mathbf{w}\|_p^{2-p} p\,\mathrm{sign}(\mathbf{w}_j)|\mathbf{w}_j|^{p-1} = \mathrm{sign}(\mathbf{w}_j)|\mathbf{w}_j|^{p-1}/\|\mathbf{w}\|_p^{p-2}.
\end{aligned}$$

The fact (7) follows from Lem. 11 as $\|\cdot\|_q$ is the dual norm of $\|\cdot\|_p$. □

We now prove each claim in turn.

F.1. DORM is ODFTRL

Fix $p \in (1, 2]$, $\lambda > 0$, and $t \geq 0$. The ODFTRL iterate with hint $\mathbf{h}_{t+1}$, $\mathbf{W} \triangleq \mathbb{R}_+^d$, $\psi(\tilde{\mathbf{w}}) = \frac{1}{2}\|\tilde{\mathbf{w}}\|_p^2$, loss gradients $\mathbf{g}_{1:t-D}^{\text{ODFTRL}} = -\mathbf{r}_{1:t-D}$, and regularization parameter λ takes the form

$$\begin{aligned}
&\operatorname*{argmin}_{\tilde{\mathbf{w}}\in\mathbb{R}_+^d} \lambda\psi(\tilde{\mathbf{w}}) + \langle \tilde{\mathbf{w}}, \mathbf{h}_{t+1} + \mathbf{r}_{1:t-D}\rangle \\
&\quad= \operatorname*{argmin}_{\tilde{\mathbf{w}}\in\mathbb{R}_+^d} \psi(\tilde{\mathbf{w}}) + \langle \tilde{\mathbf{w}}, (\mathbf{h}_{t+1} + \mathbf{r}_{1:t-D})/\lambda\rangle \\
&\quad= ((\mathbf{r}_{1:t-D} + \mathbf{h}_{t+1})/\lambda)_+^{q-1}/\|((\mathbf{r}_{1:t-D} + \mathbf{h}_{t+1})/\lambda)_+\|_q^{q-2} \quad \text{by (8)} \\
&\quad= ((\mathbf{r}_{1:t-D} + \mathbf{h}_{t+1})/\lambda)_+^{q-1}\|((\mathbf{r}_{1:t-D} + \mathbf{h}_{t+1})/\lambda)_+^{q-1}\|_p^{p-2} \quad \text{since } (p-1)(q-1) = 1 \\
&\quad= \tilde{\mathbf{w}}_{t+1}\|\tilde{\mathbf{w}}_{t+1}\|_p^{p-2}
\end{aligned}$$

proving the claim.

F.2. DORM+ is DOOMD

Fix $p \in (1, 2]$ and $\lambda > 0$, and let $(\tilde{\mathbf{w}}_t)_{t\geq 0}$ denote the unnormalized iterates generated by DORM+ with hints $\mathbf{h}_t$, instantaneous regrets $\mathbf{r}_t$, regularization parameter λ, and hyperparameter q. For $p = q/(q-1)$, let $(\bar{\mathbf{w}}_t)_{t\geq 0}$ denote the sequence generated by DOOMD with $\bar{\mathbf{w}}_0 = \mathbf{0}$, hints $\mathbf{h}_t$, $\mathbf{W} \triangleq \mathbb{R}_+^d$, $\psi(\tilde{\mathbf{w}}) = \frac{1}{2}\|\tilde{\mathbf{w}}\|_p^2$, loss gradients $\mathbf{g}_t^{\text{DOOMD}} = -\mathbf{r}_t$, and regularization parameter λ. We proceed by induction to show that, for each t, $\bar{\mathbf{w}}_t = \tilde{\mathbf{w}}_t\|\tilde{\mathbf{w}}_t\|_p^{p-2}$.

Base case By assumption, $\bar{\mathbf{w}}_0 = \mathbf{0} = \tilde{\mathbf{w}}_0\|\tilde{\mathbf{w}}_0\|_p^{p-2}$, confirming the base case.

Inductive step Fix any $t \geq 0$ and assume that for each $s \leq t$, $\bar{\mathbf{w}}_s = \tilde{\mathbf{w}}_s \|\tilde{\mathbf{w}}_s\|_p^{p-2}$. Then, by the definition of DOOMD and our ℓ^p norm facts,

$$\begin{aligned}
\bar{\mathbf{w}}_{t+1} &= \operatorname*{argmin}_{\bar{\mathbf{w}} \in \mathbb{R}_+^d} \langle \mathbf{h}_{t+1} - \mathbf{h}_t + \mathbf{r}_{t-D}, \bar{\mathbf{w}} \rangle + \mathcal{B}_{\lambda\psi}(\bar{\mathbf{w}}, \bar{\mathbf{w}}_t) \\
&= \operatorname*{argmin}_{\bar{\mathbf{w}} \in \mathbb{R}_+^d} \lambda(\psi(\bar{\mathbf{w}}) - \psi(\bar{\mathbf{w}}_t) - \langle \bar{\mathbf{w}} - \bar{\mathbf{w}}_t, \nabla\psi(\bar{\mathbf{w}}_t) \rangle) + \langle \mathbf{h}_{t+1} - \mathbf{h}_t + \mathbf{r}_{t-D}, \bar{\mathbf{w}} \rangle \\
&= \operatorname*{argmin}_{\bar{\mathbf{w}} \in \mathbb{R}_+^d} \psi(\bar{\mathbf{w}}) - \langle \bar{\mathbf{w}}, \nabla\psi(\bar{\mathbf{w}}_t) + (\mathbf{r}_{t-D} - \mathbf{h}_t + \mathbf{h}_{t+1})/\lambda \rangle \\
&= \operatorname*{argmin}_{\bar{\mathbf{w}} \in \mathbb{R}_+^d} \psi(\bar{\mathbf{w}}) - \langle \bar{\mathbf{w}}, \bar{\mathbf{w}}_t^{p-1} / \|\bar{\mathbf{w}}_t\|_p^{p-2} + (\mathbf{r}_{t-D} - \mathbf{h}_t + \mathbf{h}_{t+1})/\lambda \rangle \quad \text{by (6)} \\
&= \operatorname*{argmin}_{\bar{\mathbf{w}} \in \mathbb{R}_+^d} \psi(\bar{\mathbf{w}}) - \langle \bar{\mathbf{w}}, \tilde{\mathbf{w}}_t^{p-1} + (\mathbf{r}_{t-D} - \mathbf{h}_t + \mathbf{h}_{t+1})/\lambda \rangle \quad \text{by the inductive hypothesis} \\
&= (\tilde{\mathbf{w}}_t^{p-1} + (\mathbf{r}_{t-D} - \mathbf{h}_t + \mathbf{h}_{t+1})/\lambda)_+^{q-1} / \|(\tilde{\mathbf{w}}_t^{p-1} + (\mathbf{r}_{t-D} - \mathbf{h}_t + \mathbf{h}_{t+1})/\lambda)_+\|_q^{q-2} \quad \text{by (8)} \\
&= (\tilde{\mathbf{w}}_t^{p-1} + (\mathbf{r}_{t-D} - \mathbf{h}_t + \mathbf{h}_{t+1})/\lambda)_+^{q-1} \|(\tilde{\mathbf{w}}_t^{p-1} + (\mathbf{r}_{t-D} - \mathbf{h}_t + \mathbf{h}_{t+1})/\lambda)_+^{q-1}\|_p^{p-2} \quad \text{since } (p-1)(q-1) = 1 \\
&= \tilde{\mathbf{w}}_{t+1} \|\tilde{\mathbf{w}}_{t+1}\|_p^{p-2},
\end{aligned}$$

completing the inductive step.

G. Proof of Lem. 13: DORM and DORM+ are independent of λ

We will prove the following more general result, from which the stated result follows immediately.

Lemma 16 (DORM and DORM+ are independent of λ)**.** *Consider either* DORM *or* DORM+ *plays* $\tilde{\mathbf{w}}_t$ *as a function of* $\lambda > 0$*, and suppose that for all time points t, the observed gradient* $\mathbf{g}_t$ *and chosen 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 if* $\lambda^{q-1}\tilde{\mathbf{w}}_0$ *is independent of the choice of* $\lambda > 0$*, then so is* $\lambda^{q-1}\tilde{\mathbf{w}}_t$ *for all time points t. As a result,* $\mathbf{w}_t \propto \lambda^{q-1}\tilde{\mathbf{w}}_t$ *is also independent of the choice of* $\lambda > 0$ *at all time points.*

Proof. We prove each result by induction on t.

G.1. Scaled DORM iterates $\lambda^{q-1}\tilde{\mathbf{w}}_t$ are independent of λ

Base case By assumption, $\mathbf{h}_1$ is independent of the choice of $\lambda > 0$. Hence $\lambda^{q-1}\tilde{\mathbf{w}}_1 = (\mathbf{h}_1)_+^{q-1}$ is independent of $\lambda > 0$, confirming the base case.

Inductive step Fix any $t \geq 0$, suppose $\lambda^{q-1}\tilde{\mathbf{w}}_s$ is independent of the choice of $\lambda > 0$ for all $s \leq t$, and consider

$$\lambda^{q-1}\tilde{\mathbf{w}}_{t+1} = (\mathbf{r}_{1:t-D} + \mathbf{h}_{t+1})_+^{q-1}.$$

Since $\mathbf{r}_{1:t-D}$ depends on λ only through $\mathbf{w}_s$ and $\mathbf{g}_s$ for $s \leq t - D$, our λ dependence assumptions for $(\mathbf{g}_s, \mathbf{h}_{s+1})_{s \leq t}$; the fact that, for each s, $\mathbf{w}_s \propto \lambda^{q-1}\tilde{\mathbf{w}}_s$; and our inductive hypothesis together imply that $\lambda^{q-1}\tilde{\mathbf{w}}_{t+1}$ is independent of $\lambda > 0$.

G.2. Scaled DORM+ iterates $\lambda^{q-1}\tilde{\mathbf{w}}_t$ are independent of λ

Base case By assumption, $\lambda^{q-1}\tilde{\mathbf{w}}_0$ is independent of the choice of $\lambda > 0$, confirming the base case.

Inductive step Fix any $t \geq 0$ and suppose $\lambda^{q-1}\tilde{\mathbf{w}}_s$ is independent of the choice of $\lambda > 0$ for all $s \leq t$. Since $(p-1)(q-1) = 1$,

$$\lambda^{q-1}\tilde{\mathbf{w}}_{t+1} = (\lambda\tilde{\mathbf{w}}_t^{p-1} + \mathbf{r}_{t-D} - \mathbf{h}_t + \mathbf{h}_{t+1})_+^{q-1} = ((\lambda^{q-1}\tilde{\mathbf{w}}_t)^{p-1} + \mathbf{r}_{t-D} - \mathbf{h}_t + \mathbf{h}_{t+1})_+^{q-1}.$$

Since $\mathbf{r}_{t-D}$ depends on λ only through $\mathbf{w}_{t-D}$ and $\mathbf{g}_{t-D}$, our λ dependence assumptions for $(\mathbf{g}_s, \mathbf{h}_{s+1})_{s \leq t}$; the fact that, for each $s \leq t$, $\mathbf{w}_s \propto \lambda^{q-1}\tilde{\mathbf{w}}_s$; and our inductive hypothesis together imply that $\lambda^{q-1}\tilde{\mathbf{w}}_{t+1}$ is independent of $\lambda > 0$. $\square$

H. Proof of Cor. 14: DORM and DORM+ regret

Fix any $\lambda > 0$ and $\mathbf{u} \in \triangle_{d-1}$, consider the unnormalized DORM or DORM+ iterates $\tilde{\mathbf{w}}_t$, and define $\bar{\mathbf{w}}_t = \tilde{\mathbf{w}}_t \|\tilde{\mathbf{w}}_t\|_p^{p-2}$ for each t. For either algorithm, we will bound our regret in terms of the surrogate losses

$$\hat{\ell}_t(\tilde{\mathbf{w}}) \triangleq -\langle \mathbf{r}_t, \tilde{\mathbf{w}} \rangle = \langle \mathbf{g}_t, \tilde{\mathbf{w}} \rangle - \langle \tilde{\mathbf{w}}, \mathbf{1} \rangle \langle \mathbf{g}_t, \mathbf{w}_t \rangle$$

defined for $\tilde{\mathbf{w}} \in \mathbb{R}_+^d$. Since $\hat{\ell}_t(\mathbf{u}) = \langle \mathbf{g}_t, \mathbf{u} - \mathbf{w}_t \rangle$, $\hat{\ell}_t(\bar{\mathbf{w}}_t) = 0$, and each ℓ_t is convex, we have

$$\text{Regret}_T(\mathbf{u}) = \sum_{t=1}^T \ell_t(\mathbf{w}_t) - \ell_t(\mathbf{u}) \le \sum_{t=1}^T \langle \mathbf{g}_t, \mathbf{w}_t - \mathbf{u} \rangle = \sum_{t=1}^T \hat{\ell}_t(\bar{\mathbf{w}}_t) - \hat{\ell}_t(\mathbf{u}).$$

For DORM, Lem. 12 implies that $(\bar{\mathbf{w}}_t)_{t \ge 1}$ are ODFTRL iterates, so the ODFTRL regret bound (Thm. 5) and the fact that ψ is 1-strongly convex with respect to $\|\cdot\| = \sqrt{p-1}\|\cdot\|_p$ (see Shalev-Shwartz, 2007, Lemma 17) with $\|\cdot\|_* = \frac{1}{\sqrt{p-1}}\|\cdot\|_q$ imply

$$\text{Regret}_T(\mathbf{u}) \le \tfrac{\lambda}{2}\|\mathbf{u}\|_p^2 + \tfrac{1}{\lambda(p-1)} \sum_{t=1}^T \mathbf{b}_{t,q}^2.$$

Similarly, for DORM+, Lem. 12 implies that $(\bar{\mathbf{w}}_t)_{t \ge 0}$ are DOOMD iterates with $\bar{\mathbf{w}}_0 = \mathbf{0}$, so the DOOMD regret bound (Thm. 6) and the strong convexity of ψ yield

$$\text{Regret}_T(\mathbf{u}) \le \mathcal{B}_{\frac{\lambda}{2}\|\cdot\|_p^2}(\mathbf{u}, \mathbf{0}) + \tfrac{1}{\lambda(p-1)} \sum_{t=1}^T \mathbf{b}_{t,q}^2 = \tfrac{\lambda}{2}\|\mathbf{u}\|_p^2 + \tfrac{1}{\lambda(p-1)} \sum_{t=1}^T \mathbf{b}_{t,q}^2.$$

Since, by Lem. 13, the choice of λ does not impact the iterate sequences played by DORM and DORM+, we may take the infimum over $\lambda > 0$ in these regret bounds. The second advertised inequality comes from the identity $\frac{1}{p-1} = q - 1$ and the norm equivalence relations $\|\mathbf{v}\|_q \le d^{1/q}\|\mathbf{v}\|_\infty$ and $\|\mathbf{v}\|_p \le \|\mathbf{v}\|_1 = 1$ for $\mathbf{v} \in \mathbb{R}^d$, as shown in Lem. 17 below. The final claim follows as

$$\inf_{q' \ge 2} d^{2/q'}(q'-1) = \inf_{q' \ge 2} 2^{2\log_2(d)/q'}(q'-1) \le 2^{2\log_2(d)/(2\log_2(d))}(2\log_2(d) - 1) = 2(2\log_2(d) - 1)$$

since $d > 1$.

Lemma 17 (Equivalence of p-norms). *If $\mathbf{x} \in \mathbb{R}^n$ and $q > q' \ge 1$, then $\|\mathbf{x}\|_q \le \|\mathbf{x}\|_{q'} \le n^{(1/q' - 1/q)}\|\mathbf{x}\|_q$.*

Proof. To show $\|\mathbf{x}\|_q \le \|\mathbf{x}\|_{q'}$ for $q > q' \ge 1$, suppose without loss of generality that $\|\mathbf{x}\|_{q'} = 1$. Then, $\|\mathbf{x}\|_q^q = \sum_{i=1}^n |x_i|^q \le \sum_{i=1}^n |x_i|^{q'} = \|\mathbf{x}\|_{q'}^{q'} = 1$. Hence $\|\mathbf{x}\|_q \le 1 = \|\mathbf{x}\|_{q'}$.

For the inequality $\|\mathbf{x}\|_{q'} \le n^{1/q' - 1/q}\|\mathbf{x}\|_q$, applying Hölder's inequality yields

$$\|\mathbf{x}\|_{q'}^{q'} = \sum_{i=1}^n 1 \cdot |x_i|^{q'} \le \left(\sum_{i=1}^n 1\right)^{1 - \frac{q'}{q}} \left(\sum_{i=1}^n |x_i|^q\right)^{\frac{q'}{q}} = n^{1 - \frac{q'}{q}} \|\mathbf{x}\|_q^{q'},$$

so $\|\mathbf{x}\|_{q'} \le n^{1/q' - 1/q}\|\mathbf{x}\|_q$. $\square$

I. Self-tuned Learning with Optimism and Delay

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)

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

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 $\lambda_t \ge 0$ is a regularization parameter.

Theorem 18 (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*

$$\begin{aligned} &\text{Regret}_T(\mathbf{u}) \le \lambda_T \psi(\mathbf{u}) + \sum_{t=1}^T \min\left(\frac{\mathbf{b}_{t,F}^2}{\lambda_t}, \mathbf{a}_{t,F}\right) \\ &\textit{with} \quad \mathbf{b}_{t,F}^2 \triangleq \min\left(\tfrac{1}{2}\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^t \mathbf{g}_s\|_*^2,\right. \\ &\qquad\qquad \left.\|\mathbf{g}_t\|_* \|\mathbf{h}_t - \textstyle\sum_{s=t-D}^t \mathbf{g}_s\|_*\right) \quad \textit{and} \\ &\qquad \mathbf{a}_{t,F} \triangleq \operatorname{diam}(\mathbf{W}) \min\left(\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^t \mathbf{g}_s\|, \|\mathbf{g}_t\|_*\right). \end{aligned} \tag{9}$$

The proof of this result in App. J builds on a new regret bound for undelayed optimistic adaptive FTRL (OAFTRL) established in the appendix. In the absence of delay ($D = 0$), Thm. 18 strengthens existing regret bounds (Rakhlin & Sridharan, 2013a; Joulani et al., 2017; Mohri & Yang, 2016) for OAFTRL by providing tighter guarantees whenever the hinting error $\frac{1}{2}\|\mathbf{h}_t - \sum_{s=t-D}^{t} \mathbf{g}_t\|_*$ is larger than the gradient magnitude $\|\mathbf{g}_t\|_*$. In the presence of delay, Thm. 18 benefits both from the reduced dependence on the hinting error in the worst case and the ability to exploit accurate hints in the best case. In addition, the bounded-domain factors $\mathbf{a}_{t,F}$ will enable us to design practical λ_t-tuning strategies under delay without any prior knowledge of unobserved gradients.

We now turn to these self-tuning protocols.

I.1. Tuning λ_t with Delayed Upper Bound (DUB)

If we ignore the $\mathbf{a}_{t,F}$ terms in Thm. 18, the non-decreasing online sequence $\lambda_t = \sqrt{\frac{\sum_{s=1}^{t} \mathbf{b}_{s,F}^2}{\sup_{\mathbf{u}\in\mathbf{U}} \psi(\mathbf{u})}}$ 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 workaround is to play the conservative value $\lambda_t = \sqrt{\frac{(D+1)B_0 + \sum_{s=1}^{t-D-1} \mathbf{b}_{s,F}^2}{\sup_{\mathbf{u}\in\mathbf{U}} \psi(\mathbf{u})}}$, where B_0 is a uniform upper bound on the unobserved $\mathbf{b}_{s,F}^2$ terms (Joulani et al., 2016; McMahan & Streeter, 2014). In practice, this requires computing an *a priori* upper bound on any gradient norm that could conceivably arise and often leads to extreme over-regularization. We instead seek an adaptive setting of λ_t that only relies on previously observed losses and can be set easily in real-world settings. The following result, proved in App. K, bounds the regret of one such setting.

Theorem 19 (DUB regret). *Fix $\alpha > 0$, and, for $\mathbf{a}_{t,F}, \mathbf{b}_{t,F}^2$ as in* (9)*, consider the* delayed upper bound *(*DUB*) sequence*

$$\begin{aligned}\alpha\lambda_{t+1} = 2D \max\nolimits_{i\in[t-D]} \mathbf{a}_{i,F} \\ + \sqrt{\textstyle\sum_{i=1}^{t-D} \mathbf{a}_{i,F}^2 + 2\alpha \mathbf{b}_{i,F}^2}.\end{aligned} \tag{DUB}$$

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

$$\begin{aligned}\mathrm{Regret}_T(\mathbf{u}) \le \big(\tfrac{\psi(\mathbf{u})}{\alpha} + 1\big) \\ \big(2D \max\nolimits_{t\in[T]} \mathbf{a}_{t,F} + \sqrt{\textstyle\sum_{t=1}^{T} \mathbf{a}_{t,F}^2 + 2\alpha \mathbf{b}_{t,F}^2}\big).\end{aligned}$$

Before unpacking this bound, we introduce a second, less-conservative strategy that may be preferred in practice.

I.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 $\mathbf{b}_{t,F}^2$) relies heavily on the tightness of that bound. In practice, it is often useful to set λ_t using a tighter bound.

Our next result analyzes a new tuning sequence inspired by the popular AdaHedge algorithm (Erven et al., 2011) and based on a tighter regret bound underlying Thm. 18.

Theorem 20 (AdaHedgeD regret). *Fix $\alpha > 0$, and consider the* delayed AdaHedge-style *(*AdaHedgeD*) sequence*

$$\lambda_{t+1} = \tfrac{1}{\alpha} \textstyle\sum_{s=1}^{t-D} \delta_t \qquad \textit{for} \tag{AdaHedgeD}$$

$$\begin{aligned}&\delta_t \triangleq \min\big(F_{t+1}(\mathbf{w}_t, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t), \langle \mathbf{g}_t, \mathbf{w}_t - \bar{\mathbf{w}}_t\rangle\big)_+ \\ &\quad \textit{with} \quad \bar{\mathbf{w}}_t = \mathrm{argmin}_{\mathbf{w}\in\mathbf{W}} F_{t+1}(\mathbf{w}, \lambda_t) \\ &\quad \textit{and} \quad F_{t+1}(\mathbf{w}, \lambda_t) \triangleq \lambda_t \psi(\mathbf{w}) + \langle \mathbf{g}_{1:t}, \mathbf{w}\rangle.\end{aligned} \tag{10}$$

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

$$\begin{aligned}\mathrm{Regret}_T(\mathbf{u}) \le \big(\tfrac{\psi(\mathbf{u})}{\alpha} + 1\big) \\ \big(2D \max\nolimits_{t\in[T]} \mathbf{a}_{t,F} + \sqrt{\textstyle\sum_{t=1}^{T} \mathbf{a}_{t,F}^2 + 2\alpha \mathbf{b}_{t,F}^2}\big).\end{aligned}$$

Remarkably, Thm. 20 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 in hindsight. Although this regret bound, proved in App. L, is identical to that in Thm. 19, in practice the λ_t values produced by AdaHedgeD can be orders of magnitude smaller than those set by DUB, granting additional adaptivity. We evaluate the practical implications of these two λ_t settings in Sec. 5.

When ψ is bounded on $\mathbf{U}$, we recommend choosing $\alpha = \sup_{\mathbf{u}\in\mathbf{U}} \psi(\mathbf{u})$ so that $\frac{\psi(\mathbf{u})}{\alpha} \leq 1$. 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 second term $\langle \mathbf{g}_t, \mathbf{w}_t - \bar{\mathbf{w}}_t\rangle$ in the $\min$. This non-standard term is central to reducing the impact of delay on our regret bounds.

Previous work in adaptive regularization under delay has leveraged uniform upper bounds on the gradient norms and maximum delay to correct λ_t for the effect of delayed feedback (McMahan & Streeter, 2014; Joulani et al., 2016). Both the DUB and AdaHedgeD algorithm allow for more adaptive settings of λ_t while still bounding regret. As we will see in Sec. 5, the DUB setting is already very conservative and relying on a uniform bound would be even more extreme, motivating the need for algorithms such as AdaHedgeD that seek tighter settings for λ_t.

J. Proof of Thm. 18: ODAFTRL regret

Since ODAFTRL is an instance of OAFTRL with $\tilde{\mathbf{g}}_{t+1} = \mathbf{h}_{t+1} - \sum_{s=t-D+1}^{t} \mathbf{g}_s$, the ODAFTRL result follows immediately from the OAFTRL regret bound, Thm. 7.

K. Proof of Thm. 19: DUB Regret

Fix any $\mathbf{u} \in \mathbf{W}$. By Thm. 18, ODAFTRL admits the regret bound

$$\text{Regret}_T(\mathbf{u}) \leq \lambda_T \psi(\mathbf{u}) + \textstyle\sum_{t=1}^{T} \min(\frac{1}{\lambda_t}\mathbf{b}_{t,F}^2, \mathbf{a}_{t,F}).$$

To control the second term in this bound, we apply the following lemma proved in App. K.1.

Lemma 21 (DUB-style tuning bound)**.** *Fix any $\alpha > 0$ and any non-negative sequences $(a_t)_{t=1}^T$, $(b_t)_{t=1}^T$. If*

$$\Delta_{t+1}^* \triangleq 2\max_{j\leq t-D-1} a_{j-D+1:j} + \sqrt{\textstyle\sum_{i=1}^{t-D} a_i^2 + 2\alpha b_i} \leq \alpha\lambda_{t+1} \quad \textit{for each} \quad t$$

then

$$\textstyle\sum_{t=1}^{T} \min(b_t/\lambda_t, a_t) \leq \Delta_{T+D+1}^* \leq \alpha\lambda_{T+D+1}.$$

Since

$$D\max_{j\in[t-D]} a_j \geq \max_{j\leq t-D-1} a_{j-D+1:j}$$

and $\lambda_T \leq \lambda_{T+D+1}$, the result now follows by setting $a_t = \mathbf{a}_{t,F}$ and $b_t = \mathbf{b}_{t,F}^2$, so that

$$\text{Regret}_T(\mathbf{u}) \leq \lambda_T\psi(\mathbf{u}) + \alpha\lambda_{T+D+1} \leq (\psi(\mathbf{u}) + \alpha)\lambda_{T+D+1}.$$

K.1. Proof of Lem. 21: DUB-style tuning bound

We prove the claim

$$\Delta_t \triangleq \textstyle\sum_{i=1}^{t} \min(b_i/\lambda_i, a_i) \leq \Delta_{t+D+1}^* \leq \alpha\lambda_{t+D+1}$$

by induction on t.

Base case For $t \in [D+1]$,

$$\textstyle\sum_{i=1}^{t} \min(b_i/\lambda_i, a_i) \leq a_{1:t-1} + a_t \leq 2\max_{j\leq t-1} a_{j-D+1:j} + \sqrt{\sum_{i=1}^{t} a_i^2 + 2\alpha b_i} = \Delta_{t+D+1}^* \leq \alpha\lambda_{t+D+1}$$

confirming the base case.

Inductive step Now fix any $t+1 \geq D+2$ and suppose that

$$\Delta_i \leq \Delta^*_{i+D+1} \leq \alpha\lambda_{i+D+1}$$

for all $1 \leq i \leq t$. We apply this inductive hypothesis to deduce that, for each $0 \leq i \leq t$,

$$\begin{aligned}
\Delta^2_{i+1} - \Delta^2_i &= (\Delta_i + \min(b_{i+1}/\lambda_{i+1}, a_{i+1}))^2 - \Delta^2_i = 2\Delta_i \min(b_{i+1}/\lambda_{i+1}, a_{i+1}) + \min(b_{i+1}/\lambda_{i+1}, a_{i+1})^2 \\
&= 2\Delta_{i-D} \min(b_{i+1}/\lambda_{i+1}, a_{i+1}) + 2(\Delta_i - \Delta_{i-D}) \min(b_{i+1}/\lambda_{i+1}, a_{i+1}) + \min(b_{i+1}/\lambda_{i+1}, a_{i+1})^2 \\
&= 2\Delta_{i-D} \min(b_{i+1}/\lambda_{i+1}, a_{i+1}) + 2\sum_{j=i-D+1}^{i} \min(b_j/\lambda_j, a_j) \min(b_{i+1}/\lambda_{i+1}, a_{i+1}) + \min(b_{i+1}/\lambda_{i+1}, a_{i+1})^2 \\
&\leq 2\alpha\lambda_{i+1} \min(b_{i+1}/\lambda_{i+1}, a_{i+1}) + 2a_{i-D+1:i} \min(b_{i+1}/\lambda_{i+1}, a_{i+1}) + a^2_{i+1} \\
&\leq 2\alpha b_{i+1} + a^2_{i+1} + 2a_{i-D+1:i} \min(b_{i+1}/\lambda_{i+1}, a_{i+1}).
\end{aligned}$$

Now, we sum this inequality over $i = 0, \dots, t$, to obtain

$$\begin{aligned}
\Delta^2_{t+1} &\leq \textstyle\sum_{i=0}^{t}(2\alpha b_{i+1} + a^2_{i+1}) + 2\sum_{i=0}^{t} a_{i-D+1:i} \min(b_{i+1}/\lambda_{i+1}, a_{i+1}) \\
&= \textstyle\sum_{i=1}^{t+1}(2\alpha b_i + a^2_i) + 2\sum_{i=1}^{t+1} a_{i-D:i-1} \min(b_i/\lambda_i, a_i) \\
&\leq \textstyle\sum_{i=1}^{t+1}(a^2_i + 2\alpha b_i) + 2\max_{j\leq t} a_{j-D+1:j} \sum_{i=1}^{t+1} \min(b_i/\lambda_i, a_i) \\
&= \textstyle\sum_{i=1}^{t+1}(a^2_i + 2\alpha b_i) + 2\Delta_{t+1} \max_{j\leq t} a_{j-D+1:j}.
\end{aligned}$$

Solving this quadratic inequality and applying the triangle inequality, we have

$$\begin{aligned}
\Delta_{t+1} &\leq \textstyle\max_{j\leq t} a_{j-D+1:j} + \frac{1}{2}\sqrt{(2\max_{j\leq t} a_{j-D+1:j})^2 + 4\sum_{i=1}^{t+1} a^2_i + 2\alpha b_i} \\
&\leq \textstyle 2\max_{j\leq t} a_{j-D+1:j} + \sqrt{\sum_{i=1}^{t+1} a^2_i + 2\alpha b_i} = \Delta^*_{t+D+2} \leq \alpha\lambda_{t+D+2}.
\end{aligned}$$

L. Proof of Thm. 20: AdaHedgeD Regret

Fix any $\mathbf{u} \in \mathbf{W}$. Since the AdaHedgeD regularization sequence $(\lambda_t)_{t\geq 1}$ is non-decreasing, Thm. 7 gives the regret bound

$$\text{Regret}_T(\mathbf{u}) \leq \lambda_T\psi(\mathbf{u}) + \textstyle\sum_{t=1}^{T} \delta_t = \lambda_T\psi(\mathbf{u}) + \alpha\lambda_{T+D+1} \leq (\psi(\mathbf{u}) + \alpha)\lambda_{T+D+1},$$

and the proof of Thm. 7 gives the upper estimate (3):

$$\delta_t \leq \min\left(\frac{\mathbf{b}^2_{t,F}}{\lambda_t}, \mathbf{a}_{t,F}\right) \quad \text{for all} \quad t \in [T]. \tag{11}$$

Hence, it remains to bound λ_{T+D+1}. Our argument will make use of the following lemma, proved in App. L.1, which demonstrates how to bound one series in terms of another with index-shifted coefficients.

Lemma 22 (Index-shifted series bound). *If $a_{t+D} \geq a_t$ for each $t \geq 1$, then*

$$\textstyle\sum_{t=1}^{T} b_t a_{t+D} \leq \max_{t\in[T]} b_t(\sum_{t=T+1}^{T+D} a_t - \sum_{t=1}^{D} a_t) + \sum_{t=1}^{T} b_t a_t.$$

Since $\lambda_1 = \cdots = \lambda_{D+1} = 0$ and $\alpha(\lambda_{t+1} - \lambda_t) = \delta_{t-D}$ for $t \geq D+1$,

$$\begin{aligned}
\alpha\lambda^2_{T+D+1} &= \textstyle\sum_{t=1}^{T+D} \alpha(\lambda^2_{t+1} - \lambda^2_t) = \sum_{t=D+1}^{T+D}\big(\alpha(\lambda_{t+1} - \lambda_t)^2 + 2\alpha(\lambda_{t+1} - \lambda_t)\lambda_t\big) \\
&= \textstyle\sum_{t=1}^{T}\big(\delta^2_t/\alpha + 2\delta_t\lambda_{t+D}\big) \quad \text{by the definition of } \lambda_{t+1} \\
&\leq \textstyle\sum_{t=1}^{T}\big(\mathbf{a}^2_{t,F}/\alpha + 2\delta_t\lambda_{t+D}\big) \quad \text{by (11)} \\
&\leq \textstyle\sum_{t=1}^{T} \mathbf{a}^2_{t,F}/\alpha + \sum_{t=1}^{T} 2\delta_t\lambda_t + \sum_{t=T+1}^{T+D} \lambda_t \max_{t\in[T]} 2\delta_t \quad \text{by Lem. 22} \\
&\leq \textstyle\sum_{t=1}^{T}\big(\mathbf{a}^2_{t,F}/\alpha + 2\mathbf{b}^2_{t,F}\big) + 2D\lambda_{T+D+1} \max_{t\in[T]} \delta_t \quad \text{since } (\lambda_t)_{t\geq 1} \text{ non-decreasing.}
\end{aligned}$$

Solving the above quadratic inequality for λ_{T+D+1}, we find

$$\begin{aligned}
\alpha\lambda_{T+D+1} &\leq D \max_{t\in[T]} \delta_t + \tfrac{1}{2}\sqrt{4D^2 \max_{t\in[T]} \delta_t^2 + \sum_{t=1}^T 4\mathbf{a}_{t,F}^2 + 8\alpha\mathbf{b}_{t,F}^2} \\
&\leq 2D \max_{t\in[T]} \delta_t + \sqrt{\sum_{t=1}^T \mathbf{a}_{t,F}^2 + 2\alpha\mathbf{b}_{t,F}^2} \\
&\leq 2D \max_{t\in[T]} \mathbf{a}_{t,F} + \sqrt{\sum_{t=1}^T \mathbf{a}_{t,F}^2 + 2\alpha\mathbf{b}_{t,F}^2}.
\end{aligned}$$

L.1. Proof of Lem. 22: Index-shifted series bound

For each t, since $a_{t+D} \geq a_t$,

$$b_t a_{t+D} = b_t a_t + (a_{t+D} - a_t) b_t \leq b_t a_t + (a_{t+D} - a_t) \max_{t\in[T]} b_t.$$

Now, summing over $t \in [T]$, we have

$$\begin{aligned}
\sum_{t=1}^T b_t a_{t+D} &\leq \max_{t\in[T]} b_t (\sum_{t=1}^T a_{t+D} - \sum_{t=1}^T a_t) + \sum_{t=1}^T b_t a_t \\
&= \max_{t\in[T]} b_t (\sum_{t=D+1}^{T+D} a_t - \sum_{t=1}^T a_t) + \sum_{t=1}^T b_t a_t \\
&= \max_{t\in[T]} b_t (\sum_{t=T+1}^{T+D} a_t - \sum_{t=1}^D a_t) + \sum_{t=1}^T b_t a_t.
\end{aligned}$$

M. Experiment Details

M.1. Subseasonal Forecasting Application

We apply the online learning techniques developed in this paper to the problem of adaptive ensembling for subseasonal weather forecasting. Subseasonal forecasting is the problem predicting meteorological variables, often temperature and precipitation, 2-6 weeks in advance. These mid-range forecasts are critical for managing water resources and mitigating wildfires, droughts, floods, and other extreme weather events (Hwang et al., 2019). However, the subseasonal forecasting task is notoriously difficult due to the joint influences of short-term initial conditions and long-term boundary conditions (White et al., 2017).

To improve subseasonal weather forecasting capabilities, the US Department of Reclamation launched the Sub-Seasonal Climate Forecast Rodeo competition (Nowak et al., 2020), a yearlong real-time forecasting competition for the Western United States. Our experiments are based on Orenstein (2021), a codebase of public subseasonal forecasting models including both physics-based numerical forecasts and deep learning models. These models were developed for the subseasonal forecasting challenge and make semimonthly forecasts for the contest period (19 October 2019- 29 September 2020).

To expand our evaluation beyond the subseasonal forecasting competition, we followed the instructions provided by Orenstein (2021) to generate input model predictions in analogous year-long periods (26 semimonthly dates starting from the last Wednesday in October) for each year in 2010-2020. For each forecast date t, models in Orenstein (2021) are trained on data available at time t and model hyper-parameters are tuned to optimize average RMSE loss on the 3-year period preceding the forecast date t.

M.2. Problem Definition

Denote the set of $d = 6$ input models $\{\mathcal{M}_1, \dots \mathcal{M}_d\}$. On each semimonthly forecast date, each model $\mathcal{M}_i$ makes a prediction for each of two meteorological variables (cumulative precipitation and average temperature over 14 days) and two forecasting horizons (3-4 weeks and 5-6 weeks). For the 3-4 week and 5-6 horizons respectively, the forecaster experiences a delay of $D = 2$ and $D = 3$ forecasts. Each model makes a total of $T = 26$ semimonthly forecasts for these four tasks.

At each time t, each input model $\mathcal{M}_i$ produces a prediction at $G = 514$ gridpoints in the Western United States: $\mathbf{x}_{t,i}^c \in \mathbb{R}^G = \mathcal{M}_i(t)$ for task c at time t. Let $\mathbf{X}_t^c \in \mathbb{R}^{G\times d}$ be the matrix containing each input model's predictions as columns. The true meterological outcome for task c is $\mathbf{y}_t^c \in \mathbb{R}^G$. As online learning is performed for each task separately, we drop the task superscript c in the following.

At each timestep, the online learner makes a forecast prediction $\hat{\mathbf{y}}_t$ by playing $\mathbf{w}_t \in \mathbf{W} = \triangle_{d-1}$, corresponding to a convex combination of the individual models: $\hat{\mathbf{y}}_t = \mathbf{X}_t \mathbf{w}_t$. The learner then incurs a loss for the play $\mathbf{w}_t$ according to the root mean squared (RMSE) error over the geography of interest:

$$\ell_t(\mathbf{w}_t) = \frac{1}{\sqrt{G}} \|\mathbf{y}_t - \mathbf{X}_t \mathbf{w}_t\|_2,$$
$$\partial \ell_t(\mathbf{w}_t) \ni \mathbf{g}_t = \begin{cases} \frac{\mathbf{X}_t^\top (\mathbf{X}_t \mathbf{w}_t - \mathbf{y}_t)}{\sqrt{G}\|\mathbf{X}_t \mathbf{w}_t - \mathbf{y}_t\|_2} & \text{if} \quad \mathbf{X}_t \mathbf{w}_t - \mathbf{y}_t \neq \mathbf{0} \\ \mathbf{0} & \text{if} \quad \mathbf{X}_t \mathbf{w}_t - \mathbf{y}_t = \mathbf{0} \end{cases}$$

Our objective for the subseasonal forecasting application is to produce an adaptive ensemble forecast that competes with the best input model over the yearlong period. Hence, in our evaluation, we take the competitor set to be the set of individual models $\mathbf{U} = \{\mathbf{e}_i : i \in [d]\}$.

N. Algorithmic Details

N.1. ODAFTRL with AdaHedgeD and DUB tuning

The AdaHedgeD and DUB algorithms presented in the experiments are implementations of ODAFTRL with a negative entropy regularizer $\psi(\mathbf{w}) = \sum_{j=1}^d \mathbf{w}_j \ln \mathbf{w}_j + \ln d$, which is 1-strongly convex with respect to the norm $\|\cdot\|_1$ (Shalev-Shwartz, 2007, Lemma 16) with dual norm $\|\cdot\|_\infty$. Each algorithm optimizes over the simplex and competes with the simplex: $\mathbf{W} = \mathbf{U} = \triangle_{d-1}$. We choose $\alpha^2 = \sup_{\mathbf{u} \in \mathbf{U}} \psi(\mathbf{u}) = \ln(d)$. In the following, define $\psi_t \triangleq \lambda_t \psi$ for $\lambda_t \geq 0$. Our derivations of the update equations for AdaHedgeD and DUB make use of the follow properties of the negative entropy regularizer, proved in App. N.3.

Lemma 23 (Negative entropy properties)**.** *The negative entropy regularizer $\psi(\mathbf{w}) = \sum_{j=1}^d \mathbf{w}_j \ln \mathbf{w}_j + \ln d$ with $\psi_t = \lambda_t \psi$ for $\lambda_t \geq 0$ satisfies the following properties on the simplex $\mathbf{W} = \triangle_{d-1}$.*

$$\psi^*_{\mathbf{W}}(\theta) \triangleq \sup_{\mathbf{w} \in \mathbf{W}} \langle \mathbf{w}, \theta \rangle - \psi(\mathbf{w}) = \ln\Big(\textstyle\sum_{j=1}^d \exp(\theta_j)\Big) - \ln d,$$
$$(\lambda\psi)^*_{\mathbf{W}}(\theta) \triangleq \sup_{\mathbf{w} \in \mathbf{W}} \langle \mathbf{w}, \theta \rangle - \lambda\psi(\mathbf{w}) = \begin{cases} \lambda \psi^*_{\mathbf{W}}(\theta/\lambda) = \lambda \ln(\sum_{j=1}^d \exp(\theta_j/\lambda)) - \lambda \ln d, & \text{if } \lambda > 0 \\ \max_{j \in [d]} \theta_j & \text{if } \lambda = 0 \end{cases},$$
$$\mathbf{w}^*(\theta, \lambda) \triangleq \begin{cases} \frac{\exp(\theta/\lambda)}{\sum_{j=1}^d \exp(\theta_j/\lambda)} & \text{if } \lambda > 0 \\ \frac{\mathbb{I}[\theta = \max_j \theta_j]}{\sum_{k \in [d]} \mathbb{I}[\theta_k = \max_j \theta_j]} & \text{if } \lambda = 0 \end{cases} \in \operatorname{argmin}_{\mathbf{w} \in \mathbf{W}} \lambda\psi(\mathbf{w}) - \langle \mathbf{w}, \theta \rangle \subseteq \partial(\lambda\psi)^*_{\mathbf{W}}(\theta).$$

Our next corollary concerning optimal ODAFTRL objectives follows directly from Lem. 23.

Corollary 24 (Optimal ODAFTRL objectives)**.** *Instantiate the notation of Lem. 23 and define the functions $F_t(\mathbf{w}, \lambda) \triangleq \lambda\psi(\mathbf{w}) + \langle \mathbf{g}_{1:t-1}, \mathbf{w} \rangle$. for $\mathbf{w} \in \mathbf{W}$. Then*

$$-(\lambda\psi)^*_{\mathbf{W}}(-(\mathbf{g}_{1:t-1} + \mathbf{h})) = \inf_{\mathbf{w} \in \mathbf{W}} F_t(\mathbf{w}, \lambda) + \langle \mathbf{h}, \mathbf{w} \rangle \quad \textit{and}$$
$$\mathbf{w}^*(-(\mathbf{g}_{1:t-1} + \mathbf{h}), \lambda) = \operatorname{argmin}_{\mathbf{w} \in \mathbf{W}} F_t(\mathbf{w}, \lambda) + \langle \mathbf{h}, \mathbf{w} \rangle.$$

Using Lem. 23 and Cor. 24, we can derive an expression, proved in App. N.4, for the AdaHedgeD δ_t updates.

Proposition 25 (AdaHedgeD δ_t)**.** *The δ_t update for* AdaHedgeD *is given by the following expressions. Define $c_* = \max_{j : \mathbf{w}_{t,j} \neq 0} \mathbf{h}_{t,j} - \mathbf{g}_{t-D:t,j}$. If $\lambda_t > 0$,*

$$\begin{aligned} \delta_t &\triangleq \min\big(F_{t+1}(\mathbf{w}_t, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t), \langle \mathbf{g}_t, \mathbf{w}_t - \bar{\mathbf{w}}_t \rangle\big)_+ \quad \textit{by definition } (10) \\ &= \min\big(\lambda_t \ln(\textstyle\sum_{j \in [d]} \mathbf{w}_{t,j} \exp((\mathbf{h}_{t,j} - \mathbf{g}_{t-D:t,j})/\lambda_t)) + \langle \mathbf{g}_{t-D:t} - \mathbf{h}_t, \mathbf{w}_t \rangle, \langle \mathbf{g}_t, \mathbf{w}_t - \bar{\mathbf{w}}_t \rangle\big)_+ \\ &= \min\big(\lambda_t \ln(\textstyle\sum_{j \in [d]} \mathbf{w}_{t,j} \exp((\mathbf{h}_{t,j} - \mathbf{g}_{t-D:t,j} - c_*)/\lambda_t)) + \langle \mathbf{g}_{t-D:t} - \mathbf{h}_t, \mathbf{w}_t \rangle + c_*, \langle \mathbf{g}_t, \mathbf{w}_t - \bar{\mathbf{w}}_t \rangle\big)_+. \end{aligned}$$

for $\bar{\mathbf{w}}_t = \operatorname{argmin}_{\mathbf{w} \in \mathbf{W}} F_{t+1}(\mathbf{w}, \lambda_t) = \frac{\exp(-\mathbf{g}_{1:t}/\lambda_t)}{\sum_{j=1}^d \exp(-\mathbf{g}_{1:t,j}/\lambda_t)}$ as in Cor. 24. If $\lambda_t = 0$,

$$\delta_t = \langle \mathbf{g}_{1:t}, \mathbf{w}_t \rangle - \min_{j \in [d]} \mathbf{g}_{1:t,j}.$$

Leveraging these results, we present the pseudocode for the AdaHedgeD and DUB instantiations of ODAFTRL in Algorithm 1.

Algorithm 1 ODAFTRL with $\mathbf{W} = \triangle_{d-1}$, $\psi(\mathbf{w}) = \sum_{j=1}^{d} \mathbf{w}_j \ln \mathbf{w}_j + \ln(d)$, delay $D \geq 0$, and tuning strategy `tuning`

Parameter $\alpha^2 = \sup_{\mathbf{u} \in \triangle_{d-1}} \psi(\mathbf{u}) = \ln(d)$
Initial regularization weight: $\lambda_0 = 0$
if `tuning` is DUB **then**
    Initial regularization sum: $\Delta_0 = 0$
    Initial maximum: $\mathbf{a}^{\max} = 0$
end if
Initial gradient sum: $\mathbf{g}_{1:1} = \mathbf{0} \in \mathbb{R}^d$
Dummy losses and iterates: $\mathbf{g}_{-D} = \cdots = \mathbf{g}_0 = \mathbf{0} \in \mathbb{R}^d$, $\mathbf{w}_{-D} = \cdots = \mathbf{w}_0 = \mathbf{0} \in \mathbb{R}^d$
for $t = 1, \ldots, T$ **do**
    Receive hint $\mathbf{h}_t \in \mathbb{R}^d$
    Output $\mathbf{w}_t = \operatorname{argmin}_{\mathbf{w} \in \mathbf{W}} F_{t-D}(\mathbf{w}, \lambda_t) + \langle \mathbf{h}_t, \mathbf{w} \rangle$ as in Cor. 24
    Receive $\mathbf{g}_{t-D} \in \mathbb{R}^d$ and pay $\langle \mathbf{g}_{t-D}, \mathbf{w}_{t-D} \rangle$
    Update gradient sum $\mathbf{g}_{1:t-D} = \mathbf{g}_{1:t-D-1} + \mathbf{g}_{t-D}$
    if `tuning` is AdaHedgeD **then**
        Compute the auxiliary sequence: $\bar{\mathbf{w}}_{t-D} = \operatorname{argmin}_{\mathbf{w} \in \mathbf{W}} F_{t-D+1}(\mathbf{w}, \lambda_{t-D})$ as in Cor. 24
        Compute the drift term (10): $\delta_{t-D}^{(1)} = \langle \mathbf{g}_{t-D}, \mathbf{w}_{t-D} - \bar{\mathbf{w}}_{t-D} \rangle$
        Compute the auxiliary regret term $\delta_{t-D}^{(2)} = F_{t-D+1}(\mathbf{w}_{t-D}, \lambda_{t-D}) - F_{t-D+1}(\bar{\mathbf{w}}_{t-D}, \lambda_{t-D})$ as in Prop. 25
        Update $\lambda_{t+1} = \lambda_t + \frac{1}{\alpha^2} \min(\delta_{t-D}^{(1)}, \delta_{t-D}^{(2)})_+$ as in (10)
    else if `tuning` is DUB **then**
        Compute $\mathbf{a}_{t-D,F} = 2 \min\left(\|\mathbf{g}_{t-D}\|_\infty, \|\mathbf{h}_{t-D} - \sum_{s=t-2D}^{t-D} \mathbf{g}_s\|_\infty \right)$ as in (9)
        Compute $\mathbf{b}_{t-D,F}^2 = \min\left(\frac{1}{2} \|\mathbf{h}_{t-D} - \sum_{s=t-2D}^{t-D} \mathbf{g}_s\|_\infty^2, \|\mathbf{g}_{t-D}\|_\infty \|\mathbf{h}_{t-D} - \sum_{s=t-2D}^{t-D} \mathbf{g}_s\|_\infty \right)$ as in (9)
        Update $\Delta_{t+1} = \Delta_t + \mathbf{a}_{t-D,F}^2 + 2\mathbf{b}_{t-D,F}$
        Update maximum $\mathbf{a}^{\max} = \max(\mathbf{a}^{\max}, \mathbf{a}_{t-D,F})$
        Update $\lambda_{t+1} = 2D\mathbf{a}^{\max} + \sqrt{\Delta_{t+1}}$ as in DUB
    end if
end for

N.2. DORM and DORM+

The DORM and DORM+ algorithms presented in the experiments are implementations of ODAFTRL and DOOMD respectively that play iterates in $\mathbf{W} \triangleq \triangle_{d-1}$ using the default value $\lambda = 1$. Both algorithms use a p-norm regularizer $\psi(\mathbf{w}) = \frac{1}{2}\|\tilde{\mathbf{w}}\|_p^2$, which is 1-strongly convex with respect to $\|\cdot\| = \sqrt{p-1}\|\cdot\|_p$ (see Shalev-Shwartz, 2007, Lemma 17) with $\|\cdot\|_* = \frac{1}{\sqrt{p-1}}\|\cdot\|_q$. For the paper experiments, we choose the optimal value $q = \inf_{q' \geq 2} d^{2/q'}(q'-1)$ to obtain $\ln(d)$ scaling in the algorithm regret; for $d = 6$, $p = q = 2$. The update equations for each algorithm are given in the main text by DORM and DORM+ respectively. The optimistic hinters provide delayed negative regret hints, e.g., $\mathbf{h}_t = -D\mathbf{r}_{t-D-1}$ for `prev_g`.

N.3. Proof of Lem. 23: Negative entropy properties

The expression of the Fenchel conjugate for $\lambda > 0$ is derived by solving an appropriate constrained convex optimization problem for $\mathbf{w} = \triangle_{d-1}$, as shown in Orabona (2019, Section 6.6). The value of $\mathbf{w}^*(\theta, \lambda) \in \partial(\lambda\psi)^*_{\mathbf{W}}(\theta)$ uses the properties of the Fenchel conjugate (Rockafellar, 1970; Orabona, 2019, Theorem 5.5) and is shown in Orabona (2019, Theorem 6.6).

N.4. Proof of Prop. 25: AdaHedgeD δ_t

For $\lambda_t > 0$, the first term in the $\min$ of AdaHedgeD's δ_t setting is derived as follows:

$$
\begin{aligned}
\delta_t^{(1)} &\triangleq F_{t+1}(\mathbf{w}_t, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t) \quad \text{by definition (10)} \\
&= F_{t-D}(\mathbf{w}_t, \lambda_t) + \langle \mathbf{h}_t, \mathbf{w}_t \rangle + \langle \mathbf{g}_{t-D:t} - \mathbf{h}_t, \mathbf{w}_t \rangle - \inf_{\mathbf{w} \in \mathbf{W}} F_{t+1}(\mathbf{w}, \lambda_t) \quad \text{by definition of } \bar{\mathbf{w}}_t \\
&= F_{t-D}(\mathbf{w}_t, \lambda_t) + \langle \mathbf{h}_t, \mathbf{w}_t \rangle + \langle \mathbf{g}_{t-D:t} - \mathbf{h}_t, \mathbf{w}_t \rangle + \lambda_t \psi^*_{\mathbf{W}}(-\mathbf{g}_{1:t}/\lambda_t) \quad \text{by Cor. 24} \\
&= \lambda_t \psi^*_{\mathbf{W}}(-\mathbf{g}_{1:t}/\lambda_t) - \lambda_t \psi^*_{\mathbf{W}}((-\mathbf{h}_t - \mathbf{g}_{1:t-D-1})/\lambda_t) + \langle \mathbf{g}_{t-D:t} - \mathbf{h}_t, \mathbf{w}_t \rangle \\
&\qquad \text{because } \mathbf{w}_t \in \operatorname{argmin}_{\mathbf{w} \in \mathbf{W}} F_{t-D}(\mathbf{w}_t, \lambda_t) + \langle \mathbf{h}_t, \mathbf{w}_t \rangle \\
&= \lambda_t(\ln(\textstyle\sum_{j=1}^d \exp(-\mathbf{g}_{1:t,j}/\lambda_t)) - \lambda_t(\ln(\textstyle\sum_{j=1}^d \exp((-\mathbf{g}_{1:t-D-1,j} - \mathbf{h}_{t,j})/\lambda_t)) + \langle \mathbf{g}_{t-D:t} - \mathbf{h}_t, \mathbf{w}_t \rangle \quad \text{by Lem. 23} \\
&= \lambda_t \ln\left(\textstyle\sum_{j=1}^d \frac{\exp(-\mathbf{g}_{1:t,j}/\lambda_t)}{\sum_{j=1}^d \exp((-\mathbf{g}_{1:t-D-1,j} - \mathbf{h}_{t,j})/\lambda_t)} \right) + \langle \mathbf{g}_{t-D:t} - \mathbf{h}_t, \mathbf{w}_t \rangle \\
&= \lambda_t \ln\left(\textstyle\sum_{j=1}^d \frac{\exp((-\mathbf{g}_{1:t-D-1,j} - \mathbf{h}_{t,j})/\lambda_t) \exp((\mathbf{h}_{t,j} - \mathbf{g}_{t-D:t,j})/\lambda_t)}{\sum_{j=1}^d \exp((-\mathbf{g}_{1:t-D-1,j} - \mathbf{h}_{t,j})/\lambda_t)} \right) + \langle \mathbf{g}_{t-D:t} - \mathbf{h}_t, \mathbf{w}_t \rangle \\
&= \lambda_t \ln\left(\textstyle\sum_{j=1}^d \mathbf{w}_{t,j} \exp((\mathbf{h}_{t,j} - \mathbf{g}_{t-D:t,j})/\lambda_t) \right) + \langle \mathbf{g}_{t-D:t} - \mathbf{h}_t, \mathbf{w}_t \rangle \quad \text{by the expression for } \mathbf{w}_t \text{ in Cor. 24}
\end{aligned}
$$

The result follows by expanding the second term of the $\min$ from the definition of $\bar{\mathbf{w}}_t$ and Cor. 24.

If $\lambda_t = 0$, we have

$$
\begin{aligned}
\delta_t^{(1)} &\triangleq F_{t+1}(\mathbf{w}_t, \lambda_t) - F_{t+1}(\bar{\mathbf{w}}_t, \lambda_t) \quad \text{by definition (10)} \\
&= \langle \mathbf{g}_{1:t}, \mathbf{w}_t \rangle - \inf_{\mathbf{w} \in \mathbf{W}} F_{t+1}(\mathbf{w}, \lambda_t) \quad \text{by definition of } \bar{\mathbf{w}}_t \\
&= \langle \mathbf{g}_{1:t}, \mathbf{w}_t \rangle - \min_{j \in [d]} \mathbf{g}_{1:t,j} \quad \text{by Cor. 24.}
\end{aligned}
$$

References

Agarwal, A. and Duchi, J. C. Distributed delayed stochastic optimization. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pp. 5451–5452. IEEE, 2012.

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.

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. *Advances in Neural Information Processing Systems*, 24: 1656–1664, 2011.

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.

Hwang, J., Orenstein, P., Cohen, J., Pfeiffer, K., and Mackey, L. Improving subseasonal forecasting in the western us 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. *arXiv preprint arXiv:1709.02726*, 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. *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, pp. 2294–2302, 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 *Advances in Neural Information Processing Systems*, pp. 2915–2923, 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. URL https://proceedings.neurips.cc/paper/2003/file/feecee9f1643651799ede2740927317a-Paper.pdf.

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.

Orenstein, P. Subseasonal forecasting models, 2021. URL http://w3.impa.br/~pauloo/subseasonal_forecasting/code.zip.

Quanrud, K. and Khashabi, D. Online learning with adversarial delays. In *Advances in neural information processing systems*, pp. 1270–1278, 2015.

Rakhlin, A. and Sridharan, K. Online learning with predictable sequences. In *Conference on Learning Theory*, pp. 993–1019, 2013a.

Rakhlin, S. and Sridharan, K. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, pp. 3066–3074, 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. et al. 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 *Artificial Intelligence and Statistics*, 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 *Advances in Neural Information Processing Systems 28*, 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 *NIPS*, volume 7, pp. 1729, 2007.

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

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

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 *NIPS*, volume 7, pp. 1729, 2007.