Luckiness in Multiscale Online Learning

Muriel Felipe Pérez-Ortiz
Centrum Wiskunde & Informatica (CWI)
muriel.perez@cwi.nl

Wouter M. Koolen
CWI and University of Twente
wmkoolen@cwi.nl

Abstract

Algorithms for full-information online learning are classically tuned to minimize their worst-case regret. Modern algorithms additionally provide tighter guarantees outside the adversarial regime, most notably in the form of constant pseudoregret bounds under statistical margin assumptions. We investigate the multiscale extension of the problem where the loss ranges of the experts are vastly different. Here, the regret with respect to each expert needs to scale with its range, instead of the maximum overall range. We develop new multiscale algorithms, tuning schemes and analysis techniques to show that worst-case robustness and adaptation to easy data can be combined at a negligible cost. We further develop an extension with optimism and apply it to solve multiscale two-player zero-sum games. We demonstrate experimentally the superior performance of our scale-adaptive algorithm and discuss the subtle relationship of our results to Freund's 2016 open problem.

1 Introduction

The abstract problem of *online prediction with expert advice* [Littlestone and Warmuth, 1994, Freund and Schapire, 1997] is of fundamental importance in computational learning theory. Efficient and optimal algorithms for solving it have a substantial impact on various problems in general online convex optimization [Hazan, 2019], online model selection [Foster et al., 2017], boosting [Freund and Schapire, 1997], and maximal probabilistic inequalities [Rakhlin and Sridharan, 2017], to name a few. Concretely, a decision maker chooses among experts' advices sequentially, and the environment assigns each advice a scalar loss. If all losses have the same numerical range $[-\sigma, \sigma]$, the situation is well understood. Indeed, Freund and Schapire [1997] showed that, for K experts and t rounds, the Hedge algorithm guarantees the minimax regret (defined below) $\sigma\sqrt{2t \ln K}$. Furthermore, modern algorithms additionally guarantee lower or even constant regret when the sequence of losses is more benign [see De Rooij et al., 2014, Koolen and Van Erven, 2015, Mourtada and Gaïffas, 2019].

In the multiscale setting, where the experts' loss ranges may differ by orders of magnitude, it is natural to ask about the existence of algorithms that guarantee an optimal worst-case regret bound that scales with the loss range of the best expert instead of the maximum range. This question has been answered affirmatively [Chen et al., 2021, Bubeck et al., 2019, Cutkosky and Orabona, 2018, Foster et al., 2017]. The algorithms developed in this line of work have had a significant impact in different areas of computational learning theory and practice. Unfortunately, as we will see, the best known algorithms still fail to guarantee lower regret even for the simplest benign statistical cases. Ensuring these goals poses serious technical challenges. In particular, Bernstein's inequality, the engine of classical same-scale luckiness arguments, has no suitable multiscale upgrade. Moreover, intuitive candidate upgrades of same-scale results would contradict recent lower bounds (see Section 7). To make things worse, in order to obtain multiscale regret bounds, close attention needs to be paid to terms that are conventionally insignificant but now carry the maximum scale of the problem. This motivates our main question: *can a single algorithm have multiscale worst-case regret guarantees and, in addition, exhibit constant (pseudo)regret in stochastic lucky cases?*

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

We answer the previous question affirmatively. The key contribution in this article is MUSCADA (multiscale adaptive), a computationally efficient algorithm that simultaneously guarantees a worst-case regret that grows with the scale of the best expert, and constant expected pseudoregret under a stochastic margin condition. MUSCADA uses a refined version of Follow the Regularized Leader based on the multiscale entropy of Bubeck et al. [2019]. Its crucial improvement is a second-order variance-like adaptation, the tightest possible for the analysis of this regularizer. This second-order adaptation is close in spirit to, and an improvement of, that of AdaHedge by De Rooij et al. [2014] and those of Chen et al. [2021]. As a result of careful analysis, MUSCADA has the following attractive properties: it does not need knowledge of the length of the game in advance without resorting to any doubling trick, the presence of zero-regret rounds does not change the state of the algorithm or its regret guarantees; it is invariant both under per-round, possibly unknown, translations of each expert's losses, and under a global known scaling common to all losses and ranges.

As an application of MUSCADA and its analysis techniques, we build an optimistic variant of the algorithm and use it to solve two-person zero-sum games that have a multiscale structure. The optimistic variant makes use of a guess of what the losses in the next round will be, and achieves lower regret when the guesses are adequate. This interest originates in the fact that optimistic algorithms converge to the solutions of such games at faster rates than their nonoptimistic counterparts [Syrgkanis et al., 2015]. We find experimentally that MUSCADA outperforms existing single-scale algorithms when the payoff matrix of the game exhibits a multiscale structure.

In the rest of this introduction we lay out formally the multiscale experts problem, review existing work, present a summary of the main contributions (Section 1.1), and outline the rest of the article.

Full-information online learning. In its simplest form, we must decide sequentially in rounds how to aggregate the predictions made by a fixed number K of *experts*. At each round t, we choose an aggregation strategy, a probability distribution $\boldsymbol{w}_t \in \mathcal{P}(K)$ over experts. After choosing $\boldsymbol{w}_t$, we assess the quality of the experts' predictions with a numerical loss $\boldsymbol{\ell}_t = (\ell_{t,k})_{k\in K}$ and judge the performance of our aggregation strategy by the $\boldsymbol{w}_t$-weighted losses $\langle \boldsymbol{w}_t, \boldsymbol{\ell}_t\rangle = \sum_{k\in K} w_{t,k}\ell_{t,k}$. Our objective is to minimize the cumulative gap between the losses incurred by our aggregation strategy $t \mapsto \boldsymbol{w}_t$ and the best expert in hindsight. This cumulative gap is the *regret* $\mathcal{R}_t = \sum_{s=1}^t \langle \boldsymbol{w}_s, \boldsymbol{\ell}_s\rangle - \min_{k\in K}\sum_{s=1}^t \ell_{t,k}$. Other than range restrictions on the losses, no assumptions are made about the mechanism that generates them. More precisely, for each expert $k \in K$ and all rounds t, we only assume that $\ell_{k,t} \in [-\sigma_k, \sigma_k]$ for known nonnegative scales $\{\sigma_k\}_{k\in K}$. We call $\boldsymbol{R}_t$ the vector of regrets with respect to each expert, that is, the vector with entries $R_{t,k} = \sum_{s=1}^t \{\langle \boldsymbol{w}_s, \boldsymbol{\ell}_s\rangle - \ell_{s,k}\}$.

Existing results. Several algorithms have been proposed that achieve the worst-case regret in the multiscale setting, but none of them achieve constant regret in stochastic lucky cases. Motivated by the problem of online model selection, Foster et al. [2017] used a technique of adaptive relaxations to produce randomized algorithms that guarantee

$$\mathbf{E}_{\mathbf{P}}[R_{t,k}] = O\left(\sigma_k\sqrt{t(\ln t + \ln(1/\pi_k) + \ln(\sigma_k/\sigma_{\min}))}\right) \text{ as } t\to\infty,$$

where $\boldsymbol{\pi}$ is a prior distribution on experts that generalizes the uniform $1/K$ of the Hedge algorithm and the expectation is over the algorithm's randomness. Bubeck et al. [2019] first proposed a Follow-the-Regularized-Leader algorithm with a multiscale entropy regularization that guarantees

$$R_{t,k} = O\left(\sigma_k\sqrt{t(\ln K + \ln(\sigma_{\max}/\sigma_{\min}))}\right) \text{ as } t\to\infty$$

when the number of rounds t is known in advance. Bubeck et al. [2019, Theorem 20] also construct an instance of the $K = 2$ experts problem in which there exists a time t for which any algorithm must have $R_{t,k'} \gtrsim \sigma_{k'}\sqrt{t(\ln K + \ln(\sigma_{\max}/\sigma_{\min}))}$ for some expert k', shedding some light on the minimax picture. Recently, Chen et al. [2021] designed an optimistic algorithm that uses the same regulatization as Bubeck et al. [2019] with an additional ingredient: at each round, a second-order correction is added to the losses before computing the next round's weights. At every round, their algorithm makes use of a guess vector $\boldsymbol{m}_t$ that can depend on the losses up to time $t-1$. The scale of the guesses $\boldsymbol{m}_t$ are assumed to be the same as that of the losses; $|m_{t,k}| \le \sigma_k$. For instance, valid choices for the guess $\boldsymbol{m}_t$ are $\mathbf{0}$ and the loss $\boldsymbol{\ell}_{t-1}$ of the previous round. The algorithm of Chen et al. [2021] achieves

$$R_{t,k} = O\left(\sigma_k\sqrt{\beta_{t,k}\ln t} + \sigma_{\max}\ln t\right) \text{ as } t\to\infty,$$

now scaling with the expert-dependent "time" $\beta_{t,k} = \sum_{s=1}^{t} \frac{(\ell_{s,k}-m_{s,k})^2}{\sigma_k^2} \leq 4t$. Furthermore, they show that a different single-scale tuning of their algorithm exhibits stochastic luckiness. Namely, if the losses are sampled from a distribution with a gap $d_{\min} > 0$ between the expected loss of the best expert k^* and that of any other expert, their algorithm guarantees that

$$R_{t,k^*} = O_{\mathbf{P}}\left(\frac{\ln t}{d_{\min}}\right) \text{ as } t \to \infty,$$

where $\mathbf{P}$ is the distribution of the losses. Their technique for stochastic luckiness uses the upcoming learner's loss as the guess $m_{t,k} = \langle \boldsymbol{w}_t, \boldsymbol{\ell}_t \rangle$. Unfortunately, this approach cannot be extended to the multiscale case, as these guesses may violate the experts' loss ranges.

1.1 Main results

In this section we present succinctly the regret guarantees for MUSCADA. Firstly, we present multiscale worst-case regret guarantees. Secondly, we present the stochastic luckiness results and Massart's margin condition. We then prove analogs of these results for an optimistic modification of MUSCADA in Section 4. We close this introduction with an outline of the rest of the article.

Worst-case bounds. We propose two tunings for MUSCADA; they cover the cases where there is or is not an expert with loss range equal to zero. Our results imply Theorem 1.1 below; it contains the regret guarantees for MUSCADA, expressed in terms of v_t, an implicitly defined variance-like second-order data-dependent quantity. The quantity v_t, defined by the algorithm, is the tightest allowed by our analysis and enables our luckiness result, Theorem 3.1. We interpret v_t through the upper bounds of Theorem 1.2, also below, as an internal scale-free measure of time, as $v_t \leq 4t$.

Theorem 1.1 (Regret Bounds)**.** *Consider* MUSCADA*, $t \mapsto v_t$ defined in Figure 1, and any initial probability distribution $\boldsymbol{\pi}$.*

- *If $\sigma_{\min} = \min_{k\in K} \sigma_k > 0$, Tuning 1 guarantees, for any loss sequence,*

$$R_{t,k} \leq c\,\sigma_k \sqrt{v_t(\ln(1/\pi_k) + \ln(\sigma_k/\sigma_{\min}))} + O(1) \quad \text{as } t \to \infty, \tag{1}$$

  where c is a constant depending only on π. The constant c is well-behaved: if $\max_{k\in K} \pi_k = 1-\varepsilon$, then $c \leq 4\sqrt{2}(1 + 1/(2\ln(1+\varepsilon)))$.

- *Even if $\min_{k\in K} \sigma_k = 0$, Tuning 2 ensures, for any loss sequence,*

$$R_{t,k} \leq 2\sigma_k \sqrt{2\, v_t(\ln(1/\pi_k) + \ln(1+v_t))}(1+o(1)) \quad \text{as } t \to \infty. \tag{2}$$

The following theorem (proven in Appendix G) shows that v_t is bounded by a second-order quantity. If $w_{t,k}$ are the weights played by MUSCADA at round t and $\eta_{t-1,k}$ are its learning rates, v_t is bounded by the variance over experts of the losses w.r.t. a tilted probability distribution $\tilde{w}_{t,k} \propto w_{t,k}\eta_{t-1,k}$. The shape of this quantity may seem surprising, but it is not artificial; our analysis shows that it is the tightest and, consequently, the natural second-order quantity associated to this choice of regularization. In Appendix G, we further motivate, via a Taylor approximation, the shape of the resulting upper bound.

Theorem 1.2. *Let $\tilde{w}_{t,k}$ be the probability distribution $\tilde{w}_{t,k} \propto w_{t,k}\eta_{t-1,k}$ and let $\Delta v_t = v_t - v_{t-1}$. Then, with either tuning from Figure 2, v_t, from Figure 1, satisfies*

$$\Delta v_t \leq 4\frac{\mathrm{var}_{\tilde{\boldsymbol{w}}_t}(\boldsymbol{\ell}_t)}{\langle \tilde{\boldsymbol{w}}_t, \boldsymbol{\sigma}^2 \rangle} \leq 4, \qquad \textit{where} \qquad \mathrm{var}_{\tilde{\boldsymbol{w}}_t}(\boldsymbol{\ell}_t) = \langle \tilde{\boldsymbol{w}}_t, (\boldsymbol{\ell}_t - \langle \tilde{\boldsymbol{w}}_t, \boldsymbol{\ell}_t \rangle)^2 \rangle.$$

Stochastic luckiness. We now turn to our results for stochastic easy data. Not all stochastic scenarios are easy (in fact, worst-case regret lower bounds are proved using stochastic data). We use Massart's margin condition, a standard benchmark for easy data.

Definition 1.3 (Massart's easiness condition)**.** The losses $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots$ satisfy Massart's easiness condition if they are generated i.i.d. from a distribution $\mathbf{P}$ with the following property: there exists a constant c_{M} and an expert $k^* \in K$ such that

$$\mathbf{E}_{\mathbf{P}}[(\ell_{t,k} - \ell_{t,k^*})^2] \leq c_{\mathrm{M}} \mathbf{E}_{\mathbf{P}}[\ell_{t,k} - \ell_{t,k^*}]$$

for all $k \in K$ and $t \geq 1$. In that case, $k^* = \arg\min_{k\in K} \mathbf{E}_{\mathbf{P}}[\ell_{t,k}]$ for all t.

Massart's condition is implied by a more interpretable gap condition [Koolen et al., 2016, Lemma 3]. If there exist a gap $d_{\min} > 0$ in expectation between the loss of any expert and that of the best one k^*, that is, if, for every $k \neq k^*$, $\mathbf{E}_{\mathbf{P}}[\ell_{1,k}] \geq d_{\min} + \mathbf{E}_{\mathbf{P}}[\ell_{1,k^*}]$, Massart's condition is satisfied with $c_{\mathrm{M}} = 1/d_{\min}$. We show the following theorem.

Theorem 1.4 (Constant regret under Massart's condition)**.** *Under Massart's condition (Definition 1.3),* MUSCADA *with either Tuning 1 or 2 has constant expected pseudoregret over time, that is,*

$$\mathbf{E}_{\mathbf{P}}[R_{t,k^*}] \lesssim 1.$$

Outline. The rest of this article is organized as follows. In Section 2, we introduce and analyze MUSCADA. In Section 3, we state the main results on stochastic luckiness for MUSCADA. In Section 4, we introduce an optimistic variant of MUSCADA, give remarks about its numerical implementation in Section 5, and apply it to accelerating the solution of multiscale games in Section 6. We end this article with a discussion of our results in Section 7.

2 The MUSCADA Multiscale Online Learning Algorithm

In this section, we describe our algorithm and motivate its design. We present two useful tunings and prove the corresponding worst-case regret guarantees. For the sake of intuition, we specialize the algorithm to the case of same-scale experts with uniform prior and compare its resulting closed form to AdaHedge [De Rooij et al., 2014]. Stochastic luckiness results are found in Section 3. We begin by introducing some notation.

Notation. We use boldface type for vectors in $\mathbb{R}^K$ ($\boldsymbol{R}_t, \boldsymbol{L}_t, \boldsymbol{\mu}_t, \boldsymbol{\eta}_t, \boldsymbol{\sigma}, \boldsymbol{u}$) and distributions on K experts ($\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{\pi}$). We number rounds so that all quantities indexed by t depend on the information witnessed by the learner in the first t rounds. Exceptionally, we use weights $\boldsymbol{w}_t$ at round t. For two functions f and g we write "$f = O(g)$ as $t \to \infty$" if there exists $c > 0$ such that $\lim_{t\to\infty} f(t)/g(t) \leq c$. Similarly, we write "$f(t) \sim g(t)$ as $t \to \infty$" if $\lim_{t\to\infty} f(t)/g(t) = 1$, and $f \lesssim g$ if there is $c > 0$ so that $f \leq cg$. We denote the simplex of probability distributions on K experts by $\mathcal{P}(K)$ and use K interchangeably for a number $K \in \mathbb{N}$ and the set $\{1, \ldots, K\}$.

We define MUSCADA in Figure 1 and give its two main tunings in Figure 2. At round t, after observing cumulative corrected losses $\boldsymbol{L}_{t-1} + \boldsymbol{\mu}_{t-1}$, MUSCADA plays weights

$$w_{t,k} = u_k \mathrm{e}^{-\eta_{t-1,k}(L_{t-1,k}+\mu_{t-1,k}+a^*_{t-1})},$$

where $u_k > 0$ is a tuning parameter related to the prior weights, $\boldsymbol{\eta}_{t-1}$ are learning rates that decrease over time, $\boldsymbol{\mu}_t$ are corrections incrementally computed at every round, and the scalar a^*_{t-1} ensures normalization (see Lemma F.7). The weights $\boldsymbol{w}_t$ are reminiscent of those played by the Hedge algorithm, but the normalization a^*_t cannot be computed explicitly in general. The weights $\boldsymbol{w}_t$ are the result of a Follow-the-Regularized-Leader update on a vector of corrected losses $\boldsymbol{L}_{t-1} + \boldsymbol{\mu}_{t-1}$. The regularizer employed is the multiscale entropy: for a fixed $\boldsymbol{u} > 0$, its Bregman divergence is

$$\boldsymbol{w} \mapsto D_{\boldsymbol{\eta}}(\boldsymbol{w}, \boldsymbol{u}) = \sum_{k\in K} w_k \frac{\ln(w_k/u_k) - (1 - u_k/w_k)}{\eta_k}, \quad \boldsymbol{w} \in \mathcal{P}(K) \tag{3}$$

[see Bubeck et al., 2019, Chen et al., 2021]. The goal substracting the data-dependent second-order corrections $\boldsymbol{\mu}_t$ from the experts' regrets is to keep a scalar potential function Φ_t negative. Here, the potential $t \mapsto \Phi_t$ is defined by convex conjugacy with respect to the multiscale entropy as

$$\Phi_t := \Phi(\boldsymbol{R}_t - \boldsymbol{\mu}_t, \boldsymbol{\eta}_t) = \max_{\boldsymbol{w}\in\mathcal{P}(K)} \langle \boldsymbol{w}, \boldsymbol{R}_t - \boldsymbol{\mu}_t \rangle - D_{\boldsymbol{\eta}_t}(\boldsymbol{w}, \boldsymbol{u}), \tag{4}$$

for which $\boldsymbol{w}_{t+1}$ is the maximizer. The corrections $\boldsymbol{\mu}_t$ and the consequent negativity of the potential Φ_t are the main ingredients in the regret analysis of MUSCADA. We next motivate these choices.

The shape of the corrections $\boldsymbol{\mu}_t$. We designed MUSCADA to favor experts with low corrected regret $\boldsymbol{R}_t - \boldsymbol{\mu}_t$. For the sake of informal discussion, our goal is to obtain $\mu_{t,k} \approx \sigma_k \sqrt{v_t \ln(1/\pi_k)}$. The algorithm achieves this by additively correcting the regrets in each round. Indeed, from the analysis of entropy-regularized algorithms, one would expect learning rates of the shape $\eta_{t,k} \approx \frac{1}{\sigma_k}\sqrt{\frac{\ln(1/\pi_k)}{v_t}}$

Parameters: A vector $u_k > 0$ of initial weights, initial strictly positive learning rates $\eta_{0,k} \leq 1/(2\sigma_k)$, and real, continuous nonincreasing functions $H_k : \mathbb{R}^+ \mapsto \mathbb{R}$ with $H_k(0) = 1$.
Initialization: Let $\mu_{0,k} = 0$, $v_0 = 0$, $R_{0,k} = 0$ and $L_{0,k} = 0$. For each round $t = 1, 2, 3, \ldots$

1. Play (follow the multiscale-entropy regularized leader of the corrected losses)

$$\boldsymbol{w}_t = \operatorname*{arg\,min}_{\boldsymbol{w} \in \mathcal{P}(K)} \langle \boldsymbol{w}, \boldsymbol{L}_{t-1} + \boldsymbol{\mu}_{t-1} \rangle + D_{\boldsymbol{\eta}_{t-1}}(\boldsymbol{w}, \boldsymbol{u}), \tag{5}$$

   where $D_{\boldsymbol{\eta}}$ is the multiscale relative entropy given in (3).
2. Observe loss $\boldsymbol{\ell}_t$. Update $R_{t,k} = R_{t-1,k} + \langle \boldsymbol{w}_t, \boldsymbol{\ell}_t \rangle - \ell_{t,k}$ and $L_{t,k} = L_{t-1,k} + \ell_{t,k}$.
3. Compute Δv_t, the value $\Delta v \geq 0$ such that

$$\Phi(\boldsymbol{R}_t - \boldsymbol{\mu}_{t-1} - \boldsymbol{\sigma}^2 \boldsymbol{\eta}_{t-1} \Delta v, \boldsymbol{\eta}_{t-1}) = \Phi(\boldsymbol{R}_{t-1} - \boldsymbol{\mu}_{t-1}, \boldsymbol{\eta}_{t-1}), \tag{6}$$

   where Φ is the potential function defined in (4).
4. Compute $\Delta \mu_{t,k} = \sigma_k^2 \eta_{t-1,k} \Delta v_t$. Update $\mu_{t,k} = \mu_{t-1,k} + \Delta \mu_{t,k}$ and $v_t = v_{t-1} + \Delta v_t$.
5. Set the new learning rate $\eta_{t,k} = \eta_{0,k} H_k(v_t)$.

Figure 1: MUSCADA

to be optimal. With this learning rates in mind, the desired correction $\boldsymbol{\mu}_t$ can be approximated using a Riemann-sum approximation of $\sqrt{v_t} = \int_0^{v_t} \frac{1}{2\sqrt{v}} \mathrm{d}v$. Indeed, for the conjectured learning rates, our target $\mu_{t,k}$ satisfies $\mu_{t,k} \approx \sigma_k^2 \sum_{s \leq t} \eta_{s-1,k} \Delta v_s$, where $\Delta v_t = v_t - v_{t-1}$. This implies that the choice $\Delta \mu_{t,k} = \sigma_k^2 \eta_{t-1,k} \Delta v_t$ as our per-round additive correction is helpful for achieving our goal. We discuss our precise choice of learning rates after the formal statement of Proposition 2.2 below.

Negativity of Φ. Our regret bounds are a direct consequence of the negativity of the potential $t \mapsto \Phi_t$. Indeed, by its definition, $\Phi_0 \leq 0$, and, because of our choice of nonincresing learning rates and corrections, the change in potential $\Delta \Phi_t = \Phi_t - \Phi_{t-1}$ can be bounded by

$$\Delta \Phi_t \leq \Phi(\boldsymbol{R}_t - \boldsymbol{\mu}_t, \boldsymbol{\eta}_{t-1}) - \Phi(\boldsymbol{R}_{t-1} - \boldsymbol{\mu}_{t-1}, \boldsymbol{\eta}_{t-1}) = 0,$$

where the last equality follows from (6), the choice of corrections $\Delta \boldsymbol{\mu}_t$. This implies the following lemma, of which we give a more general proof in Section C.1.

Lemma 2.1. The potential $t \mapsto \Phi_t$ starts at $\Phi_0 \leq 0$ and is decreasing for $t \geq 0$.

Once we prove that the potential Φ_t is negative, we are ready to derive regret guarantees for MUSCADA. The maximal nature of the potential $t \mapsto \Phi_t$ and its nonpositivity together imply that, *simultaneously* for all distributions $\boldsymbol{p} \in \mathcal{P}(K)$,

$$\langle \boldsymbol{p}, \boldsymbol{R}_t - \boldsymbol{\mu}_t \rangle \leq D_{\boldsymbol{\eta}_t}(\boldsymbol{p}, \boldsymbol{u}). \tag{7}$$

We choose $\boldsymbol{p}$ concentrated on each expert $k \in K$ to deduce the next proposition (proof in Section C.1).

Proposition 2.2. *Assume that the learning rates $t \mapsto \boldsymbol{\eta}_t$ are decreasing.* MUSCADA *guarantees that, for any $t = 1, 2, 3, \ldots$ and all $k \in K$,*

$$R_{t,k} \leq \mu_{t,k} + \frac{\ln(1/u_k)}{\eta_{t,k}} + \sum_{j \in K} \frac{u_j}{\eta_{t,j}} - \frac{1}{\eta_{t,k}}, \tag{8}$$

where $\mu_{t,k} = \sigma_k^2 \sum_{s \leq t} \eta_{s-1,k} \Delta v_s$. Furthermore, for $\eta_{t,k} = \eta_0 H_k(v_t)$ as in Figure 1, $\boldsymbol{\mu}_t$ satisfies

$$\mu_{t,k} \leq \sigma_k^2 \eta_{0,k} \int_0^{v_t} H_k(x) \mathrm{d}x + \sigma_k^2 (\eta_{0,k} - \eta_{t,k}) \max_{s \leq t} \Delta v_s. \tag{9}$$

Choice of learning rates. Proposition 2.2 guides us in choosing the learning rates presented in Figure 2. The starting value of the learning rates influences our ability to control v_t in terms of the variance of the losses of the algorithm while their behavior for large v_t determines the long-term growth of the regret bounds. The learning rates presented in Figure 2 interpolate smoothly

Let $\boldsymbol{\pi} \in \mathcal{P}(K)$ be a probability distribution on K experts.

Tuning 1 Requires $\sigma_{\min} > 0$. Set $u_k = \pi_k \frac{\sigma_{\min}}{\sigma_k}$, $\eta_{0,k} = \frac{1}{2\sigma_{\max}}$, $\gamma_k = 8\frac{\sigma_{\max}^2}{\sigma_k^2}\ln(1/u_k)$ and

$$H_{1,k}(v) = \frac{\mathrm{d}}{\mathrm{d}v}\left[\frac{v}{\sqrt{1+v/\gamma_k}}\right] = \frac{v/\gamma_k+2}{2(1+v/\gamma_k)^{3/2}}.$$

Tuning 2 Set $u_k = \pi_k, \eta_{0,k} = \frac{1}{2\sigma_{\max}}$, $\alpha_k = 32\frac{\sigma_{\max}^2}{\sigma_k^2}$, $\gamma_k = \alpha_k \ln(1/u_k)$ and

$$H_{2,k}(v) = \frac{\mathrm{d}}{\mathrm{d}v}\left[\sqrt{\alpha_k^2\left\{(1+v/\alpha_k)\ln\left(1+v/\alpha_k\right) - v/\alpha_k\right\} + \frac{v^2}{2(1+v/(2\gamma_k))}}\right]$$
$$= \frac{\alpha_k \ln\left(1+v/\alpha_k\right) + \frac{1}{2}\frac{2v+v^2/(2\gamma_k)}{(1+v/(2\gamma_k))^2}}{2\sqrt{\alpha_k^2\left\{(1+v/\alpha_k)\ln\left(1+v/\alpha_k\right) - v/\alpha_k\right\} + \frac{v^2}{2(1+v/(2\gamma_k))}}}.$$

If, for some k, $\sigma_k = 0$, define $H_{2,k}$ to be the limit value $\lim_{\sigma\downarrow 0} H_{2,k}(v_t) = 1$.

Figure 2: Tunings

between these two regimes by taking the form $\eta_{t,k}^{(1)} = \eta_{0,k}H_{1,k}(v_t)$ and $\eta_{t,k}^{(2)} = \eta_{0,k}H_{2,k}(v_t)$. Here, the starting learning rates are set to $\eta_{0,k} = 1/(2\sigma_{\max})$. The functions $H_{1,k}, H_{2,k} \leq 1$ decrease monotonically from their initial values $H_{1,k}(0) = H_{2,k}(0) = 1$ in such a way that, as $v_t \to \infty$,

$$\eta_{t,k}^{(1)} \sim \frac{\sqrt{2}}{\sigma_k}\sqrt{\frac{\ln(1/\pi_k)}{v_t}} \qquad \text{and} \qquad \eta_{t,k}^{(2)} \sim \frac{\sqrt{2}}{\sigma_k}\sqrt{\frac{\ln(1/\pi_k)+\ln v_t}{v_t}}.$$

The asymptotic expresion for $\eta_{t,k}^{(1)}$ is reminiscent of the optimal learning rates for the Hedge algorithm with the number of rounds t replaced by the refined v_t and the uniform $\ln K$ replaced by $\ln(1/\pi_k)$. Finally, with the Riemann sum bound (9) from Proposition 2.2 in mind, the learning rates were chosen as the derivatives of functions that will become the dominant term in the regret guarantees.

Tuned regret bounds. The learning rates from Figure 2 can be readily used in Proposition 2.2 to derive regret guarantees for MUSCADA. However, to facilitate interpretation, we bound the learning rates and their reciprocals in order to obtain the regret bounds contained in the following proposition (proof in Appendix C.2). After its statement, we prove Theorem 1.1 from the introduction.

Proposition 2.3. *Let $\boldsymbol{\pi}$ be a probability distribution on K.*

- MUSCADA *run with Tuning 1 depicted in Figure 2 guarantees that, for any $t = 1, 2, \ldots,$*

$$R_{t,k} \leq 2\sigma_k\sqrt{2v_t\ln(1/u_k)} + c_{\boldsymbol{\sigma},\boldsymbol{\pi}}\sigma_{\min}\sqrt{2v_t} + 8\sigma_{\max}\ln(1/u_k) + 4\sigma_{\max} + \frac{\sigma_k}{2}\max_{s\leq t}\Delta v_s, \quad (10)$$

  where the constant $c_{\boldsymbol{\sigma},\boldsymbol{\pi}} = \sum_{k\in K}\pi_k(1/\sqrt{\ln(1/u_k)})$ and $u_k = \pi_k\frac{\sigma_{\min}}{\sigma_k}$.

- MUSCADA *run with Tuning 2 depicted in Figure 2 guarantees that, for any $t = 1, 2, \ldots,$*

$$R_{t,k} \leq 2\sigma_k\sqrt{2v_t\left(\ln\left(1+\frac{\sigma_k^2 v_t}{32\sigma_{\max}^2}\right)+\ln(1/\pi_k)\right)} + \sigma_k\ln(1/\pi_k)Z_k + \sum_{j\in K}\pi_j\sigma_j Z_j + \frac{\sigma_k}{2}\max_{s\leq t}\Delta v_t, \quad (11)$$

  where $Z_k = \sqrt{\frac{v_t}{2\ln\left(1+\frac{\sigma_k^2 v_t}{32\sigma_{\max}^2}\right)}}\left(1+\sqrt{\frac{\min\{\ln(1/\pi_k), \frac{\sigma_k^2 v_t}{16\sigma_{\max}^2}\}}{\ln\left(1+\frac{\sigma_k^2 v_t}{32\sigma_{\max}^2}\right)}}\right) = O\left(\sqrt{\frac{v_t}{\ln v_t}}\right)$ *as* $v_t \to \infty$.

Proof of Main Theorem 1.1. With Proposition 2.3 at hand, we can prove the claims made in Section 1.1. Use the fact that $\sigma_{\min} \leq \sigma_k$ to conclude from (10) that, as $t \to \infty$,

$$R_{t,k} \leq 2\sigma_k \sqrt{2v_t \ln(1/u_k)} + 2c_{\boldsymbol{\sigma},\boldsymbol{\pi}} \sigma_k \sqrt{2v_t} + O(1).$$

We can bound $c_{\boldsymbol{\sigma},\boldsymbol{\pi}} / \sqrt{\ln(1/u_k)} \leq 1/\ln(1/\pi_{\max})$, where $\pi_{\max} = \max_{k \in K} \pi_k$. Consequently,

$$R_{t,k} \leq 2\sigma_k \left\{1 + 1/(2\ln(1+\varepsilon))\right\} \sqrt{2v_t \ln(1/u_k)} + O(1)$$

as $t \to \infty$ any time that $\pi_{\max} = 1 - \varepsilon$. This coincides with (1). Similarly, (11) implies (2). □

2.1 Closed-form solutions in the single-scale uniform-prior case

To help in the interpretation and to illustrate the challenges of the multiscale problem, we instantiate MUSCADA to a situation where all calculations can be carried out in closed form: when all scales are the same and equal to σ, and the initial weights $\boldsymbol{\pi}_{\mathrm{Unif}}$ are uniform; $\pi_{\mathrm{Unif},k} = 1/K$. This is the setting in which AdaHedge by De Rooij et al. [2014] operates. In this case, the learning rates and corrections of MUSCADA are the same for all experts; $\eta_{t,k} = \eta_t$ and $\Delta\mu_{t,k} = \Delta\mu_t$. The potential Φ_t and the corrections $\Delta\mu_t$ take the familiar form

$$\Phi_t = \frac{1}{\eta_t} \ln\left(\frac{1}{K} \sum_{k \in K} \mathrm{e}^{\eta_t (R_{t,k} - \mu_{t,k})}\right), \qquad \text{and} \qquad \Delta\mu_t = \frac{1}{\eta_{t-1}} \ln \sum_{k \in K} w_{t,k} \mathrm{e}^{\eta_{t-1}(\langle \boldsymbol{w}_t, \boldsymbol{\ell}_t \rangle - \boldsymbol{\ell}_t)}.$$

These two quantities play a central role in the analysis of AdaHedge, where De Rooij et al. [2014] called $\Delta\mu_t$ the *mixability gap*, the difference between the average $\langle \boldsymbol{w}_t, \boldsymbol{\ell}_t \rangle$ and the *mixed average* $-\frac{1}{\eta_{t-1}} \ln \sum_{k \in K} w_{t,k} \mathrm{e}^{-\eta_{t-1} \ell_{t,k}}$. The main quantity in our analysis, Δv_t, becomes

$$\Delta v_t = \frac{1}{\eta_{t-1}^2 \sigma^2} \ln \sum_{k \in K} w_{t,k} \mathrm{e}^{\eta_{t-1}(\langle \boldsymbol{w}_t, \boldsymbol{\ell}_t \rangle - \ell_{t,k})}.$$

Using well-known estimates for cumulant generating functions, Δv_t can be bounded by the ratio $\mathrm{var}_{\boldsymbol{w}_t}(\boldsymbol{\ell}_t)/\sigma^2$. Indeed, Hoeffding's inequality implies the worst-case bound $\Delta v_t \leq \frac{1}{2}$; Bernstein's, the second-order $\Delta v_t \lesssim \mathrm{var}_{\boldsymbol{w}_t}(\boldsymbol{\ell}_t)/\sigma^2$. Since it is v_t that appears in the regret bounds in Proposition 2.3, they are a refinement over those of AdaHegde[1]. Additionally, the present analysis yields improvements that are apparent in lower-order terms. Indeed, the last two terms in the regret bound (8) in Proposition 2.2 vanish, and the analysis used in the proof of Proposition 2.3 with $\eta_0 = \sqrt{2}/\sigma$ and the instantiation of H_1 from Figure 2, $H_1(x) = \frac{x/\ln(K)+2}{2(1+x/\ln(K))^{3/2}}$, give the regret bound

$$\mathcal{R}_t \leq \begin{cases} c_1 \sigma v_t + c_2 \sigma \ln K + \sigma/2 & \text{if } v_t \leq \ln K, \\ 2\sigma\sqrt{2v_t \ln K} + \sigma/2 & \text{if } v_t > \ln K \end{cases}$$

with $c_1 = 3/\sqrt{2}$ and $c_2 = 1/\sqrt{2}$. Unfortunately, multiscale analogs of Bernstein and Hoeffding's inequalities on Δv_t are not available; considerably more technical work needs to be carried out to prove Theorem 1.2. A multiscale analog of Bernstein's estimate for Δv_t is only available when all the learning rates are smaller than $1/(2\sigma_{\max})$ (see the proof of Theorem 1.2 in Appendix G).

3 Multiscale Stochastic Luckiness

In this section we show, under easiness conditions, that the expected pseudoregret of MUSCADA is constant. Assume that the loss vectors $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots$ are i.i.d. and are generated according to a distribution $\mathbf{P}$ that satisfies Massart's easiness condition (see Definition 1.3). For Tuning 1, assume that the minimum scale among experts $\sigma_{\min}$ is strictly positive. The analysis technique in this case is similar to that of Koolen et al. [2016] with an extra step. A use of Theorem 1.2 shows that Δv_t can be estimated in terms of $\mathrm{var}_{\boldsymbol{w}_t}(\boldsymbol{\ell}_t)$. This estimate possibly incurs in a multiplicative factor that can be as high as $1/\sigma_{\min}^2$. There are examples for which this constant is necessary (not shown). After this, standard arguments show that the expected pseudoregret is constant. See Appendix E for proofs.

[1] Our algorithm with learning rate tuning function $H(v) = \sqrt{\frac{\ln K}{4v}}$ comes closest to AdaHedge.

Theorem 3.1. *Under Massart's condition and using Tuning 1 from Figure 2, the expected pseudoregret of* MUSCADA *is bounded by a constant in the number of rounds. Specifically, for any* $t \geq 0$*,*

$$\mathbf{E}_{\mathbf{P}}[R_{t,k^*}] \lesssim a^2 c_{\mathrm{M}} + b,$$

where $a = \sqrt{2 \max_{i,j \in K} \left\{ \frac{1}{\sigma_i \sigma_j} \frac{\ln(1/\pi_i) + \ln(\sigma_i/\sigma_{\min})}{\ln(1/\pi_j) + \ln(\sigma_j/\sigma_{\min})} \right\}} \left(4\sigma_{k^*} \sqrt{2 \ln(1/u_{k^*})} + 2\sqrt{2} c_{\boldsymbol{\sigma},\boldsymbol{\pi}} \sigma_{\min} \right)$ *and* $b = 8\sigma_{\max} \ln(1/u_{k^*}) + 4\sigma_{\max} + 2\sigma_{k^*}$.

For Tuning 2, where we do not assume that $\sigma_{\min} > 0$, still $\mathbf{E}_{\mathbf{P}}[R_{t,k^*}] \lesssim 1$ using a different proof technique. Using the expression for the weights of the algorithm, we show that they concentrate on the best expert k^*. The analysis here is similar to that of Mourtada and Gaïffas [2019], but the lack of an expression for the normalizing a_t^* presents with an additional technical difficulty. The result is the following theorem.

Theorem 3.2. *Let* $d_k = \mathbf{E}_{\mathbf{P}}[\ell_{t,k} - \ell_{t,k^*}]$ *and assume that* $\min_{k \neq k^*} d_k > 0$*. Using Tuning 2 in Figure 2,* MUSCADA *guarantees constant expected pseudoregret. Specifically,*

$$\mathbf{E}_{\mathbf{P}}[R_{t,k^*}] \leq \sum_{k \in K} f(d_k), \qquad \textit{where} \qquad f(d) = O\left(\tfrac{\sigma_{\max}^2}{d} \ln\left(\tfrac{\sigma_{\max}^2}{d^2} \right) \right) \textit{ as } d \to 0.$$

Standard modifications of the arguments presented may be used to prove that the pseudoregret is constant with $\mathbf{P}$-high probability (not shown).

4 Optimism

In this section we show an optimistic variant of MUSCADA. Suppose that, before round t, we count on guesses $\boldsymbol{m}_t$ for what $\boldsymbol{\ell}_t$ will be. Assume that $\boldsymbol{m}_t$ is of the same scale as $\boldsymbol{\ell}_t$, that is, $|m_{t,k}| \leq \sigma_k$. In particular, this entails that $|\ell_{t,k} - m_{t,k}| \leq 2\sigma_k$. A modification of MUSCADA, presented in Figure 1, puts these guesses to good use. These modifications allow for regret guarantees similar to those contained in Proposition 2.3, but in this case $\Delta v_t^\circ \lesssim \mathrm{var}_{\tilde{\boldsymbol{w}}_t^\circ}(\boldsymbol{\ell}_t - \boldsymbol{m}_t) / \langle \tilde{\boldsymbol{w}}_t^\circ, \boldsymbol{\sigma}^2 \rangle$, where the superscript $\circ$ signals the optimistic analogs of the quantities from MUSCADA. These modifications are shown in Figure 3 and the regret bounds in the following proposition (proofs in Appendix D).

Proposition 4.1. *If* $t \mapsto v_t^\circ$ *is the variance process defined by Optimistic* MUSCADA *in Figure 3, the same regret bounds presented Proposition 2.3 hold with two modifications:* v_t° *instead of* v_t *and all scales doubled, that is,* $2\boldsymbol{\sigma}$ *instead of* $\boldsymbol{\sigma}$*. Furthermore, for each* $t = 1, 2, \ldots$*,* $\Delta v_t^\circ \leq 4 \, \mathrm{var}_{\tilde{\boldsymbol{w}}_t^\circ}(\boldsymbol{\ell}_t - \boldsymbol{m}_t) / \langle \tilde{\boldsymbol{w}}_t^\circ, \boldsymbol{\sigma}^2 \rangle \leq 4t$*, where* $\tilde{w}_{t,k}^\circ \propto w_{t,k}^\circ \eta_{t-1,k}$*.*

1' Compute the guess $\boldsymbol{m}_t$ and play

$$\boldsymbol{w}_t^\circ = \operatorname*{arg\,min}_{\boldsymbol{w} \in \mathcal{P}(K)} \langle \boldsymbol{w}, \boldsymbol{L}_{t-1} + \boldsymbol{m}_t + \boldsymbol{\mu}_{t-1} \rangle - D_{\boldsymbol{\eta}_{t-1}}(\boldsymbol{w}, \boldsymbol{u}).$$

3' Let Δv_t° be the value $\Delta v^\circ \geq 0$ such that

$$\Phi(\boldsymbol{R}_t - \boldsymbol{\mu}_{t-1} - \boldsymbol{\eta}_{t-1} \boldsymbol{\sigma}^2 \Delta v^\circ, \boldsymbol{\eta}_{t-1}) = \Phi(\boldsymbol{R}_{t-1} + \langle \boldsymbol{w}_t^\circ, \boldsymbol{m}_t \rangle - \boldsymbol{m}_t - \boldsymbol{\mu}_{t-1}, \boldsymbol{\eta}_{t-1}). \quad (12)$$

Tuning 1' and Tuning 2'. As in Figure 2 but with halved starting learning rate $\eta_{0,k} = \frac{1}{4\sigma_{\max}}$.

Figure 3: Optimistic MUSCADA, given as update w.r.t. Figure 1.

5 Computation

At each round, MUSCADA requires two computations. We now argue that both can be executed to machine precision in $O(K)$ time. First, computing the weights (5) given the losses $\boldsymbol{L}_{t-1}$ and correction terms $\boldsymbol{\mu}_{t-1}$ can be reduced, by Lemma F.6, to a single scalar convex minimization problem. Cancelling the derivative of the objective amounts to searching for the normalizing offset a_t. To that

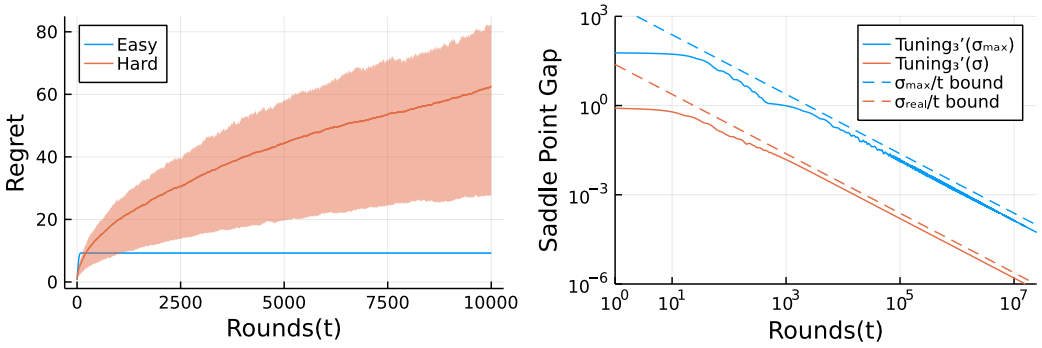

Figure 4: Left: empirical mean and quartiles of 2000 realizations of the regret $t \mapsto R_{t,k^*}$ of MUSCADA. For easy i.i.d. Massart distribution, the regret is constant; for a hard distribution without a gap, $\Omega(\sqrt{t})$. Right: optimistic MUSCADA (solid red) achieves an iterate-average saddle-point gap of σ_{real}/t where $\sigma_{\text{real}} = \sigma_{\max}/100$ is the relevant scale of the Nash equilibrium. Other methods scale as $\sigma_{\max}/t$.

end, binary search to machine precision takes $O(K)$ time per round. Notice that this also allows us to compute the potential value. Second, for computing the variance contribution (6), we observe that the right hand side of (6) is decreasing in Δv_t. Since the potential can be computed in $O(K)$ time, we can use an outer binary search to compute Δv_t to machine precision in $O(K)$ time as well. Alternatively, Newton's method may be employed; both of the previous problems require finding a root of a convex function. When deferring to a convex optimization library, a convenient expression is the jointly convex minimization form (see Lemma F.6)

$$\Delta v_t \;=\; \inf_{a,\Delta v} \; \Delta v \quad \text{subject to} \quad a + \sum_{k\in K} w_{t,k} \frac{\mathrm{e}^{\eta_{t-1,k}(\langle \boldsymbol{w}_t, \boldsymbol{\ell}_t\rangle - \ell_{t,k} - a) - \eta_{t-1,k}^2 \sigma_k^2 \Delta v} - 1}{\eta_{t-1,k}} \;\le\; 0.$$

6 Experiments on Synthetic Data

We investigate the performance of our multiscale method on two experiments: one for illustrating the performance of MUSCADA under Massart's condition, another for solving multiscale two-player zero-sum games.

The aim of the first experiment is to compare the performance of MUSCADA in easy and hard stochastic data sequences. To this end, we compared a sequence of hard stochastic data with no gap vs. easy data sampled i.i.d. from a distribution satisfying Massart's condition. We witnessed constant regret for the easy data, as shown in Figure 4 (Left). We take $K = 50$ experts and set $\sigma_k = 1/k$ for each $k \in K$. To generate our data, we fix some mean $\lambda_k \in [-\sigma_k, \sigma_k]$ and generate binary expert losses $\ell_{t,k} \in \{-\sigma_k, +\sigma_k\}$ independently between rounds and experts, with probability $\mathbf{P}\{\ell_{t,k} = \sigma_k\} = \frac{\sigma_k + \lambda_k}{2\sigma_k}$. For the hard case, we set $\lambda_k = 0$ for all k. For the lucky case, we set $\lambda_2 = -1/5$ instead. Generating this figure with the code in the supplementary material takes 3 seconds on an Intel i7-7700 processor.

The aim of the second experiment is to show the performance of MUSCADA for solving multiscale zero-sum games. Here, the payoff matrix is unknown, but row and column scales are available and vastly different. As detailed in Appendix A, we run two instances of appropriately tuned Optimistic MUSCADA against each other. As shown in Figure 4 (Right), the pair of time-average iterates converges to the saddle point with a suboptimality gap of order σ_{real}/t instead of the worst-case $\sigma_{\max}/t$, where σ_{real} is the maximum range within the support of the saddle point. In Appendix A, we conjecture that this rate holds for any such game and prove a weaker result: without optimism, the slower but scale-adaptive rate $\sigma_{\text{real}}/\sqrt{t}$ is achieved.

7 Discussion

We developed a new algorithm for multiscale online learning that is both worst-case safe and achieves constant pseudoregret in stochastic lucky cases. Our method is a refinement of the Follow-the-Regularized-Leader template with a weighted entropy. The main innovation is in the correction terms added to the losses, which are the tightest the technique admits. This suggests that these variance-like terms are in fact intrinsic to the problem of obtaining scale-dependent regret bounds. Lastly, we relate this newfound variance to the variance asked for by Freund [2016], we comment on the advantage of second-order guarantees over zeroth-order ones, and we state an open problem.

Quantile bounds and solving Freund's problem. Freund [2016] asked whether quantile adaptivity and variance adaptivity are compatible, that is, whether one can have $\langle \boldsymbol{p}, \boldsymbol{R}_t \rangle \leq \sqrt{\mathrm{KL}(\boldsymbol{p}, \boldsymbol{u}) \sum_{s \leq t} \mathrm{var}_{\boldsymbol{w}_s}(\boldsymbol{\ell}_s)}$ for all comparator distributions $\boldsymbol{p} \in \mathcal{P}(K)$ simultaneously. Even though our tuning of $\boldsymbol{\eta}_t$ does not yield quantile bounds, these can, however, be added employing a now-standard method [Koolen and Van Erven, 2015]. Namely, instead of only including every expert with a private learning rate tuned to its prior complexity level (the typical $\ln K$ or $\ln(1/\pi_k)$ term), we include multiple copies of each expert, each with a learning rate tuned to a smaller complexity level. We then start from (7) with comparator distribution $\boldsymbol{p}$ concentrated on the ε-quantile of interest and carry out all future steps (from Proposition 2.2 on), ending up with the quantile regret bound $\langle \boldsymbol{p}, \boldsymbol{R}_t \rangle \leq \max_{k: p_k > 0} \sigma_k \sqrt{v_t \left(\ln C + D_{\boldsymbol{\eta}_0}(\boldsymbol{p}, \boldsymbol{u}) \right)}$, where C is the number of learning rates thus created. As these learning rates can be exponentially spaced in an interval of width $\ln K$, C is of order $\ln \ln K$. Does this procedure answer Freund's question? For our notion of variance, v_t, which our results suggest is a rather useful notion, the answer is yes. However, to relate Δv_t to $\mathrm{var}_{\boldsymbol{w}_t}(\boldsymbol{\ell}_t)$, we incur a multiplicative ratio $\eta_{t,\max}/\eta_{t,\min}$, which, for the quantile case, is of order $\sqrt{\ln K}$, turning the prior-in-the-square-root bound into a prior-outside-the-square-root bound. The latter was already achievable by not tuning $\boldsymbol{\eta}$ to the prior complexities at all. This problem does not arise in the same-scale uniform-prior case; there, Δv_t is bounded by a small multiple of $\mathrm{var}_{\boldsymbol{w}_t}(\boldsymbol{\ell}_t)$ [De Rooij et al., 2014]. Note that this problem is present even when K is fixed while t grows, which is narrowly outside the scope of the impossibility results of Marinov and Zimmert [2021]. This discussion sheds light from another angle on why Freund's problem is hard; we present a desirable multiscale alternative.

Luckiness, gap, and Massart's condition. We now address the advantage of MUSCADA's refined second-order measure of time v_t over the zeroth-order number of rounds t. Multiscale zeroth-order regret bounds (growing with t) can be guaranteed either by tuning MUSCADA crudely to a constant multiple of t or by building an any-time improvement of the algorithm of Bubeck et al. [2019], also tuned to t. Both t-tuned and v_t-tuned algorithms have constant expected pseudoregret in stochastic lucky cases, but the constant can be widely different. Indeed, the constant for t-tuned algorithms scales with the inverse $1/d_{\min}$ of the gap $d_{\min} = \min_{k \neq k^*} \mathbf{E}[\ell_{t,k} - \ell_{t,k^*}]$, while the constant for v_t-tuned algorithms scales with the constant c_{M} from Massart's condition (see Definition 1.3). The difference stems from the fact that c_{M} is at most $1/d_{\min}$, but it can be arbitrarily smaller. This separation appears to be fundamental. In the single-scale uniform-prior case, the above t-tuned algorithms are closely related to Decreasing Hedge [Mourtada and Gaïffas, 2019], just as MUSCADA is related to AdaHedge (see Section 2.1). Mourtada and Gaïffas [2019] show that, in the single-scale case, even under Massart's condition with $c_{\mathrm{M}} = 1$, Decreasing Hedge and, consequently, Bubeck et al.'s algorithm with decreasing learning rates, has expected pseudoregret $\mathbf{E}[R^B_{t,k}] \gtrsim 1/d_{\min}$. If the smallest scale $\sigma_{\min} > 0$, by taking $d_{\min}$ small, this lower bound can be made arbitrarily worse than the guarantee of MUSCADA, $\mathbf{E}[R^{\text{MUSCADA}}_{t,k^*}] \lesssim c_{\mathrm{M}} + 1$, from Theorem 3.1.

Open problem. Our ability to incorporate an arbitrary prior suggests that the results should extend to countably many experts. However, the current techniques do break down. When $\max_{k \in \mathbb{N}} \sigma_k < \infty$ MUSCADA with Tuning 1 (if $\inf_{k \in \mathbb{N}} \sigma_k > 0$) or Tuning 2 would still deliver the worst-case bound. Yet our luckiness result currently requires $\max_{k,l,t} \frac{\eta_{t,k}}{\eta_{t,l} \sigma_l^2} < \infty$. Even with a common scale σ, this is never the case due to the dependence of $\boldsymbol{\eta}_t$ on the prior $\boldsymbol{\pi}$, which is necessarily decreasing. Is luckiness actually possible, for example, in the online learning analog of the elegant challenge example presented by Talagrand [2014, Chapter 2]?

References

Sébastien Bubeck, Nikhil R. Devanur, Zhiyi Huang, and Rad Niazadeh. Multi-scale online learning: Theory and applications to online auctions and pricing. *Journal of Machine Learning Research*, 20 (62):1–37, 2019.

Liyu Chen, Haipeng Luo, and Chen-Yu Wei. Impossible Tuning Made Possible: A New Expert Algorithm and Its Applications. *arXiv:2102.01046 [cs]*, June 2021. arXiv: 2102.01046.

Ashok Cutkosky and Francesco Orabona. Black-Box Reductions for Parameter-free Online Learning in Banach Spaces. In *Conference On Learning Theory*, pages 1493–1529. PMLR, July 2018.

Dylan J. Foster, Satyen Kale, Mehryar Mohri, and Karthik Sridharan. Parameter-Free Online Learning via Model Selection. In *Advances in Neural Information Processing Systems 30*, pages 6020–6030. Curran Associates, Inc., 2017.

Yoav Freund. Open Problem: Second order regret bounds based on scaling time. In *Conference on Learning Theory*, pages 1651–1654, 2016.

Yoav Freund and Robert E. Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.

Elad Hazan. Introduction to Online Convex Optimization. *arXiv:1909.05207 [cs, math, stat]*, September 2019. arXiv: 1909.05207.

Yu-Guan Hsieh, Kimon Antonakopoulos, and Panayotis Mertikopoulos. Adaptive learning in continuous games: Optimal regret bounds and convergence to nash equilibrium. In *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 2388–2422. PMLR, 15–19 Aug 2021.

Wouter M. Koolen and Tim van Erven. Second-order Quantile Methods for Experts and Combinatorial Games. In *Conference on Learning Theory*, pages 1155–1175. PMLR, June 2015.

Wouter M Koolen, Peter Grünwald, and Tim van Erven. Combining Adversarial Guarantees and Stochastic Fast Rates in Online Learning. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Nick Littlestone and Manfred K. Warmuth. The Weighted Majority Algorithm. *Information and Computation*, 108(2):212–261, February 1994.

Teodor Vanislavov Marinov and Julian Zimmert. The pareto frontier of model selection for general contextual bandits. In *Advances in Neural Information Processing Systems*, 2021.

Jaouad Mourtada and Stéphane Gaïffas. On the optimality of the Hedge algorithm in the stochastic regime. *Journal of Machine Learning Research*, 20(83):1–28, 2019.

Alexander Rakhlin and Karthik Sridharan. On Equivalence of Martingale Tail Bounds and Deterministic Regret Inequalities. In *Conference on Learning Theory*, pages 1704–1722. PMLR, June 2017.

Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

Steven de Rooij, Tim van Erven, Peter D. Grünwald, and Wouter M. Koolen. Follow the Leader If You Can, Hedge If You Must. *Journal of Machine Learning Research*, 15(37):1281–1316, 2014.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E. Schapire. Fast Convergence of Regularized Learning in Games. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

Michel Talagrand. *Upper and Lower Bounds for Stochastic Processes*, pages 1–12. 01 2014. ISBN 978-3-642-54074-5.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] **We include in Section 7 a discussion on open problems.**
   (c) Did you discuss any potential negative societal impacts of your work? [No]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets... **We do not use any existing assets.**
5. If you used crowdsourcing or conducted research with human subjects... **We did not use crowdsourcing or human subjects.**

A Saddle-Point Computation in Multiscale Games

One application of online learning is computing approximate mixed-strategy Nash equilibria in finite two-player zero-sum games (and more generally, to approximate saddle points of convex-concave functions). Here, we investigate a multiscale version of that problem. Our main focus is to find methods whose performance does not depend on the *maximum scale*, but on the *relevant scale* to the problem instance at hand. In this case, this means the scale of the payoffs in the subset of rows and columns in the support of the Nash equilibrium. In Section A.1 we lay out the setup of two-player zero-sum finite games. In Section A.2 we define the suboptimality gap, the main measure of performance in judging the solution to these games. In Section A.3 we define the payoff matrices used in the experiments that produced Figure 4. We conjecture that MUSCADA achieves fast scale-dependent convergence in Section A.5 and provide the additional details of the experiments that produced Figure 4(right) in Section A.6.

A.1 Two-player zero-sum finite games

Given a payoff matrix $A \in \mathbb{R}^{K \times M}$ (specifying losses for the row player and gains for the column player) we are looking for the mixed-strategy saddle point $(\boldsymbol{p}_*, \boldsymbol{q}_*) \in \mathcal{P}(K) \times \mathcal{P}(M)$ such that

$$\min_i \boldsymbol{e}_i^\intercal A \boldsymbol{q}_* \;\geq\; \max_j \boldsymbol{p}_*^\intercal A \boldsymbol{e}_j .$$

Our approach will be based on oracle access to the matrix-vector products $\boldsymbol{q} \mapsto A\boldsymbol{q}$ and $\boldsymbol{p} \mapsto A^\intercal \boldsymbol{p}$. We will use the scheme of running two online learners against each other, with loss vectors $\boldsymbol{\ell}_t^{\text{row}} = A\boldsymbol{q}_t$ and $\boldsymbol{\ell}_t^{\text{col}} = -A^\intercal \boldsymbol{p}_t$ and optimistic estimates given by the past loss vector $\boldsymbol{m}_t^{\text{row/col}} = \boldsymbol{\ell}_{t-1}^{\text{row/col}}$. For the same-scale case, Rakhlin and Sridharan [2013] show that uncoupled adaptive schemes benefit from convergence of the gap of the pair of iterate averages at rate $O(\sigma_{\max} \frac{\ln K + \ln M}{T})$, while recently Hsieh et al. [2021] showed last iterate convergence as well. Here we investigate the advantage of using adaptive multiscale learners to improve the dependence in $\sigma_{\max}$.

A.2 The metric of success: suboptimality gap

We are looking for the equilibrium in mixed strategies, i.e. $\min_{\boldsymbol{p}} \max_{\boldsymbol{q}} \boldsymbol{p}^\intercal A \boldsymbol{q}$. The social exploitability of a candidate saddle point pair $\boldsymbol{p}, \boldsymbol{q}$ is defined as the gap

$$\mathrm{gap}(\boldsymbol{p}, \boldsymbol{q}) \;=\; \max_j \boldsymbol{p}^\intercal A \boldsymbol{e}_j - \min_i \boldsymbol{e}_i^\intercal A \boldsymbol{q}.$$

We use the common technique of employing online learning with linear loss functions $\boldsymbol{p} \mapsto A\boldsymbol{q}_t$ and $\boldsymbol{q} \mapsto -A^\intercal \boldsymbol{p}_t$. A standard analysis [Freund and Schapire, 1997] bounds the gap of the iterate averages $\bar{\boldsymbol{p}}_t = \frac{1}{t}\sum_{s\le t} \boldsymbol{p}_s$ and $\bar{\boldsymbol{q}}_t = \frac{1}{t}\sum_{s\le t} \boldsymbol{q}_s$ from above by the social (sum-of) regret

$$\begin{aligned}\mathrm{gap}(\bar{\boldsymbol{p}}_t, \bar{\boldsymbol{q}}_t) \;&=\; \max_j \bar{\boldsymbol{p}}_t^\intercal A \boldsymbol{e}_j - \min_i \boldsymbol{e}_i^\intercal A \bar{\boldsymbol{q}}_t \;=\; \frac{1}{t}\left(\max_j \sum_{s\le t} \boldsymbol{p}_s^\intercal A \boldsymbol{e}_j - \min_i \sum_{s\le t} \boldsymbol{e}_i^\intercal A \boldsymbol{q}_s\right) \\ &=\; \frac{1}{t}\max_{i,j}\Big(\underbrace{\sum_{s\le t} \boldsymbol{p}_s^\intercal A \boldsymbol{e}_j - \sum_{s\le t} \boldsymbol{p}^\intercal A \boldsymbol{p}_s}_{R_t^{\boldsymbol{q}}(j)} + \underbrace{\sum_{s\le t} \boldsymbol{p}^\intercal A \boldsymbol{p}_s - \sum_{s\le t} \boldsymbol{e}_i^\intercal A \boldsymbol{q}_s}_{R_t^{\boldsymbol{p}}(i)}\Big).\end{aligned}$$

Having multiscale regret bounds at our disposal, it is natural to look at multiscale payoff matrices.

A.3 Multiscale structure

We will assume that our payoff matrix is multiscale in the sense that we are given row and column range vectors $\boldsymbol{\sigma}^{\text{row}}$ and $\boldsymbol{\sigma}^{\text{col}}$ such that $|A_{ij}| \le \min\{\sigma_i^{\text{row}}, \sigma_j^{\text{col}}\}$. The main point is to learn the saddle point faster if the maximum range is much larger than the range in the support of the saddle point, i.e. $\sigma_{\max}^{\text{row}} \gg \sigma_{\text{real}}^{\text{row}} := \max\{\sigma_i^{\text{row}} \mid \boldsymbol{e}_i^\intercal \boldsymbol{p}_* > 0\}$ and/or $\sigma_{\max}^{\text{col}} \gg \sigma_{\text{real}}^{\text{col}} := \max\{\sigma_j^{\text{col}} \mid \boldsymbol{e}_j^\intercal \boldsymbol{q}_* > 0\}$. We will denote that largest relevant scale by $\sigma_{\text{real}} = \max\left\{\sigma_{\text{real}}^{\text{row}}, \sigma_{\text{real}}^{\text{col}}\right\}$. Our aim is to get gap bounds that scale with σ_{real}, not $\sigma_{\max}$.

Example A.1 (Simple multiscale Game)**.** For the purpose of our experiment, we will construct our multiscale payoff matrices following the template

$$A \;=\; \begin{bmatrix} B & -\mathbf{1}\mathbf{1}^\intercal \\ \mathbf{1}\mathbf{1}^\intercal & C \end{bmatrix}$$

where B_{ij} are i.i.d. Rademacher $\{\pm 1\}$ and C_{ij} are i.i.d. Rademacher $\{\pm\sigma_{\max}\}$ for some pre-specified $\sigma_{\max} \gg 1$. By construction, any saddle point for the submatrix B is (upon padding with zeros) also a saddle point for the full matrix A. Moreover, it is a strict saddle point for A if it is a strict saddle point for B with value $\min_{\boldsymbol{p}} \max_{\boldsymbol{q}} \boldsymbol{p}^\intercal B \boldsymbol{q} \in (\pm 1)$. We will assume throughout that we are in this latter strict case. Here $\sigma_{\text{real}} = 1$ regardless of $\sigma_{\max}$.

A.4 What can one hope to achieve?

Throughout the remainder we assume for simplicity that the saddle point $\boldsymbol{p}_*, \boldsymbol{q}_*$ of the payoff matrix A is unique (a common situation). We define the *optimality gap* of row i by $\delta^{\text{row}}(i) = (\boldsymbol{e}_i - \boldsymbol{p}_*)^\intercal A \boldsymbol{q}_* \ge 0$ and of column j by $\delta^{\text{col}}(j) = \boldsymbol{p}_*^\intercal A(\boldsymbol{q}_* - \boldsymbol{e}_j) \ge 0$. We are interested in scenarios where at least one player has strictly positive optimality gap on the action(s) of largest scale. We will show that multiscale regret bounds allow the learning to accelerate. Moreover, the learner does not need to know about this structure and will adapt automatically.

Let us assume without loss of generality that $\delta^{\text{row}}(k) > 0$ while $\sigma_k^{\text{row}} = \max_i \sigma_i^{\text{row}}$ where $\sigma_i^{\text{row}} = \max_j |A_{i,j}|$. The general idea now is to use that $\bar{\boldsymbol{p}}_T \to \boldsymbol{p}_*$. This means that from some point t on,

$$\max_j \bar{\boldsymbol{p}}_t^\intercal A \boldsymbol{e}_j \;=\; \max_{j:\boldsymbol{q}_*(j)>0} \bar{\boldsymbol{p}}_t^\intercal A \boldsymbol{e}_j \;=\; \frac{1}{t}\max_{j:\boldsymbol{q}_*(j)>0} \sum_{s\le t} \boldsymbol{p}_s^\intercal A \boldsymbol{e}_j \;\le\; \frac{1}{t}\sum_{s\le t} \boldsymbol{p}_s^\intercal A \boldsymbol{q}_s + \max_{j:\boldsymbol{q}_*(j)>0} \frac{1}{t} R_t^{\text{col}}(j)$$

A similar argument for the row player then allows us to conclude

$$\begin{aligned}\mathrm{gap}(\bar{\boldsymbol{p}}_t, \bar{\boldsymbol{q}}_t) \;&\le\; \frac{1}{t}\left(\sum_{s\le t} \boldsymbol{p}_s^\intercal A \boldsymbol{q}_s + \max_{j:\boldsymbol{q}_*(j)>0} R_t^{\text{col}}(j) - \sum_{s\le t} \boldsymbol{p}_s^\intercal A \boldsymbol{q}_s + \max_{i:\boldsymbol{p}_*(i)>0} R_t^{\text{row}}(i)\right) \\ &=\; \frac{1}{t}\left(\max_{j:\boldsymbol{q}_*(j)>0} R_t^{\text{col}}(j) + \max_{i:\boldsymbol{p}_*(i)>0} R_t^{\text{row}}(i)\right).\end{aligned}$$

The main point is that this bound scales with $\max_{i:\boldsymbol{p}_*(i)>0} \sigma_i^{\text{row}} + \max_{j:\boldsymbol{q}_*(j)>0} \sigma_j^{\text{col}}$ and not with the respective unconstrained maxima.

Proposition A.2. *Any pair of multiscale online learning algorithms with bounds of order $R_t^i \leq O(\sigma_i\sqrt{T})$, including* MUSCADA *with Tuning 3 (see Lemma B.1), ensures iterate average gap*

$$\text{gap}(\bar{p}_t, \bar{q}_t) \;=\; O(\sigma_{\text{real}}/\sqrt{t})$$

as $t \to \infty$.

Note that same-scale algorithms would only deliver the weaker guarantee $O(\sigma_{\max}/\sqrt{t})$.

A.5 Why our approach may achieve the hope optimistically

Rakhlin and Sridharan [2013] show that using optimism in saddle point interactions can improve the rate to $O(\sigma_{\max}/t)$. We first show that this is true for MUSCADA as well, after which we will investigate achieving $O(\sigma_{\text{real}}/t)$. The mechanism for this proof is to show that the social regret is constant. Technically, one would explicitly keep track of the slack in (17) and (18), and use these harvested slacks to cancel the $\sqrt{t}$ term of the regret bound. Only the constant-order term measuring the entropy of the initial weights remains. For this to be a constant, we further need that the learning rate stops decreasing once the regret stabilizes. Following exactly the steps of Rakhlin and Sridharan [2013], we can prove the following proposition.

Proposition A.3. *For same-scale games, the optimistic version (see Figure 3) of* MUSCADA *with Tuning 3 and uniform prior (see Lemma B.1) achieves average iterate gap $gap(\bar{\boldsymbol{p}}_t, \bar{\boldsymbol{q}}_t) = O(\sigma_{\max}/t)$ as $t \to \infty$.*

The same-scale assumption makes all $\boldsymbol{\sigma}$ equal, while the uniform-prior assumption in addition makes all $\boldsymbol{\eta}$ equal. This makes the standard argument from the literature apply.

We further forward the natural conjecture that we state next.

Conjecture A.4. *For the multiscale case, the optimistic version (see Figure 3) of* MUSCADA *with Tuning 3 and any nondegenerate prior (see Lemma B.1) achieves average iterate gap bounded by $gap(\bar{\boldsymbol{p}}_t, \bar{\boldsymbol{q}}_t) = O(\sigma_{\text{real}}/t)$.*

The reason that our Tuning 3 has any chance here is that *no* terms (not even the additive constant) in the regret bound scale with $\sigma_{\max}$. This in contrast to the algorithms of Foster et al. [2017], Cutkosky and Orabona [2018], Bubeck et al. [2019], Chen et al. [2021], whose existing multiscale analyses all result in a lower-order term scaling with $\sigma_{\max}$.[2] We next provide empirical support for our conjecture.

A.6 Numerical results

We investigate three algorithms: Hedge with classic time-decreasing learning rate $\eta_t = \sqrt{\frac{\ln(K)}{\sigma_{\max}^2 t}}$, MUSCADA with all scales set to $\sigma_{\max}$ and MUSCADA with actual knowledge of the multiscale vectors. All algorithms are run in optimistic mode with guesses $\boldsymbol{m}_t = \boldsymbol{\ell}_{t-1}$, the loss vector of the previous round (and $\boldsymbol{m}_{1,k} = 0$). We choose a matrix of structure given in Example A.1, with B and C of size 10×10, and pick $\sigma_{\max} = 100$. We give all algorithms the uniform prior $\pi_k = 1/20$. The results are displayed in Figure 5, where we show the saddle point gap for the average iterate, the last iterate and the theoretical regret bounds that we obtain from the analysis. In the main text, Figure 4(right) shows only the saddle point gap for the average iterate of optimistic MUSCADA with the optimistic modification of Tuning 3 from Figure 6. Generating this figure with the code from the supplementary material takes 30 minutes on an Intel i7-7700 processor. Memory usage is negligible.

We see in Figure 5 that the gap of optimistic Hedge decays at the slow rate $O(\sigma_{\max}/\sqrt{t})$. This means that optimism alone is insufficient to obtain a faster $O(\sigma_{\max}/t)$ convergence rate; it is also necessary that the learning rates stop decreasing when the regret plateaus. It is also apparent that MUSCADA tuned to $\sigma_{\max}$ has the fast $O(1/t)$ rate, but at the $\sigma_{\max}$ scale. Finally, the numerical experiments show evidence that our multiscale algorithm does exploit the small scale of the actions in the support of the saddle point, exhibiting the desired $O(\sigma_{\text{real}}/t)$ regret conjectured above. The plot also includes the quality of the last iterate. Hsieh et al. [2021] prove convergence of the last iterate for the common

[2] Which is hard to spot in some of the literature because of a global $\sigma_{\max} = 1$ convention.

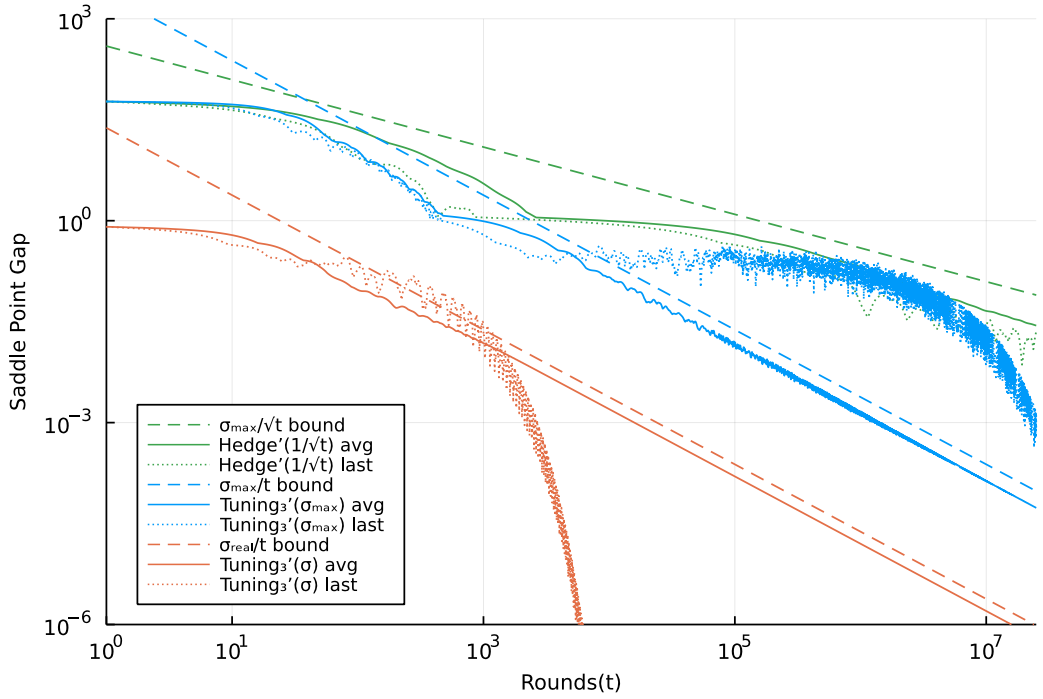

Figure 5: Quality of average iterate (solid) and last iterate (dotted) for three optimistic algorithms, compared to their relevant bounds (dashed). The multiscale-aware algorithm (red) outperforms the non-scale-aware competitors by the factor $\sigma_{\max}/\sigma_{\text{real}} = 100$. See Section A.6 for further discussion.

Tuning 3 $u = \pi \frac{\sigma_{\min}}{\sigma_k}$, $\eta_{0,k} = \frac{1}{2\sigma_k}$, $\gamma = 8\ln(1/u_k)$, and

$$H_{1,k}(v_t) = \frac{\mathrm{d}}{\mathrm{d}v_t}\left[\frac{v_t}{\sqrt{1+v_t/\gamma_k}}\right] = \frac{v_t/\gamma_k + 2}{2(1+v_t/\gamma_k)^{3/2}}.$$

Figure 6: Tuning 3 for MUSCADA

scale, common prior case. In our experiment the iterate average can be seen to converge quickly in the multiscale case, but convergence is terribly slow in the same-scale case. This is not inconsistent; no rates are currently known for the last iterate.

B Tuning 3

In this section we describe a third tuning, defined in Figure 6. In contrast to Tunings 1 and 2, the learning rates in Tuning 3 start *higher*, namely at $1/(2\sigma_k)$ instead of $1/(2\sigma_{\max})$. The downside of this aggressive tuning is that the variance bound is not available (though the weaker, uncentered second-moment analog is). The upside is that the resulting regret bound compared to expert k features only σ_k and has *no* occurrence of $\sigma_{\max}$ whatsoever, not even in the additive constants.

Lemma B.1. Let $\boldsymbol{\pi}$ be a probability distribution on K experts. MUSCADA run with Tuning 3 depicted in Figure 6 guarantees that, for any $t = 1, 2, \ldots,$

$$R_{t,k} \le 2\sigma_k\sqrt{2v_t\ln(1/u_k)} + c_{\boldsymbol{\sigma},\boldsymbol{\pi}}\sigma_{\min}\sqrt{2v_t} + 8\sigma_k\ln(1/u_k) + 4\sigma_{\min} + \frac{\sigma_k}{2}\max_{s\le t}\Delta v_s, \qquad (13)$$

where $c_{\boldsymbol{\sigma},\boldsymbol{\pi}} = \sum_{k\in K} \pi_k(1/\sqrt{\ln(1/u_k)})$ and $u_k = \pi_k \frac{\sigma_{\min}}{\sigma_k}$. Additionally, $v_t \le 4\sum_{s\le t} \frac{\langle \tilde{\boldsymbol{w}}_s, \boldsymbol{\ell}_s^2\rangle}{\langle \tilde{\boldsymbol{w}}_s, \boldsymbol{\sigma}^2\rangle} \le 4t$, where, for each $t = 1, 2, \dots$, the weights are $\tilde{w}_{t,k} \propto w_{t,k}\eta_{t-1,k}$.

Proof. Follow the same steps as in the proof of the regret bound for Tuning 1 in Lemma 2.3. Obtain that

$$\mu_{t,k} \le \sigma_k\sqrt{2v_t\ln(1/u_k)} + 4\sigma_k\ln(1/u_k) \tag{14}$$

$$\frac{\ln(1/u_k)}{\eta_{t,k}} \le \sigma_k\sqrt{2v_t\ln(1/u_k)} + 4\sigma_k\ln(1/u_k), \text{ and} \tag{15}$$

$$\sum_{k\in K}\frac{u_k}{\eta_k} \le c_{\boldsymbol{\sigma},\boldsymbol{\pi}}\sigma_{\min}\sqrt{2v_t} + 4\sigma_{\min} \tag{16}$$

with $c_{\boldsymbol{\sigma},\boldsymbol{\pi}} = \sum_{k\in K}\pi_k\left(\frac{1}{\sqrt{\ln(1/u_k)}}\right)$. Use Proposition 2.3 to conclude the first claim. For the additional claim, use Lemma G.2 with $\lambda = 0$. □

C Algorithm Analysis

The only step in the algorithm that may be problematic is the definition of Δv_t at every round, which one might think can take infinite values. We show in Proposition G.1 that this is not the case and that consequently $t \mapsto v_t$ is well defined.

C.1 Untuned regret bound, proof of Proposition 2.2

We prove that the potential $t \mapsto \Phi_t$ is decreasing for optimistic MUSCADA. The result for the nonoptimistic version follows by setting the guesses $\boldsymbol{m}_t$ to $\mathbf{0}$. Recall from (4) in Section 2 that the potential Φ_t is defined by

$$\Phi_t \;=\; \Phi(\boldsymbol{R}_t - \boldsymbol{\mu}_t, \boldsymbol{\eta}_t) \;=\; \max_{\boldsymbol{w}\in\mathcal{P}(K)} \langle \boldsymbol{w}, \boldsymbol{R}_t - \boldsymbol{\mu}_t\rangle - D_{\boldsymbol{\eta}_t}(\boldsymbol{w}, \boldsymbol{u}).$$

Proof of Lemma 2.1. We prove the result in the optimistic case. The nonoptimistic case is recovered for $\boldsymbol{m}_t = \mathbf{0}$ and replacing $4\sigma_k^2$, which is a bound on $|m_{t,k} - \ell_{t,k}|^2$, by σ_k^2, which bounds $|\ell_{t,k}|^2$. The result is a consequence of the following inequalities:

$$\begin{aligned}
\Phi_t \;&\le\; \Phi(\boldsymbol{R}_t - \boldsymbol{\mu}_t, \boldsymbol{\eta}_{t-1}) && \boldsymbol{\eta}\mapsto D_{\boldsymbol{\eta}} \text{ decr.} && (17)\\
&=\; \Phi(\boldsymbol{R}_t - \boldsymbol{\mu}_{t-1} - 4\boldsymbol{\eta}_{t-1}\boldsymbol{\sigma}^2\Delta v_t, \boldsymbol{\eta}_{t-1}) && \text{by def. of } \boldsymbol{\mu}_t\\
&=\; \Phi(\boldsymbol{R}_{t-1} + \langle \boldsymbol{w}_t, \boldsymbol{\mu}_t\rangle - \boldsymbol{m}_t - \boldsymbol{\mu}_{t-1}, \boldsymbol{\eta}_{t-1}) && \text{by def. of } \Delta v_t\\
&=\; \max_{\boldsymbol{w}\in\mathcal{P}(K)} \langle \boldsymbol{w}, \boldsymbol{R}_{t-1} + \langle \boldsymbol{w}_t, \boldsymbol{m}_t\rangle - \boldsymbol{m}_t - \boldsymbol{\mu}_{t-1}\rangle - D_{\boldsymbol{\eta}_{t-1}}(\boldsymbol{w}, \boldsymbol{u}) && \text{by def. of } \Phi\\
&=\; \langle \boldsymbol{w}_t, \boldsymbol{R}_{t-1} + \langle \boldsymbol{w}_t, \boldsymbol{m}_t\rangle - \boldsymbol{m}_t - \boldsymbol{\mu}_{t-1}\rangle - D_{\boldsymbol{\eta}_{t-1}}(\boldsymbol{w}_t, \boldsymbol{u}) && \text{by def. of } \boldsymbol{w}_t\\
&=\; \langle \boldsymbol{w}_t, \boldsymbol{R}_{t-1} - \boldsymbol{\mu}_{t-1}\rangle - D_{\boldsymbol{\eta}_{t-1}}(\boldsymbol{w}_t, \boldsymbol{u}) && \langle \boldsymbol{w}_t, \boldsymbol{m}_t\rangle \text{ cancels}\\
&\le\; \max_{\boldsymbol{w}\in\mathcal{P}(K)} \langle \boldsymbol{w}, \boldsymbol{R}_{t-1} - \boldsymbol{\mu}_{t-1}\rangle - D_{\boldsymbol{\eta}_{t-1}}(\boldsymbol{w}, \boldsymbol{u}) && \text{since } \boldsymbol{w}_t \in \mathcal{P}(K) && (18)\\
&=\; \Phi(\boldsymbol{R}_{t-1} - \boldsymbol{\mu}_{t-1}, \boldsymbol{\eta}_{t-1}) \;=\; \Phi_{t-1} && \text{by def. of } \Phi, \Phi_t.
\end{aligned}$$

Hence, $\Phi_t \le \Phi_{t-1}$, as we were to show. □

Proof of Proposition 2.2. Lemma 2.1 shows that the potential $t \mapsto \Phi(\boldsymbol{R}_t - \boldsymbol{\mu}_t, \boldsymbol{\eta}_t)$ is decreasing in t and that consequently $\Phi(\boldsymbol{R}_t - \boldsymbol{\mu}_t, \boldsymbol{\eta}_t) \le \Phi(\boldsymbol{R}_0 - \boldsymbol{\mu}_0, \boldsymbol{\eta}_0) = -D_{\boldsymbol{\eta}_0}(\boldsymbol{w}_1, \boldsymbol{u})$. The maximal nature of the definition of Φ implies that, for any probability distribution $\boldsymbol{p} \in \mathcal{P}(K)$,

$$\langle \boldsymbol{p}, \boldsymbol{R}_t\rangle \le \langle \boldsymbol{p}, \boldsymbol{\mu}_t\rangle + D_{\boldsymbol{\eta}_t}(\boldsymbol{p}, \boldsymbol{u}) - D_{\boldsymbol{\eta}_0}(\boldsymbol{w}_1, \boldsymbol{u}). \tag{19}$$

The second claim contained in (8) follows from the special case where $\boldsymbol{p} = \boldsymbol{\delta}_k$, the probability distribution that puts all of its mass on expert k, and by bounding the last term in (19) by zero. The last statement contained in (9) is proven in Lemma F.3. This is all that we had set ourselves to prove. □

C.2 Tuning, proof of Proposition 2.3

Proof of Proposition 2.3. The main tool that is employed here to derive the regret bounds is Proposition 2.2. The fact that the learning rates at hand are decreasing is a consequence of Lemma F.4; we give more details in the following. A slightly stronger result than what we claim could be obtained by replacing directly the learning rates in Proposition 2.2. However, the result is not amenable to an easy interpretation, and we use upper bounds on the learning rates and their reciprocals. Recall that $\gamma_k = 8\frac{\sigma_{\max}^2}{\sigma_k^2}\ln(1/u_k)$. The learning rate is of the form $\eta_{t,k} = \eta_{0,k}H_{1,k}(v_t) = \eta_{0,k}h(v_t/\gamma_k)$ with $h(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left[\sqrt{\frac{x^2}{1+x}}\right] = \frac{x+2}{2(1+x)^{3/2}}$ and $\eta_{0,k} = 1/(2\sigma_{\max})$. That this choice of learning rate is indeed nondecreasing can be proven using Lemma F.4. We use the following two elementary inequalities in relation to this specific choice of function h.

Lemma C.1. Let $x \geq 0$. The function $h(x) = \frac{x+2}{2(1+x)^{3/2}}$ satisfies

$$\int_0^x h(x')\mathrm{d}x' \leq \min\left\{x, \sqrt{x}\right\} \leq \max\left\{1, \sqrt{x}\right\}, \quad \text{and} \tag{20}$$

$$\frac{1}{h(x)} \leq \left\{\begin{array}{ll} 1+x & \text{if } x \leq 1 \\ 2\sqrt{x} & \text{if } x > 1 \end{array}\right\} \leq 2\max\left\{1, \sqrt{x}\right\}, \tag{21}$$

where the first minimum is equalized at $x = 1$.

Using these upper bounds and the choice $u_k = \pi_k\frac{\sigma_{\min}}{\sigma_k}$ in Proposition 2.2 gives the claimed result. Indeed, recall that Proposition 2.2 implies that

$$R_{t,k} \leq \sigma_k^2\eta_{0,k}\int_0^{v_t} h(x/\gamma_k)\mathrm{d}x + \frac{\ln(1/u_k)}{\eta_{t,k}} + \sum_{j\in K}\frac{u_j}{\eta_{t,j}} + \sigma_k^2\eta_{0,k}\max_{s\leq t}\Delta v_s. \tag{22}$$

We now focus on bounding each term. First,

$$\int_0^{v_t} h(x/\gamma_k)\mathrm{d}x = \gamma_k\int_0^{v_t/\gamma_k} h(x')\mathrm{d}x' \leq \max\left\{\gamma_k, \sqrt{v_t\gamma_k}\right\}.$$

Consequently,

$$\sigma_k^2\eta_{0,k}\int_0^{v_t} h(x/\gamma_k)\mathrm{d}x \leq \sigma_k\sqrt{2v_t\ln(1/u_k)} + 4\sigma_{\max}\ln(1/u_k). \tag{23}$$

Next,

$$\frac{1}{\eta_k} = \frac{2\sigma_{\max}}{h(v/\gamma_k)} \leq 4\sigma_{\max}\max\left\{1, \sqrt{\frac{v_t}{\gamma_k}}\right\} \leq 4\sigma_{\max} + \sigma_k\sqrt{\frac{2v_t}{\ln(1/u_k)}}.$$

With this at hand, the second and third term on the right hand side of (22) can be bounded by

$$\frac{\ln(1/u_k)}{\eta_{t,k}} \leq \sigma_k\sqrt{2v_t\ln(1/u_k)} + 4\sigma_{\max}\ln(1/u_k), \text{ and} \tag{24}$$

$$\sum_{j\in K}\frac{u_j}{\eta_j} \leq c_{\boldsymbol{\sigma},\boldsymbol{\pi}}\sigma_{\min}\sqrt{2v_t} + 4\sigma_{\max} \tag{25}$$

with $c_{\boldsymbol{\sigma},\boldsymbol{\pi}} = \sum_{k\in K}\pi_k\left(\frac{1}{\sqrt{\ln(1/u_k)}}\right)$. Replace (23), (24), and (25) in the the regret bound (22) to obtain the result. In order to prove the second claim we follow a similar path; we use Proposition 2.2 as our main tool. Recall that in this case the learning rate is of the form $\eta_{t,k} = \eta_{0,k}H_{2,k}(v_t)$ with $\eta_{0,k} = 1/(2\sigma_{\max})$ and

$$H_{2,k}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left[\sqrt{\alpha_k^2\left\{\left(1+\frac{x}{\alpha_k}\right)\ln\left(1+\frac{x}{\alpha_k}\right) - \frac{x}{\alpha_k}\right\} + \frac{x^2}{2(1+x/(2\gamma_k))}}\right]$$

with $\alpha_k = 32\frac{\sigma_{\max}^2}{\sigma_k^2}$ and $\gamma_k = \alpha_k\ln(1/\pi_k)$. The fact that $k \mapsto H_{2,k}(x)$ is decreasing follows from Lemma F.4 after performing the change of variable $x' = x/\alpha_k$. We use the inequalities for $H_{2,k}$ that are proven in the following lemma.

Lemma C.2. Let $\beta_k = \ln(1/\pi_k)$. The function $H_{2,k}$ satisfies

$$\int_0^x H_{2,k}(x')\mathrm{d}x' \le \sqrt{\alpha_k x(\ln(1+x/\alpha_k)+\beta_k)}, \text{ and} \tag{26}$$

$$\frac{1}{H_{2,k}(x)} \le 2\sqrt{\frac{x/\alpha_k}{\ln(1+x/\alpha_k)}}\sqrt{1+\frac{\min\{\beta_k, \frac{1}{2}\frac{x}{\alpha_k}\}}{\ln(1+x/\alpha_k)}}. \tag{27}$$

We can now compute the analogs of (23), (24), and (25) to obtain that

$$\sigma_k^2\eta_{0,k}\int_0^{v_t} H_{2,k}(x)\mathrm{d}x \le 2\sigma_k\sqrt{2v_t\left(\ln\left(1+\tfrac{\sigma_k^2}{32\sigma_{\max}^2}v_t\right)+\ln(1/\pi_k)\right)},$$

$$\frac{\ln(1/\pi_k)}{\eta_{t,k}} \le \sigma_k\ln(1/\pi_k)\sqrt{\frac{v_t}{2\ln\left(1+\frac{\sigma_k^2}{32\sigma_{\max}^2}v_t\right)}}\left(1+\sqrt{\frac{\min\{\ln(1/\pi_k), \frac{\sigma_k^2}{16\sigma_{\max}^2}v_t\}}{\ln\left(1+\frac{\sigma_k^2}{32\sigma_{\max}^2}v_t\right)}}\right),$$

$$\sum_{j\in K}\frac{u_j}{\eta_j} \le \sum_{j\in K}\pi_j\left(\sigma_j\sqrt{\frac{v_t}{2\ln\left(1+\frac{\sigma_j^2}{32\sigma_{\max}^2}v_t\right)}}\left(1+\frac{\sqrt{\min\{\ln(1/\pi_j), \frac{\sigma_j^2}{16\sigma_{\max}^2}v_t\}}}{\sqrt{\ln\left(1+\frac{\sigma_j^2}{32\sigma_{\max}^2}v_t\right)}}\right)\right),$$

and employ them in Proposition 2.2 to obtain the result. □

Proof of Lemma C.1. The relations are clear for $x = 0$. Let $x > 0$. Recall that $\int_0^x h(x')\mathrm{d}x' = \frac{x}{\sqrt{1+x}}$. We start by proving (20). The fact that $\frac{x}{\sqrt{1+x}} \le x$ is clear. The inequality $\frac{x}{\sqrt{1+x}} \le \sqrt{x}$ follows from dividing both sides of the inequality $x \le \sqrt{x^2+x}$ by $\sqrt{1+x}$. Thus, the first inequality in (20) follows, and the second is direct after observing that $x \le \sqrt{x} \le 1$ for $x \le 1$. We now turn to proving (21). Recall that $1/h(x) = \frac{2(1+x)^{3/2}}{2+x}$. We start by showing that $1/h(x) \le 1+x$ for all $x > 0$. Note that $\frac{2(1+x)^{3/2}}{2+x} = (1+x)\frac{2\sqrt{1+x}}{2+x}$. Thus, the claim holds if and only if $2\sqrt{1+x} \le 2+x$, which is easily checked to be the case. Now let $x > 1$. Observe that the second claim in the first inequality holds if and only if $2(1+x)^{3/2} \le 2\sqrt{x}(2+x)$. Square both members and rearrange to conclude that the sought relation holds if and only if $0 \le 4x^2+4x-4$, which is the case as $x > 1$. The second inequality in (21) is clear. □

Proof of Lemma C.2. The inequalities contained in (26) and (27) are a consequence of the fact that

$$\int_0^x H_2(x')\mathrm{d}x' = \sqrt{\alpha^2\left\{\left(1+\frac{x}{\alpha}\right)\ln\left(1+\frac{x}{\alpha}\right)-\frac{x}{\alpha}\right\}+\frac{x^2}{2(1+x/(2\gamma))}}$$

and the inequalities

$$(1+x')\ln(1+x')-x' \le x'\ln(1+x') \text{ and } \frac{a^2x'^2}{2(1+x'/(2b))} \le \min\{bx', \tfrac{1}{2}a^2x'^2\},$$

that hold for $x', a, b \ge 0$. From this, (26) is immediate once we use the substitutions $x' = x/\alpha$, $a = \alpha$, and $b = \beta$. To prove (27), use the same substitution and estimate

$$\begin{aligned}
\frac{1}{H(x')} &= 2\frac{\sqrt{(1+x')\ln(1+x')-x'+\frac{x'^2}{2(1+x'/(2b))}}}{\ln(1+x')+\frac{1}{2}\frac{2x'+x'^2/(2b)}{(1+x'/(2b))^2}} \\
&\le 2\frac{\sqrt{x'\ln(1+x')+\min\{bx', \frac{1}{2}x'^2\}}}{\ln(1+x')} \\
&= 2\sqrt{\frac{x'}{\ln(1+x')}}\sqrt{1+\frac{\min\{b, \frac{1}{2}x'\}}{\ln(1+x')}} \\
&\le 2\sqrt{\frac{x'}{\ln(1+x')}}\left(1+\sqrt{\frac{\min\{b, \frac{1}{2}x'\}}{\ln(1+x')}}\right).
\end{aligned}$$

This is all we set ourselves to prove. □

D Optimism, proof of Proposition 4.1

Proof of Proposition 4.1. In Lemma 2.1 we show that the potential $t \mapsto \Phi_t$ is decreasesing. The rest of the proof is identical to that of Proposition 2.3 after multiplying all scales by 2. The "furthermore" claim follows from a direct modification of Proposition G.1. □

E Luckiness

This appendix contains the proofs of the luckiness results in Section 3.

E.1 Proof of Theorem 3.1

Proof of Theorem 3.1. Let $s_t = \sum_{s\le t} \frac{\mathrm{var}_{\tilde{\boldsymbol{w}}_s}(\boldsymbol{\ell}_t)}{\langle \tilde{\boldsymbol{w}}_s, \boldsymbol{\sigma}^2\rangle}$. It is shown in Proposition G.1 that v_t can be bounded in terms of s_t. Indeed, in any case $v_t \le 4$, and because the learning rates are low enough at the start of the protocol, namely $\eta_{t,k} \le 1/(2\sigma_{\max})$, the upper bound $v_t \le 4s_t$ also holds. A verification of the regret bound obtained in Proposition 2.2 shows that it is increasing in v_t, and consequently the same regret bound holds once we replace v_t with the larger quantity $4s_t$, and the proof of Proposition 2.3 can be repeated with no problems. Consequently the regret bounds in Proposition 2.3 are available with $4s_t$ occupying the place of v_t. The next step that we follow is to show that $\mathbf{E}_{\mathbf{P}}[s_t] \lesssim \mathbf{E}_{\mathbf{P}}[R_{t,k^*}]$, which is done in the following lemma.

Lemma E.1. Under Massart's condition (see Definition 1.3),

$$\mathbf{E}_{\mathbf{P}}[s_t] \le k_{\mathrm{M}} \mathbf{E}_{\mathbf{P}}[R_{t,k^*}],$$

where $k_{\mathrm{M}} = c_{\mathrm{M}} \max_{i,j\in K} \sup_{v\ge 0} \left\{ \frac{\eta_{0,i} H_i(v)}{\eta_{0,j} H_j(v) \sigma_j^2} \right\}$ satisfies

$$k_{\mathrm{M}} \le 2c_{\mathrm{M}} \max_{i,j\in K} \left\{ \frac{1}{\sigma_i \sigma_j} \frac{\ln(1/\pi_i) + \ln(\sigma_i/\sigma_{\min})}{\ln(1/\pi_j) + \ln(\sigma_j/\sigma_{\min})} \right\}.$$

From the previous discussion, a small modification of Proposition 2.3 shows that this tuning guarantees a regret bound of the form

$$R_{t,k^*} \le a' \sqrt{s_t} + b \tag{28}$$

with $a' = 4\sigma_{k^*}\sqrt{2\ln(1/u_{k^*})} + 2\sqrt{2} c_{\boldsymbol{\sigma},\boldsymbol{\pi}} \sigma_{\min}$ and $b = 8\sigma_{\max}\ln(1/u_{k^*}) + 4\sigma_{\max} + 2\sigma_{k^*}$. Take $\mathbf{P}$-expectations in the last display, use the concavity of $x \mapsto \sqrt{x}$ to invoke Jensen's inequality, and use Lemma E.1 to obtain that

$$\mathbf{E}_{\mathbf{P}}[R_{t,k^*}] \le a' \sqrt{k_{\mathrm{M}} \mathbf{E}_{\mathbf{P}}[R_{t,k^*}]} + b. \tag{29}$$

This implies that the expected regret satisfies $\mathbf{E}_{\mathbf{P}}[R_{t,k^*}] \lesssim 1$. Indeed, using Lemma E.2 yields that

$$\mathbf{E}_{\mathbf{P}}[R_{t,k^*}] \le a'^2 k_{\mathrm{M}} + b. \tag{30}$$

The upper bound for k_{M} is contained in Lemma E.3. Given the definition of a in the claim, this is what we set ourselves to prove. □

Proof of Lemma E.1. Recall that $s_t = \sum_{s\le t} \Delta s_s = \sum_{s\le t} \frac{\mathrm{var}_{\tilde{\boldsymbol{w}}_s}(\boldsymbol{\ell}_s)}{\langle \tilde{\boldsymbol{w}}_s, \boldsymbol{\sigma}^2\rangle}$ with the weights $\tilde{w}_{t,k} \propto w_{t,k}\eta_{t-1,k}$. Define $\ell_s^* = \ell_{s,k^*}$ to be the loss of the best expert k^*, and use that the variance $\mathrm{var}_{\tilde{\boldsymbol{w}}_s}(\boldsymbol{\ell}_s)$ satisfies $\mathrm{var}_{\tilde{\boldsymbol{w}}_s}(\boldsymbol{\ell}_s) \le \langle \tilde{\boldsymbol{w}}_s, (\boldsymbol{\ell}_s - \ell_s^*)^2\rangle$ to obtain the estimate

$$\Delta s_s \le \frac{\langle \tilde{\boldsymbol{w}}_s, (\boldsymbol{\ell}_s - \ell_s^*)^2\rangle}{\langle \tilde{\boldsymbol{w}}_s, \boldsymbol{\sigma}^2\rangle}.$$

Recall that, under $\mathbf{P}$, the loss vector $\boldsymbol{\ell}_s$ is assumed to be independent of $\boldsymbol{\ell}_{s-1}$. This implies that

$$\begin{aligned}\mathbf{E}_{\mathbf{P}}[\Delta s_s] &\le \sum_{k\in K} \left(\mathbf{E}_{\mathbf{P}}\left[\frac{\tilde{w}_{s,k}}{\langle \tilde{\boldsymbol{w}}_s, \boldsymbol{\sigma}^2\rangle}\right] \mathbf{E}_{\mathbf{P}}\left[(\ell_{s,k} - \ell_s^*)^2\right] \right) \\ &\le c_{\mathrm{M}} \sum_{k\in K} \left(\mathbf{E}_{\mathbf{P}}\left[\frac{\tilde{w}_{s,k}}{\langle \tilde{\boldsymbol{w}}_s, \boldsymbol{\sigma}^2\rangle}\right] \mathbf{E}_{\mathbf{P}}\left[\ell_{s,k} - \ell_s^*\right] \right).\end{aligned}$$

Sum the last display over rounds, and use the fact that the weights $\tilde{w}_{t,k} \propto w_{t,k}\eta_{t-1,k}$ to deduce that

$$\mathbf{E}_{\mathbf{P}}\left[s_t\right] \leq c_{\mathrm{M}} \left\| \max_{s\leq t} \left\{ \frac{\max_{k\in K} \eta_{s-1,k}}{\min_{k\in K} \eta_{s-1,k}\sigma_k^2} \right\} \right\|_{\infty} \mathbf{E}_{\mathbf{P}}\left[R_{t,k^*}\right],$$

where $\|\cdot\|_\infty$ is the infinity norm w.r.t. $\mathbf{P}$ (recall that $\eta_{t-1,k}$ depend on the random losses $\boldsymbol{\ell}_{t-1}$). Since, for any $s = 1, \ldots,$ and $k \in K$, the learning rate $\eta_{s-1,k} = \eta_{0,k}H_k(v)$, we can deduce that $c_{\mathrm{M}} \left\| \max_{s\leq t} \left\{ \frac{\max_{k\in K} \eta_{s-1,k}}{\min_{k\in K} \eta_{s-1,k}\sigma_k^2} \right\} \right\|_{\infty} \leq k_{\mathrm{M}}$, where k_{M} is as defined in the claim of the proposition. This implies what we set ourselves to prove. □

Lemma E.2. Let $y, a, b \geq 0$. If $y^2 \leq ay + b$ then $y \leq b + \sqrt{a}$.

Proof. The quadratic polynomial $y^2 - ay - b$ has a zero at $y^* = \frac{b+\sqrt{b^2+4a}}{2} \leq b + \sqrt{a}$. Hence, if $y^2 \leq ay + b$, then $y \leq y^*$, and the result follows. □

E.2 Proof of Theorem 3.2

Proof of Theorem 3.2. Call $\Delta_{t,k} = L_{s,k} - L_{s,k^*}$, and $d_k = \mathbf{E}_{\mathbf{P}}[\Delta_{t,k}]$. Since $\boldsymbol{\ell}_s$ and $\boldsymbol{\ell}_{s-1}$ are independent, the expected value of the increment of the regret R_{t,k^*} is

$$\mathbf{E}_{\mathbf{P}}[\Delta R_{t,k^*}] = \sum_{k\neq k^*} \mathbf{E}_{\mathbf{P}}[w_{t,k}]\mathbf{E}_{\mathbf{P}}[\ell_{t,k} - \ell_{t,k^*}] \tag{31}$$

$$= \sum_{k\neq k^*} \mathbf{E}_{\mathbf{P}}[w_{t,k}]d_k. \tag{32}$$

We seek to prove that for $k \neq k^*$, in an event $\Omega_{t,k}$ which we define next, the weight $w_{t,k}$ is small. Define, for each $k \neq k^*$ and $t \geq 1$, the event $\Omega_{t,k}$ by

$$\Omega_{t,k} = \left\{ L_{t,k^*} - L_{t,k} \leq \mu_{t,k} - \mu_{t,k^*} - \frac{1}{\eta_{t,k^*}} \ln\left(1/\pi_{k^*}\right) - \frac{1}{\eta_{t,k}} \ln\left(\frac{1}{\pi_k \varepsilon_t}\right) \right\},$$

for deterministic constants $\varepsilon_t = 1/t^2$. Recall that the weights have the form $w_{t,k} = \pi_k \mathrm{e}^{-\eta_{t,k}(L_{t,k}+\mu_{t,k}+a_t^*)}$, where a_t^* is such that $\sum_k w_{t,k} = 1$. Next, we show that, in each event $\Omega_{t,k}$, for carefully chosen $\tilde{a}_t = -\frac{1}{\eta_{t,k^*}} \ln\left(1/\pi_{k^*}\right) - L_{t,k^*} - \mu_{t,k^*}$, it holds that $a_t^* \geq \tilde{a}_t$. Indeed, this follows because, by design $\pi_{k^*}\mathrm{e}^{-\eta_{t,k}(L_{t,k}+\mu_{t,k}+\tilde{a}_t)} = 1$, and consequently,

$$\sum_{k\in K} \pi_k(\mathrm{e}^{-\eta_{t,k}(L_t+\mu_{t,k}+\tilde{a}_{t,k})}) \geq 1 = \sum_{k\in K} \pi_k(\mathrm{e}^{-\eta_{t,k}\left(L_t+\mu_{t,k}+a_{t,k}^*\right)}),$$

which implies $a_t^* \geq \tilde{a}_t$. We use this in the weight $w_{t,k}$ of expert k to conclude that

$$\mathbf{E}_{\mathbf{P}}[w_{t,k}\mathbf{1}\left\{\Omega_{t,k}\right\}] \leq \pi_k(\mathrm{e}^{-\eta_{t,k}(L_t+\mu_{t,k}+\tilde{a}_{t,k})}) = \pi_k\varepsilon_t,$$

hence

$$\mathbf{E}_{\mathbf{P}}[w_{t,k}] \leq \pi_k\varepsilon_t + \mathbf{P}\left\{\Omega_{t,k}^c\right\}.$$

Consequently, using (32),

$$\mathbf{E}_{\mathbf{P}}[R_{t,k^*}] = \sum_{s\leq t}\sum_{k\neq k^*} d_k\mathbf{E}_{\mathbf{P}}[w_{t,k}]d_k \tag{33}$$

$$\leq \sum_{s\leq t}\sum_{k\neq k^*} \{\pi_k d_k\varepsilon_s + d_k\mathbf{P}\{\Omega_{s,k}\}\} \tag{34}$$

$$\leq 2\sum_{k\in K} \pi_k d_k + \sum_{s\leq t}\sum_{k\neq k^*} d_k\mathbf{P}\{\Omega_{s,k}^c\}, \tag{35}$$

where we used that $\sum_s \varepsilon_s \leq \pi^2/6 \leq 2$. We now focus on bounding the probabilities $\mathbf{P}\{\Omega_{s,k}^c\}$. We use that $\Delta\mu_{t,k} \geq 0$ to deduce that

$$\mathbf{P}\left\{\Omega_{t,k}^c\right\} = \mathbf{P}\left\{ L_{t,k^*} - L_{t,k} > \mu_{t,k} - \mu_{t,k^*} - \frac{1}{\eta_{t,k^*}} \ln\left(1/\pi_{k^*}\right) - \frac{1}{\eta_{t,k}} \ln\left(1/\varepsilon_t\right) \right\}$$

$$\leq \mathbf{P}\left\{ L_{t,k^*} - L_{t,k} > -\mu_{t,k^*} - \frac{1}{\eta_{t,k^*}} \ln(1/\pi_{k^*}) - \frac{1}{\eta_{t,k}} \ln(1/\varepsilon_t) \right\}.$$

In order to continue, we derive an upper bound on μ_{t,k^*}, and a lower bound on $\eta_{t,k}$, and η_{t,k^*} consisting of deterministic functions of time. Recall from Lemma G.1 that $v_t \leq 4t$, and that Lemma C.2 can be used to bound μ_{t,k^*} in terms of the integral of the function $x \mapsto H_{2,k^*}(x)$ (see proof of Proposition 2.3) to obtain that

$$\begin{aligned}\mu_{t,k^*} &\leq \sigma_{k^*}^2 \eta_{0,k^*} \int_0^{4t} H_{2,k^*}(v)\mathrm{d}v + 4\sigma_{k^*}^2 \eta_{0,k^*} \\ &\leq 4\sigma_{k^*}\sqrt{2t(\ln(1+t/8)+\ln(1/\pi_*))} + 4\sigma_{k^*}\end{aligned}$$

Now fix $k \in K$, and use again that $v_t \leq 4t$ and that $x \mapsto H_{2,k}(x)$ is decreasing (see Lemma F.4) to deduce that $\eta_{t,k} = \eta_{0,k} H_k(v_t) \geq \eta_{0,k} H_k(4t)$. From these observations, $\mathbf{P}\left\{\Omega_{t,k}^c\right\}$ can be further bounded by

$$\mathbf{P}\left\{\Omega_{t,k}^c\right\} \leq \mathbf{P}\left\{L_{t,k^*} - L_{t,k} > -F_k(t)\right\},$$

where the complicated $F_k(t) = 4\sigma_{k^*}\sqrt{2t(\ln(1+t/8)+\ln(1/\pi_{k^*}))} + 4\sigma_{k^*} + \frac{\ln(1/\pi_{k^*})}{\eta_{0,k^*}H_{2,k^*}(4t)} + \frac{\ln(1/\varepsilon_t)}{\eta_{0,k}H_{2,k}(4t)}$ is a deterministic function of time. Recall that the gap $d_{\min}$ was defined as $d_{\min} = \min_{k\neq k^*} d_k$ and that is assumed to be strictly positive. Recall that $\Delta_{t,k} = L_{t,k} - L_{t,k^*}$ is the gap in losses between expert k and the best expert k^*. Hoeffding's inequality implies that

$$\begin{aligned}\mathbf{P}\left\{L_{t,k^*} - L_{t,k} > F_k(t)\right\} &= \mathbf{P}\left\{td_k - \Delta_{t,k} > td_k - F_k(t)\right\} \\ &\leq \exp\left(-\frac{t}{2\sigma_{\max}^2}((d_k - F_k(t)/t)_+)^2\right) \\ &= \exp\left(-\frac{td_k^2}{2\sigma_{\max}^2}((1 - F_k(t)/(d_k t))_+)^2\right),\end{aligned}$$

where $x \mapsto (x)_+ = \max\{0, x\}$. We now seek a bound on the point $t_k^\star$ at which $F_k(t_k^\star)/t_k^\star = d_k/2$. For these values $t_k^\star$, we have, using (35), that

$$\mathbf{E}_{\mathbf{P}}[R_{t,k^*}] \leq 2\sum_{k\in K}\pi_k d_k + \sum_{k\neq k^*}(t_k^\star d_k + \sum_{s\geq t^\star} d_k \mathbf{P}\{\Omega_{t,k}^c\}). \tag{36}$$

We now concentrate on bounding $t_k^\star$ and the probability of the event $\Omega_{t,k}^c$ for each k. In the limit that $d_k \to 0$, the time $t_k^\star \to \infty$. A quick computation shows that, as $t \to \infty$, $H_{2,k}(t) \sim \sqrt{\frac{2\ln t}{t}}$, and, in the same limit, $4\sigma_{k^*}\sqrt{2t(\ln(1+t/8)+\ln(1/\pi_*))} + 4\sigma_{k^*} \sim 4\sigma_{k^*}\sqrt{2t\ln t}$. Hence, as $t \to \infty$, the function F_k satisfies $F_k(t) \sim (4\sigma_{k^*} + 2\sigma_{\max})\sqrt{2t\ln t}$. We now give a bound on the solution $x_k^\star$ to the equation $xd_k/2 = (4\sigma_{k^*} + 2\sigma_{\max})\sqrt{2x\ln x}$ that holds asymptotically as $d_k \to 0$. Call $c = d_k/(2\sqrt{2}(4\sigma_{k^*} + 2\sigma_{\max}))$. Our equation of interest can be rewritten as $xc^2 = \ln x$. Linearize $x\ln x$ around $x = 2/c^2$, and use its concavity to obtain that $\ln(x) \leq \ln(2/c^2) + (c^2/2)(x - 2/c^2)$. With this estimate at hand, the solution to the simpler, linear equation $xc^2 = \ln(2/c^2) + (c^2/2)(x - 2/c^2)$ is an upper bound on $x_k^\star$. From this discussion it follows that the point $t_k^\star$ of interest satisfies $t_k^\star \leq 2\frac{\ln(1/c^2)}{c^2} - \frac{2}{c^2}$. Hence, as $d_k \to 0$,

$$d_k t_k^\star \leq \frac{2(4\sigma_{k^*} + 2\sigma_{\max})^2}{d_k}\left\{\ln\left(\frac{8(4\sigma_{k^*} + 2\sigma_{\max})^2}{d_k^2}\right) - 1\right\} = O\left(\frac{\sigma_{\max}^2}{d_k}\ln\left(\frac{\sigma_{\max}^2}{d_k^2}\right)\right). \tag{37}$$

We deduce that, as $d_k \to 0$, for $t \geq t_k^\star$ and any $k \neq k^*$, the probability $\mathbf{P}\{\Omega_{t,k}^c\} \leq \exp\left(-\frac{t}{8\sigma_{\max}^2}d_k^2\right)$. We sum $\mathbf{P}\{\Omega_{t,k}^c\}$ over rounds to conclude that

$$\mathbf{E}_{\mathbf{P}}[R_{t,k^*}] \leq 2\sum_{k\in K}\pi_k d_k + \sum_{k\neq k^*}(t_k^\star d_k + \sum_{s\geq t^\star} d_k \mathbf{P}\{\Omega_{t,k}^c\}). \tag{38}$$

We now concentrate on bounding $t_k^\star$ and the probability of the event $\Omega_{t,k}^c$ for each k. In the limit that $d_k \to 0$, the time $t_k \to \infty$. A quick computation shows that, as $t \to \infty$, $H_{2,k}(t) \sim \sqrt{\frac{2\ln t}{t}}$, and, in the same limit, $4\sigma_{k^*}\sqrt{2t(\ln(1+t/8)+\ln(1/\pi_*))} + 4\sigma_{k^*} \sim 4\sigma_{k^*}\sqrt{2t\ln t}$. Hence, as $t \to \infty$, the function F_k satisfies $F_k(t) \sim (4\sigma_{k^*} + 2\sigma_{\max})\sqrt{2t\ln t}$. We know give a bound on the solution $x_k^\star$

to the equation $xd_k/2 = (4\sigma_{k^*} + 2\sigma_{\max})\sqrt{2x\ln x}$ that holds asymptotically as $d_k \to 0$. Call $c = d_k/(2\sqrt{2}(4\sigma_{k^*} + 2\sigma_{\max}))$. Our equation of interest can be rewritten as $xc^2 = \ln x$. Linearize $x\ln x$ around $x = 2/c^2$, and use its concavity to obtain that $\ln(x) \le \ln(2/c^2) + (c^2/2)(x - 2/c^2)$. With this estimate at hand, the solution to the simpler, linear equation $xc^2 = \ln(2/c^2) + (c^2/2)(x - 2/c^2)$ is an upper bound on $x_k^\star$. From this discussion it follows that the point $t_k^\star$ of interest satisfies

$$t_k^\star \le 2\frac{\ln(1/c^2)}{c^2} - \frac{2}{c^2} = O\left(\frac{\sigma_{\max}}{d_k^2}\ln\frac{\sigma_{\max}^2}{d_k^2}\right), \tag{39}$$

as $d_k \to 0$. Hence, again, as $d_k \to 0$,

$$\sum_{t\ge t^\star} d_k \mathbf{P}\left\{\Omega_{t,k}^c\right\} \le \sum_{t\ge t^\star} d_k \mathrm{e}^{-td_k^2/(8\sigma_{\max}^2)} \le \frac{d_k}{1-\mathrm{e}^{-d_k^2/(8\sigma_{\max}^2)}}. \tag{40}$$

We use (39) and (40) in (38), and the fact that $d/(1+\mathrm{e}^{d^2/\sigma^2}) = O(\sigma^2/d)$ as $d \to 0$ to conclude the proof. □

E.3 In Lemma E.1, k_{M} is bounded

Lemma E.3. In Lemma E.1, the constant k_{M} is bounded for Tuning 1, shown in Figure 2. More precisely,

$$k_{\mathrm{M}} \le 2\max_{i,j\in K}\left\{\frac{1}{\sigma_i\sigma_j}\frac{\ln(1/\pi_i)+\ln(\sigma_i/\sigma_{\min})}{\ln(1/\pi_j)+\ln(\sigma_j/\sigma_{\min})}\right\}.$$

Proof. Recall that in both tunings of the algorithm we use the starting learning rate $\eta_{0,k} = 1/(2\sigma_{\max})$, a constant over the experts. As long as this is the case, the constant of interest k_{M} can be bounded by

$$k_{\mathrm{M}} \le \max_{i,j\in K}\sup_v \frac{H_i(v)}{\sigma_j^2 H_j(v)}. \tag{41}$$

Recall from Figure 2 that $H_{1,k}(v) = \frac{v/\gamma_k+2}{2(1+v/\gamma_k)^{3/2}}$ with $\gamma_k = 8\frac{\sigma_{\max}^2}{\sigma_k^2}(\ln(1/\pi_k) + \ln(\sigma_k/\sigma_{\min}))$. We can estimate the ratio

$$\begin{aligned}
\frac{H_i(v)}{H_j(v)} &= \frac{v/\gamma_i+2}{(1+v/\gamma_i)^{3/2}}\frac{(1+v/\gamma_j)^{3/2}}{v/\gamma_j+2} \\
&\le \frac{2v/\gamma_i+2}{(1+v/\gamma_i)^{3/2}}\frac{(1+v/\gamma_j)^{3/2}}{v/\gamma_j+1} \\
&= 2\sqrt{\frac{1+v/\gamma_j}{1+v/\gamma_i}} \\
&\le 2\max\left\{1, \sqrt{\frac{\gamma_i}{\gamma_j}}\right\}.
\end{aligned}$$

Hence

$$k_{\mathrm{M}} \le 2\max_{i,j\in K}\left\{\frac{1}{\sigma_i\sigma_j}\frac{\ln(1/\pi_i)+\ln(\sigma_i/\sigma_{\min})}{\ln(1/\pi_j)+\ln(\sigma_j/\sigma_{\min})}\right\},$$

as it was to be shown. □

F Technical Lemmas

In this appendix we gather technical results used in previous sections.

F.1 For showing that the potential decreases

Lemma F.1. For fixed $\boldsymbol{X}$, the function $\boldsymbol{\eta} \mapsto \Phi(\boldsymbol{X}, \boldsymbol{\eta})$ is increasing, that is, if $\eta_k \le \eta_k'$, then, for fixed X, it holds that $\Phi(\boldsymbol{X}, \boldsymbol{\eta}) \le \Phi(\boldsymbol{X}, \boldsymbol{\eta}')$.

Proof. It follows from the definition of Φ and the fact that, for all $x \geq 0$, the function $x \mapsto -\ln(x) - 1 + x$ is nonnegative. Indeed, for any $w \in \mathcal{P}(K)$, it holds that

$$\begin{aligned} D_{\boldsymbol{\eta}}(\boldsymbol{w}, \boldsymbol{u}) &= \sum_{k \in K} w_k \left(\frac{\ln(w_k/u_k) - (1 - u_k/w_k)}{\eta_k} \right) \\ &\geq \sum_{k \in K} w_k \left(\frac{\ln(w_k/u_k) - (1 - u_k/w_k)}{\eta'_k} \right) \\ &= D_{\boldsymbol{\eta}'}(\boldsymbol{w}, \boldsymbol{u}). \end{aligned}$$

The result follows from the definition of Φ contained in (4). □

Lemma F.2. Fix vectors $\boldsymbol{X}, \boldsymbol{m} \in \mathbb{R}^K$ and $\boldsymbol{u}, \boldsymbol{\eta} \in \mathbb{R}^K_+$. Let $\boldsymbol{w}$ be the optimum value $\boldsymbol{w} = \arg\max_{\boldsymbol{p} \in \mathcal{P}(K)} \langle \boldsymbol{p}, \boldsymbol{X} + \boldsymbol{m} \rangle - D_{\boldsymbol{\eta}}(\boldsymbol{p}, \boldsymbol{u})$. Then,

$$\Phi(\boldsymbol{X} + \boldsymbol{m} - \langle \boldsymbol{w}, \boldsymbol{m} \rangle, \boldsymbol{\eta}) \leq \Phi(\boldsymbol{X}, \boldsymbol{\eta})$$

Proof. The result follows from the chain of inequalities

$$\begin{aligned} \Phi(\boldsymbol{X} + \boldsymbol{m} - \langle \boldsymbol{w}, \boldsymbol{m} \rangle, \boldsymbol{u}) &= \langle \boldsymbol{w}, \boldsymbol{X} + \boldsymbol{m} - \langle \boldsymbol{w}, \boldsymbol{m} \rangle \rangle - D_{\boldsymbol{\eta}}(\boldsymbol{w}, \boldsymbol{u}) \\ &= \langle \boldsymbol{w}, \boldsymbol{X} \rangle - D_{\boldsymbol{\eta}}(\boldsymbol{w}, \boldsymbol{u}) \\ &\leq \Phi(\boldsymbol{X}, \boldsymbol{\eta}). \end{aligned}$$

□

F.2 For bounding μ with v

The following is the consequence of a standard result in the theory of Riemann integration.

Lemma F.3. Let $x \mapsto H(x)$ be a decreasing, positive, real, and continuous function such that $H(x) < \infty$ on $0 \leq x < \infty$. If $\Delta v_s \geq 0$ for $s = 1, 2, \ldots, t$ then

$$\sum_{s \leq t} H(v_{s-1}) \Delta v_s \leq \int_0^{v_t} H(x) \mathrm{d}x + (H(0) - H(v_t)) \max_{s \leq t} \Delta v_t,$$

where $v_t = \sum_{s \leq t} \Delta v_s$.

Proof. Because H is decreasing and $t \mapsto v_t = \sum_{s \leq t} \Delta v_t$ is nondecreasing,

$$\int_0^{v_t} H(x) \mathrm{d}x \geq \sum_{s \leq t} H(v_s) \Delta v_s.$$

Use this observation to deduce that

$$\begin{aligned} \sum_{s=1} H(v_s) \Delta v_s - \int_0^{v_t} H(x) \mathrm{d}x &\leq \sum_{s \leq t} (H(v_{s-1}) - H(v_s)) \Delta v_s \\ &\leq (H(0) - H(v_s)) \max_{s \leq t} \Delta v_s, \end{aligned}$$

which is what we set ourselves to prove. □

F.3 The learning rates decrease

Lemma F.4. The functions $f(x) = \frac{x+2}{2(1+x)^{3/2}}$ and $g(x) = \frac{\ln(1+x) + \frac{2x + x^2/a}{(1+x/a)^2}}{\sqrt{(1+x)\ln(1+x) - x + \frac{x^2}{2(1+x/a)}}}$ are decreasing in $x \geq 0$ for any fixed $a > 0$.

Proof. The function f is differentiable in $x \geq 0$, and its derivative is $f'(x) = -\frac{x+4}{(1+x)^{5/2}}$, a negative function. Thus, f is decreasing. We turn our attention to the function g. Let $h_1(x) = \ln(1 + x)$, $h_2(x) = \frac{2x + x^2/a}{(1+x/a)^2}$, and let $H_1(x) = \int_0^x h_1(s) \mathrm{d}s = (1+x)\ln(1+x) - x$, and $H_2(x) = \int_0^x h_2(s) \mathrm{d}s =$

$\frac{x^2}{2(1+x/a)}$. Then, the function g is of the form $h/(2\sqrt{H})$ with $h = h_1 + h_2$, and $H = H_1 + H_2$. Since $g(x)$ is differentiable in $x \geq 0$, it is enough to prove that $g' \leq 0$. We compute the derivative $g' = \frac{h'(x)\sqrt{H(x)} - h^2(x)/(2\sqrt{H(x)})}{H(x)}$ and conclude that $g' \leq 0$ if and only if

$$h'(x)H(x) \leq \frac{1}{2}h^2(x). \tag{42}$$

Since $h_1/\sqrt{H_1} = \frac{\sqrt{2}}{2}f(x/a)$, the analog of the last display holds for the pair h_1, H_1. We will show that the same holds true for the pair h_2, H_2 at the end of the proof. For now, use that (42) holds for both pairs, replace the definition of h and H, and conclude that it is enough to show that

$$h_1' H_2 + h_2' H_1 \leq h_1 h_2.$$

We now focus on showing that $\delta^\star = h_1 h_2 - h_1' H_2 - h_2' H_1$ is nonnegative. Define $\delta(x) = (1 + x/a)^3(x+1)2a^3\delta^\star(x)$. It is clear that it is sufficient to our purposes to show that $\delta(x) \geq 0$ for $x \geq 0$. Computation shows that

$$\begin{aligned}\delta(x) = a^3x^2 - 2\,a^2x^3 - ax^4 + 2\,a^3x + \\ \left((4\,a+1)x^4 + x^5 - 2\,a^3x + 5\,a^2x^2 + \left(5\,a^2 + 4\,a\right)x^3 - 2\,a^3\right)\ln\left(x+1\right).\end{aligned}$$

Since $\delta(0) = 0$, it is enough to show that its derivative is positive; that $\delta'(x) \geq 0$ for $x \geq 0$. Computation shows that

$$\begin{aligned}\delta'(x) = 2\,a^3x - a^2x^2 + x^4 + \\ \left(4\,(4\,a+1)x^3 + 5\,x^4 - 2\,a^3 + 10\,a^2x + 3\left(5\,a^2 + 4\,a\right)x^2\right)\ln\left(x+1\right).\end{aligned}$$

We now pay attention to the first three summands of the previous display. We use that $2a^3x - a^2x^2 + x^4 = x(2a^3 - a^2x + x^3) \geq \ln(1+x)(2a^3 - a^2x + x^3)$, which follows from the fact that last factor of the last equation is a depressed cubic that is nonnegative for $x, a \geq 0$. This fact, the previous display, and a short computation together imply that

$$\frac{\delta'(x)}{\ln(1+x)} \geq (16\,a+5)x^3 + 5\,x^4 + 9\,a^2x + 3\left(5\,a^2 + 4\,a\right)x^2,$$

which shows that $\delta'(x) \geq 0$ for $x \geq 0$. This in turn shows that the function δ is positive, that consequently the relation (42) holds, and finally, that the original function of interest g is decreasing. □

F.4 For bounding Δv in terms of Δs

Lemma F.5. Let $y, x, b \in \mathbb{R}$ be such that $b \geq 0$, $x \leq b$, and $y > 0$. Let $\varphi = \frac{\mathrm{e}^b - 1 - b}{\frac{1}{2}b^2} \geq 1$. Then the following statements hold.

1. For $g(y) = \frac{\varphi - 1 - \sqrt{(\varphi-1)^2 + 2\varphi y}}{\varphi} - \ln\left(\varphi - \sqrt{(\varphi-1)^2 + 2\varphi y}\right)$, we have

$$\mathrm{e}^{x - g(y)} - 1 - x \leq \frac{1}{2}\varphi x^2 - y$$

   any time that $y \leq \frac{2\varphi - 1}{\varphi}$.

2. Let $c = \varphi/(\varphi - 1)$. For any $0 < s < 1/c$ it holds that

$$\mathrm{e}^{x - s - h(cs)} - 1 - x \leq \frac{1}{2}\varphi x^2 - s,$$

   where

$$h(u) = -u - \ln(1-u) \leq \frac{1}{2}\frac{u^2}{1-u}$$

   for $0 < u < 1$.

Proof. Proving our claim is equivalent to proving that

$$g(z) \geq x - \ln\left(1 - z + x + \frac{1}{2}\varphi x^2\right).$$

The condition that $z < \frac{2\varphi-1}{2\varphi}$ ensures that the logarithm is well defined. The first claim follows because g was chosen as the maximizer over $x \leq b$ of the right hand side of the previous display. Indeed, the maximizer is $-x^\star(z)$ with $x^\star(z) = -\frac{\varphi-1-\sqrt{(\varphi-1)^2+2\varphi z}}{\varphi} \geq 0$. Now we turn to proving the second claim, which will follow from a series of rewritings of the first claim. The previous display can be rewritten as

$$g(z) = -x^\star(z) - \ln(1 - \varphi x^\star(z)).$$

Let $s' = x^\star(z)$ so that $z = \frac{1}{2}\varphi s'^2 + (\varphi - 1)s'$. If we let $h(u) = -u - \ln(1 - u)$, the previous display can be rewritten as

$$g(z) = (\varphi - 1)s' + h(\varphi s').$$

In these terms, the first claim that we already proved takes the shape

$$\mathrm{e}^{x-(\varphi-1)s'-h(\varphi s')} - 1 - x \leq \frac{1}{2}\varphi x^2 - (\varphi - 1)s' - \frac{1}{2}\varphi s'^2$$

any time that $s' \leq 1/\varphi$. Define $s = (\varphi - 1)s'$. Replace this in the last display and bound the last, negative term by 0 to obtain that, as long as $s \leq \frac{\varphi-1}{\varphi}$,

$$\mathrm{e}^{x-s-h(cs)} - 1 - x \leq \frac{1}{2}\varphi x^2 - s.$$

This is our claim. The additional bound on h is well known and can be proven with a term-wise bound on the Taylor expansion of $u \mapsto -u - \ln(1 - u)$. □

F.5 Dual formulation of $\Delta\Phi$

Recall from the definitions in Section 2 that the Bregman divergence $D_{\boldsymbol{\eta}}(\boldsymbol{p}, \boldsymbol{u})$ between $\boldsymbol{p}$ and $\boldsymbol{u}$, two vectors in $\mathbb{R}_+^K$, was defined in (3) as

$$D_{\boldsymbol{\eta}}(\boldsymbol{p}, \boldsymbol{u}) = \sum_{k \in K} p_k \left(\frac{\ln(p_k/u_k) - (p_k - u_k)}{\eta_k}\right);$$

and the corresponding potential Φ, in (4) as

$$\Phi(\boldsymbol{X}, \boldsymbol{\eta}) = \sup_{\boldsymbol{p} \in \mathcal{P}(K)} \langle \boldsymbol{p}, \boldsymbol{X} \rangle - D_{\boldsymbol{\eta}}(\boldsymbol{p}, \boldsymbol{u}).$$

In the implementation of the algorithm, we rely on the dual formulation of the potential Φ and its change $\Delta\Phi$ between rounds. We compute these in the following two lemmas.

Lemma F.6 (Potential difference in dual form)**.** Let $\boldsymbol{X}, \Delta\boldsymbol{X} \in \mathbb{R}^K$ and $\boldsymbol{u}, \boldsymbol{\eta} \in \mathbb{R}_+^K$, $\Delta\Phi = \Phi(\boldsymbol{X} + \Delta\boldsymbol{X}, \boldsymbol{\eta}) - \Phi(\boldsymbol{X}, \boldsymbol{\eta})$, and $\boldsymbol{w} = \arg\max_{\boldsymbol{p} \in \mathcal{P}(K)} \langle \boldsymbol{p}, \boldsymbol{X} \rangle - D_{\boldsymbol{\eta}}(\boldsymbol{p}, \boldsymbol{u})$. Then

$$\Delta\Phi = \inf_{\Delta a \in \mathbb{R}} \sum_{k \in K} w_k \left(\frac{\mathrm{e}^{\eta_k(\Delta X_k - \Delta a)} + \eta_k \Delta a - 1}{\eta_k}\right).$$

Proof. From Lemma F.7 we know that

$$w_k = u_k \mathrm{e}^{\eta_k(X_k - a^*)},$$

where a^* is such that $\sum_{k \in K} w_k = 1$, and that

$$\Phi(\boldsymbol{X}, \boldsymbol{\eta}) = a^* + \sum_{k \in K} u_k \left(\frac{\mathrm{e}^{\eta_k(X_k - a^*)} - 1}{\eta_k}\right).$$

Use the same lemma and the change of variable $a = a^* + \Delta a$ to obtain that

$$\Phi(\boldsymbol{X} + \Delta\boldsymbol{X}, \boldsymbol{\eta}) = \inf_{\Delta a \in \mathbb{R}} \left\{ a^* + \Delta a + \sum_{k \in K} u_k \left(\frac{\mathrm{e}^{\eta_k(X_k - a^* + \Delta X_k - \Delta a)} - 1}{\eta_k}\right) \right\}.$$

Substract these two displays and use the explicit expresion for w. In this way, we obtain the result. □

Lemma F.7 (Potential Dual)**.** Let $\boldsymbol{X} \in \mathbb{R}^K$ be a vector, and let $\boldsymbol{u}, \boldsymbol{\eta} \in \mathbb{R}^K_+$ be positive vectors. Then

1. The potential Φ satisfies
$$\Phi(\boldsymbol{X}, \boldsymbol{\eta}) = \langle \boldsymbol{p}^*, \boldsymbol{X} \rangle - D_{\boldsymbol{\eta}}(\boldsymbol{p}^*, \boldsymbol{u}),$$
where $p^*_k = u_k \mathrm{e}^{\eta_k (X_k - a^*)}$, and a^* is such that $\sum_{k \in K} p^*_k = 1$.
2. The potential Φ satisfies the identity
$$\Phi(\boldsymbol{X}, \boldsymbol{\eta}) = \inf_{a \in \mathbb{R}} \left\{ a + \sum_{k \in K} u_k \left(\frac{\mathrm{e}^{\eta (X_k - a)} - 1}{\eta_k} \right) \right\}.$$

Proof. Consider the optimization problem
$$\sup_{p \in \mathcal{P}(K)} \langle \boldsymbol{p}, \boldsymbol{X} \rangle - D_{\boldsymbol{\eta}}(\boldsymbol{p}, \boldsymbol{u}).$$
Its Lagrangian function is
$$\mathcal{L}(a, \boldsymbol{p}) = \langle \boldsymbol{p}, \boldsymbol{X} \rangle - D_{\boldsymbol{\eta}}(\boldsymbol{p}, \boldsymbol{u}) - a \left(\sum_{k \in K} p_k - 1 \right).$$
The strong duality relation
$$\sup_{\boldsymbol{p} \in \mathcal{P}(K)} \langle \boldsymbol{p}, \boldsymbol{X} \rangle + D_{\boldsymbol{\eta}}(\boldsymbol{p}, \boldsymbol{u}) = \inf_{a \in \mathbb{R}} \sup_{\boldsymbol{p} \in \mathbb{R}^K} \mathcal{L}(a, \boldsymbol{p}) \tag{43}$$
holds, and the maximum on the right hand side can be computed by diferentiation. The gradient with respect to $\boldsymbol{p}$ is
$$\nabla_{\boldsymbol{p}} \mathcal{L}_k = X_k - a - \frac{\ln(p_k / u_k)}{\eta_k},$$
which is zero at
$$p^*_k = u_k \mathrm{e}^{\eta_k (X_k - a)}.$$
Replace $\boldsymbol{p}^*$ in the Lagrangian $\mathcal{L}$ to conclude that
$$\mathcal{L}(a, \boldsymbol{p}^*) = a + \sum_{k \in K} u_k \left(\frac{\mathrm{e}^{\eta_k (X_k - a)} - 1}{\eta_k} \right).$$
Replace this in (43) to obtain the second claim. For the first claim, differentiate $\inf_{a \in \mathbb{R}} \mathcal{L}(a, p^*)$ with respect to a and equate to 0. □

G Proof of Theorem 1.2

Recall that Δv_t is implicitly specified in the definition of MUSCADA, in Figure 1. The main intuition driving the result contained in Theorem 1.2 stems from a Taylor approximation of the increment of the potential function at round t for small learning rates. The duality computation for the potential increment $\Delta\Phi$ of Lemma F.6 implies that, at round t, Δv_t is the value of Δv that satisfies
$$\inf_{\lambda \in \mathbb{R}} \sum_{k \in K} w_{t,k} \left(\frac{\mathrm{e}^{-\eta_{t-1,k}(\ell_{t,k} - \lambda) - \eta^2_{t-1,k} \sigma^2_k \Delta v} + \eta_{t-1,k}(\ell_{t,k} - \lambda) - 1}{\eta_{t-1,k}} \right) = 0, \tag{44}$$
where, in the notation of Lemma F.6, we used $\Delta \boldsymbol{X} = \Delta \boldsymbol{R}_t$ and reparametrized by $\lambda = \langle \boldsymbol{w}_t, \boldsymbol{\ell}_t \rangle - \Delta a$. For small values of η, the Taylor approximation $\mathrm{e}^{\eta x - \eta^2 b} = 1 + \eta x + \frac{1}{2}\eta^2 (x^2 - 2b) + O(\eta^3)$ gives that, if all the learning rates are small, the quantity being minimized in the previous display can be approximated as
$$\begin{aligned} \sum_{k \in K} w_{t,k} \left(\frac{\mathrm{e}^{-\eta_{t-1,k}(\ell_{t,k} - \lambda) - \eta^2_{t-1,k} \sigma^2_k \Delta v} + \eta_{t-1,k}(\ell_{t,k} - \lambda) - 1}{\eta_{t-1,k}} \right) \approx \\ \frac{1}{2} \sum_{k \in K} w_{t,k} \eta_{t-1,k} (\ell_{t,k} - \lambda)^2 - \Delta v \sum_{k \in K} w_{t,k} \eta_{t-1,k} \sigma^2_k. \end{aligned} \tag{45}$$

If this approximate expression could be plugged into (44), we could solve the infimum and obtain that

$$\Delta v_t \approx \frac{1}{2}\frac{\mathrm{var}_{\tilde{\boldsymbol{w}}}(\boldsymbol{\ell}_t)}{\langle \tilde{\boldsymbol{w}}, \boldsymbol{\sigma}^2\rangle}$$

with $\tilde{w}_{t,k} \propto w_{t,k}\eta_{t-1,k}$. However, this approximation is only valid under range restrictions in the values of λ. This is the subject of Lemma G.2, whose main technical ingredient is the inequality obtained in Lemma F.5, which contains an estimate that makes (45) precise. We gather theses results in the following proposition. Used with $b = 1$, it implies Theorem 1.2 because the learning rates from Figure 2 are all smaller than $1/(2\sigma_{\max})$.

Proposition G.1. *Fix $t \geq 1$. Let $\tilde{w}_{t,k} \propto w_{t,k}\eta_{t-1,k}$, where $\boldsymbol{w}_t$ are the weights played by* MUSCADA *at round t, and $\boldsymbol{\eta}_{t-1}$ its learning rates. The following statements hold.*

1. *If $\max_k 2\eta_{t-1,k}\sigma_k \leq b$ and $b \leq 1$, then*

$$\Delta v_t \leq c_0 \frac{\langle \tilde{\boldsymbol{w}}_t, \boldsymbol{\ell}_t^2\rangle}{\langle \tilde{\boldsymbol{w}}_t, \boldsymbol{\sigma}^2\rangle} \leq c_0, \tag{46}$$

   where the constant c_0 satisfies $c_0 \leq 3.1$ and depends only on b.

2. *If $\max_k 2\eta_{t-1,k}\sigma_{\max} \leq b$ for some $b \leq 1$, and*

$$\Delta s_t = \frac{\mathrm{var}_{\tilde{\boldsymbol{w}}_t}(\boldsymbol{\ell}_t)}{\langle \tilde{\boldsymbol{w}}_t, \boldsymbol{\sigma}^2\rangle},$$

   then

$$\Delta v_t \leq c_1 \Delta s_t + c_2 \Delta s_t^2, \tag{47}$$

   and consequently

$$v_t \leq c_3 s_t,$$

   where $c_1 \leq 0.72$, $c_2 \leq 2.4$, and $c_3 = c_1 + c_2 \leq 3.1$ depend on b only.

Proof of Proposition G.1. First, we prove 1. Assume that $\max_k 2\eta_{t-1,k}\sigma_k \leq b'$ and that $b' \leq 1$. Our objective is to use Lemma G.2 with $\lambda = 0$. To this end, let $\varphi' = \frac{\mathrm{e}^{b'} - b' - 1}{\frac{1}{2}b'^2} \geq 1$, $c_1' = \frac{b'^2\varphi'^2}{8(\varphi'-1)}$, and $c_2' = \frac{\varphi'^4 b'^2}{8(\varphi'-1)^2} - \frac{\varphi'^3 b'^2}{8(\varphi'-1)}$ be as in Lemma G.2. Since we assumed that $b \leq 1$, we have that $c_1' \leq 1/2$, and we can conclude that

$$\Delta v_t \leq \frac{\varphi'}{2}\Delta s_{t,0} + \frac{1}{2}\frac{c_2'\Delta s_{t,0}^2}{1 - c_1'\Delta s_{t,0}}$$

with $\Delta s_{t,0} = \frac{\langle \tilde{\boldsymbol{w}}_t, \boldsymbol{\ell}_t^2\rangle}{\langle \tilde{\boldsymbol{w}}_t, \boldsymbol{\sigma}^2\rangle} \leq 1$. Use this to conclude that

$$\Delta v_t \leq \frac{\varphi'}{2} + \frac{1}{2}\frac{c_2'}{1 - c_1'}.$$

This last display is exactly our first claim once we set $c_0 = \frac{\varphi'}{2} + \frac{1}{2}\frac{c_2'}{1-c_1'}$. The value of c_0' depends monotonically on that of b'. Compute the value of c_0' for $b' = 1$ to confirm that $c_0' \leq 3.1$.

We now turn our attention to the second claim. We proceed in a similar fashion as before. Assume that $\max_k 2\eta_{t-1,k}\sigma_{\max} \leq b$ for some $b \geq 1$. Let φ, c_1, c_2 be defined as before but now in terms of b. Use Lemma G.2 to obtain that

$$\Delta v_t \leq \frac{\varphi}{2}\Delta s_t + \frac{1}{2}\frac{c_2\Delta s_t^2}{1 - c_1\Delta s_t}$$

with $\Delta s_t = \frac{\mathrm{var}_{\tilde{\boldsymbol{w}}_t}(\boldsymbol{\ell}_t)}{\langle \tilde{\boldsymbol{w}}_t, \boldsymbol{\sigma}^2\rangle} \leq 1$. Use this to conclude that

$$\Delta v_t \leq \frac{\varphi}{2}\Delta s_t + \frac{1}{2}\frac{c_2}{1 - c_1}\Delta s_t^2.$$

This is exactly the second claim up to a redefinition of constants. The "consequenlty" part of the claim follows from the observation that $\Delta s_t^2 \leq \Delta s_t$ and a summation over time. The computation of the upper bound on the constants is similar as before. □

Lemma G.2. Let $t \geq 1$, $\lambda \in \mathbb{R}$, and let

$$\Delta s_t = \Delta s_t(\lambda) = \frac{\langle \tilde{\boldsymbol{w}}_t, (\boldsymbol{\ell}_t - \lambda)^2 \rangle}{\langle \tilde{\boldsymbol{w}}_t, \boldsymbol{\sigma}^2 \rangle} \tag{48}$$

with $\tilde{w}_{t,k} \propto w_{t,k}\eta_{t-1,k}$. Then, whenever $\max_k \eta_{t-1,k}(\ell_{k,t} - \lambda) \leq b$ and $\max_k(2\eta_{t-1,k}\sigma_k) \leq b$ for some $b \geq 0$, we have that

$$\Delta v_t \leq \frac{\varphi}{2}\Delta s_t + c_1 \Delta v_t \Delta s_t + \frac{1}{2}c_2 \Delta s_t^2, \tag{49}$$

where $\varphi = \frac{\mathrm{e}^b - b - 1}{\frac{1}{2}b^2} \geq 1$, $c_1 = \frac{b^2\varphi^2}{8(\varphi-1)}$, and $c_2 = \frac{\varphi^4 b^2}{8(\varphi-1)^2} - \frac{\varphi^3 b^2}{8(\varphi-1)}$. If additionally $c_1 \Delta v_t < 1$, then

$$\Delta v_t \leq \frac{\varphi}{2}\Delta s_t + \frac{1}{2}\frac{c_2 \Delta s^2}{1 - c_1 \Delta s_t}. \tag{50}$$

Proof. Let $t \geq 1$. First note that if $c_1 \Delta s_t \geq 1$, our claim becomes trivial. We can safely assume that that $c_1 \Delta s_t < 1$. We proceed in the following steps. Use Lemma F.6 to express the increase in the potential function $\Delta\Phi_t(\Delta v) = \Phi(\boldsymbol{R}_t - \boldsymbol{\mu}_{t-1} - \boldsymbol{\eta}\boldsymbol{\sigma}^2\Delta v, \boldsymbol{\eta}_{t-1}) - \Phi(\boldsymbol{R}_t - \boldsymbol{\mu}_t, \boldsymbol{\eta}_{t-1})$ in dual form as

$$\Delta\Phi_t(\Delta v) = \inf_{\lambda \in \mathbb{R}} \sum_{k \in K} w_{t,k} \left(\frac{\mathrm{e}^{-\eta_{t-1,k}(\ell_{t,k} - \lambda) - \eta_{t-1,k}^2 \sigma_k^2 \Delta v} + \eta_{t-1,k}(\ell_{t,k} - \lambda) - 1}{\eta_{t-1,k}} \right).$$

From now and until the end of the proof, omit the time indexes for readability.

Because of our assumption that $\eta_k|\ell_k - \lambda| \leq b$, Lemma F.5 can be used to obtain that

$$\Delta\Phi(\Delta v) \leq \frac{1}{2}\varphi \sum_{k \in K} w_k[\eta_k(\ell_k - \lambda)^2] - \sum_{k \in K} w_k \left(\frac{g^{-1}(\eta_k^2 \sigma_k^2 \Delta v)}{\eta_k} \right)$$

where $g(x) = x + h(cx)$, $h(u) = \frac{1}{2}\frac{u^2}{1-u}$ and $c = \varphi/(\varphi - 1)$. Use the concavity of $x \mapsto g^{-1}(\Delta v x)/x$ and Jensen's inequality to deduce that

$$\begin{aligned} \sum_{k \in K} w_k \left(\frac{g^{-1}(\eta_k^2 \sigma_k^2 \Delta v)}{\eta_k} \right) &= \sum_{k \in K} w_k \left(\eta_k \sigma_k^2 \frac{g^{-1}(\eta_k^2 \sigma_k^2 \Delta v)}{\eta_k^2 \sigma_k^2} \right) \\ &\geq \langle \boldsymbol{w}, \boldsymbol{\eta}\boldsymbol{\sigma}^2 \rangle \frac{g^{-1}\left(\Delta v \langle \hat{\boldsymbol{w}}, \boldsymbol{\eta}^2\boldsymbol{\sigma}^2 \rangle\right)}{\langle \hat{\boldsymbol{w}}, \boldsymbol{\eta}^2\boldsymbol{\sigma}^2 \rangle}, \end{aligned}$$

where we defined $\hat{w}_k \propto w_k \eta_k \sigma_k^2$. This is useful for obtaining the bound

$$\Delta\Phi(\Delta v) \leq \frac{1}{2}\varphi \sum_{k \in K} w_k \left(\eta_k(\ell_k - \lambda)^2 \right) - \langle \boldsymbol{w}, \boldsymbol{\eta}\boldsymbol{\sigma}^2 \rangle \frac{g^{-1}\left(\Delta v \langle \hat{\boldsymbol{w}}, \boldsymbol{\eta}^2\boldsymbol{\sigma}^2 \rangle\right)}{\langle \hat{\boldsymbol{w}}, \boldsymbol{\eta}^2\boldsymbol{\sigma}^2 \rangle}.$$

Consequenlty, $\Delta\Phi(\Delta v^\star) \leq 0$ for

$$\Delta v^\star = \frac{1}{\langle \hat{\boldsymbol{w}}, \boldsymbol{\eta}^2\boldsymbol{\sigma}^2 \rangle} g \left(\frac{1}{2}\varphi \langle \hat{\boldsymbol{w}}, \boldsymbol{\eta}^2\boldsymbol{\sigma}^2 \rangle \frac{\langle \tilde{\boldsymbol{w}}, (\boldsymbol{\ell} - \lambda)^2 \rangle}{\langle \tilde{\boldsymbol{w}}, \boldsymbol{\sigma}^2 \rangle} \right),$$

where $\tilde{w}_k \propto w_k \eta_k$. Use the definition of Δv and the continuity of $\Delta\Phi$ to conclude that $\Delta v \leq \Delta v^\star$. Unpack the definition of g to obtain that

$$\Delta v \leq \frac{1}{2}\varphi \Delta s + \frac{1}{2}\frac{\langle \hat{\boldsymbol{w}}, \boldsymbol{\eta}^2\boldsymbol{\sigma}^2 \rangle (c' \Delta s)^2}{1 - c' \langle \hat{\boldsymbol{w}}, \boldsymbol{\eta}^2\boldsymbol{\sigma}^2 \rangle \Delta s}$$

with $c' = \frac{1}{2}\frac{\varphi^2}{\varphi - 1}$. Next, we will use use that $\langle \hat{\boldsymbol{w}}, \boldsymbol{\eta}^2\boldsymbol{\sigma}^2 \rangle \leq \frac{1}{4}b^2$ to bound further Δv. Use this observation and the definition of c_1 to deduce the inequality $\langle \hat{\boldsymbol{w}}, \boldsymbol{\eta}^2\boldsymbol{\sigma}^2 \rangle c' \Delta s \leq c_1 \Delta s < 1$. Plug this in the previous display and rearrange to obtain the result:

$$\Delta v \leq \frac{1}{2}\varphi \Delta s + \frac{1}{2}\Delta s^2 \left(\frac{1}{4}c'^2 b^2 - \frac{1}{4}\varphi c' b^2 \right) + \frac{1}{4}c' b^2 \Delta v \Delta s,$$

exactly what we claimed. □