Proceedings of Machine Learning Research vol 134:1–44, 2021 34th Annual Conference on Learning Theory

Impossible Tuning Made Possible: A New Expert Algorithm and Its Applications

Liyu Chen* LIYUC@USC.EDU
University of Southern California

Haipeng Luo* HAIPENGL@USC.EDU
University of Southern California

Chen-Yu Wei* CHENYU.WEI@USC.EDU
University of Southern California

Editors: Mikhail Belkin and Samory Kpotufe

Abstract

We resolve the long-standing "impossible tuning" issue for the classic expert problem and show that, it is in fact possible to achieve regret $\tilde{\mathcal{O}}\left(\sqrt{(\ln d)\sum_{t=1}^{T}\ell_{t,i}^2}\right)$ simultaneously for all expert i in a T-round d-expert problem where $\ell_{t,i}$ is the loss for expert i in round t. Our algorithm is based on the Mirror Descent framework with a correction term and a weighted entropy regularizer. While natural, the algorithm has not been studied before and requires a careful analysis. We also generalize the bound to $\tilde{\mathcal{O}}\left(\sqrt{(\ln d)\sum_{t=1}^{T}(\ell_{t,i}-m_{t,i})^2}\right)$ for any prediction vector m_t that the learner receives, and recover or improve many existing results by choosing different m_t. Furthermore, we use the same framework to create a master algorithm that combines a set of base algorithms and learns the best one with little overhead. The new guarantee of our master allows us to derive many new results for both the expert problem and more generally Online Linear Optimization.

1. Introduction

In the classic expert problem (Freund and Schapire, 1997), a learner interacts with an adversary for T rounds, where in each round t, the learner first decides a distribution $w_t \in \Delta_d$ over a fixed set of d experts, and then the adversary decides a loss vector $\ell_t \in \mathbb{R}^d$. The learner suffers loss $\langle w_t, \ell_t\rangle$ and observes ℓ_t at the end of round t. The regret against a fixed strategy $u \in \Delta_d$ is defined as $\textsc{Reg}(u) = \sum_{t=1}^{T}\langle w_t - u, \ell_t\rangle$. Many existing algorithms achieve $\max_u \textsc{Reg}(u) = \max_i \textsc{Reg}(e_i) = \mathcal{O}(\sqrt{T\ln d})$, which is known to be minimax optimal.

In particular, both the $\textsc{Prod}$ algorithm (Cesa-Bianchi et al., 2007), which sets $w_{t+1,i} \propto w_{t,i}(1-\eta\ell_{t,i})$, and a variant of the classic multiplicative-weight (Steinhardt and Liang, 2014), which sets $w_{t+1,i} \propto w_{t,i}e^{-\eta\ell_{t,i}-\eta^2\ell_{t,i}^2}$, achieve a regret bound $\textsc{Reg}(e_i) \le \frac{\ln d}{\eta} + \eta\sum_{t=1}^{T}\ell_{t,i}^2$ for some learning rate η. With the optimal tuning of η, this gives an adaptive bound $\textsc{Reg}(e_i) = \mathcal{O}\left(\sqrt{(\ln d)\sum_{t=1}^{T}\ell_{t,i}^2}\right)$, potentially much better than the minimax bound. However, since different expert i requires a different tuning, no method is known to achieve this bound *simultaneously for all* i. Several works discuss the difficulty of doing so even with different η for different experts and why all standard

* Authors are listed in alphabetical order.

Table 1: Summary of main results. $w_t \in \mathbb{R}^d$ is the decision of the learner, ℓ_t is the loss vector, m_t is a prediction for ℓ_t, $\mathcal{L}_T = \sum_{t=1}^T (\ell_t - m_t)(\ell_t - m_t)^\top$, and r is the rank of $\mathcal{L}_T$.

	Results	Notes
Expert $\textsc{Reg}(e_i)$	$\tilde{\mathcal{O}}\left(\sqrt{(\ln d)\sum_{t=1}^T (\ell_{t,i} - m_{t,i})^2}\right)$ With different m_t, $(\ell_{t,i} - m_{t,i})^2$ becomes: • $\ell_{t,i}^2$ • $(\ell_{t,i} - \frac{1}{T}\sum_{s=1}^T \ell_{s,i})^2$ • $(\ell_{t,i} - \ell_{t-1,i})^2$ • $(\ell_{t,i} - \ell_{t,1})^2$ • $\langle w_t - e_i, \ell_t \rangle^2$ • $\langle w_t - e_i, \ell_t - m_t \rangle^2$	• $\ln d$ can be generalized to $\mathrm{KL}(u, \pi)$ for competitor u and prior π • all results generalize to switching regret and unknown loss range • analogue for interval regret or bandits is impossible
OLO $\textsc{Reg}(u)$	$\tilde{\mathcal{O}}\left(\sqrt{r\sum_{t=1}^T \langle u, \ell_t - m_t \rangle^2}\right)$ $\tilde{\mathcal{O}}\left(\lVert u \rVert \sqrt{\sum_{t=1}^T \lVert \ell_t - m_t \rVert_\star^2}\right)$ $\tilde{\mathcal{O}}\left(\sqrt{\left(\lVert u \rVert_2^2 + u^\top \mathcal{L}_T^{1/2} u\right)\mathrm{tr}\left(\mathcal{L}_T^{1/2}\right)}\right)$ $\tilde{\mathcal{O}}\left(\sqrt{r\sum_{t=1}^T \langle u - w_t, \ell_t - m_t \rangle^2}\right)$	• first three bounds hold simultaneously • all results generalize to unconstrained learning and unknown Lipschitzness

tuning techniques fail (Cesa-Bianchi et al., 2007; Hazan and Kale, 2010). Indeed, the problem is so challenging that it has been referred to as the "impossible tuning" issue (Gaillard et al., 2014).

Our first main contribution is to show that, perhaps surprisingly, this impossible tuning is in fact *possible* (up to an additional $\ln T$ factor), via an algorithm combining ideas that mostly appear before already. More concretely, we achieve this via Mirror Descent with a correction term similar to (Steinhardt and Liang, 2014) and a weighted negative entropy regularizer with different learning rates for each expert (and each round) similar to (Bubeck et al., 2017). Note that while natural, this algorithm has not been studied before,[1] and is *not equivalent* to using different learning rates for different experts in PROD or multiplicative-weight, as it does not admit a closed "proportional" form (and instead needs to be computed via a line search). Crucially, our analysis carefully utilizes a negative term in the regret bound to achieve the claimed result.

We present our result in a more general setting where the learner receives a predicted loss vector m_t before deciding w_t (Rakhlin and Sridharan, 2013b), and show a bound $\textsc{Reg}(e_i) = \tilde{\mathcal{O}}\left(\sqrt{(\ln d)\sum_{t=1}^T (\ell_{t,i} - m_{t,i})^2}\right)$ simultaneously for all i (setting $m_t = 0$ resolves the original impossible tuning issue). Using different m_t, we achieve various regret bounds summarized in Table 1, which either recover the guarantees of existing algorithms such as $(\mathcal{A}, \mathcal{B})$-PROD (Sani et al., 2014), ADAPT-ML-PROD (Gaillard et al., 2014), OPTIMISTIC-ADAPT-ML-PROD (Wei et al., 2016), or improve over existing variance/path-length bounds in (Steinhardt and Liang, 2014). We also show that the bound $\tilde{\mathcal{O}}\left(\sqrt{(\ln d)\sum_{t=1}^T \langle w_t - e_i, \ell_t - \ell_{t-1} \rangle^2}\right)$, obtained by (Wei et al., 2016) and our work, simultaneously ensures the "fast rate" consequences discussed in (Koolen et al., 2016) for stochastic settings and the path-length bound useful for fast convergence in games (Syrgkanis et al., 2015). See Section 2.1 for detailed discussions.

1. Except that a simpler version is used in a concurrent work (Chen et al., 2021) by the same authors for a different problem (learning stochastic shortest path).

Our second main contribution is to use the same algorithmic framework to create a master algorithm that combines a set of base algorithms and learns the best for different environments (Section 2.2). Although similar ideas appear in many prior works with different masters (Koolen et al., 2014; van Erven and Koolen, 2016; Foster et al., 2017; Cutkosky, 2019b; Bhaskara et al., 2020), the new guarantee of our master allows us to derive many new results that cannot be achieved before, for both the expert problem and more generally Online Linear Optimization (OLO).

Specifically, for the expert problem, using the master to combine different instances of *itself*, we further generalize the aforementioned bound from different aspects, including replacing the $\ln d$ factor with $\mathrm{KL}(u, \pi)$ when competing against u with a prior distribution π, adapting to the scale of each expert, extending the results to switching regret, and dealing with unknown loss range. These results improve over (Luo and Schapire, 2015; Koolen and Van Erven, 2015), (Bubeck et al., 2017; Foster et al., 2017; Cutkosky and Orabona, 2018), (Cesa-Bianchi et al., 2012), and (Mhammedi et al., 2019) respectively. See Section 3 for detailed discussions.

Next, we consider the more general OLO problem where the learner's decision set generalizes from Δ_d to an arbitrary closed convex set $\mathcal{K} \subset \mathbb{R}^d$ (other than this change, the learning protocol and the regret definition remain the same). Using our master to combine different types of base algorithms, we achieve four different and incomparable bounds on $\textsc{Reg}(u)$ simultaneously for all u, listed in Table 1. Importantly, the first three bounds can be achieved at the same time *with one single algorithm*. These bounds improves over a line of recent advances in OLO (van Erven and Koolen, 2016; Cutkosky and Orabona, 2018; Cutkosky, 2019a,b; Mhammedi et al., 2019; Mhammedi and Koolen, 2020; Cutkosky, 2020). See Section 4 for detailed discussions.

Notation Throughout the paper, Δ_d denotes the $d-1$ dimensional simplex; $e_i, \mathbf{0}, \mathbf{1} \in \mathbb{R}^d$ are respectively the i-th standard basis vector, the all-zero vector, and the all-one vector; $[n]$ denotes the set $\{1, \ldots, n\}$; $\mathrm{KL}(\cdot, \cdot)$ denotes the KL divergence; $\|u\|_A = \sqrt{u^\top A u}$ is the quadratic norm with respect to a matrix A; $D_\psi(u, w) = \psi(u) - \psi(w) - \langle \nabla\psi(w), u - w\rangle$ is the Bregman divergence of u and w with respect to a convex function ψ, and $\tilde{\mathcal{O}}(\cdot)$ hides logarithmic dependence on T.

2. An Algorithmic Framework

Consider the expert problem and recall that the learner sequentially decides a distribution $w_t \in \Delta_d$ (with the help of a prediction $m_t \in \mathbb{R}^d$) and then observes the loss vector $\ell_t \in \mathbb{R}^d$. Note that we do not make the typical assumption $\ell_{t,i} \in [0, 1]$ or $|\ell_{t,i}| \leq 1$; instead, the requirement (if any) on the range of the losses will be stated either explicitly or implicitly in the conditions of each lemma or theorem.

We start by proposing a general algorithmic framework called Multi-scale Multiplicative-weight with Correction (MsMwC), shown in Algorithm 1. In Section 2.1, we instantiate the framework in a specific way to resolve the impossible tuning issue, and in Section 2.2, we instantiate it differently to obtain a new master algorithm, with more applications discussed in following sections.

MsMwC is a variation of the standard Optimistic-Mirror-Descent (OMD) framework, which maintains two sequences $w_1, \ldots, w_T$ and $w'_1, \ldots, w'_T$ updated according to Line 3 and Line 5. The key new ingredients are the following. First, we adopt a time-varying decision subset $\Omega_t \subseteq \Delta_d$ to which w_t and w'_{t+1} belong. This is decided at the beginning of each round t and is useful for applications discussed in Section 3.4 and Appendix D.5, where we need to eliminate some experts on-the-fly. (For other applications, Ω_t is either Δ_d or its truncated version throughout all T rounds.)

Algorithm 1 Multi-scale Multiplicative-weight with Correction (MSMWC)

Initialize: $w'_1 \in \Delta_d$.

for $t = 1, \ldots, T$ **do**

1 Receive prediction $m_t \in \mathbb{R}^d$.

2 Decide a compact convex decision subset $\Omega_t \subseteq \Delta_d$ and learning rates $\eta_t \in \mathbb{R}^d_{\geq 0}$.

3 Compute $w_t = \operatorname{argmin}_{w \in \Omega_t} \langle w, m_t \rangle + D_{\psi_t}(w, w'_t)$ where $\psi_t(w) = \sum_{i=1}^d \frac{1}{\eta_{t,i}} w_i \ln w_i$.

4 Play w_t, receive ℓ_t, and construct correction term $a_t \in \mathbb{R}^d$ with $a_{t,i} = 32\eta_{t,i}(\ell_{t,i} - m_{t,i})^2$.

5 Compute $w'_{t+1} = \operatorname{argmin}_{w \in \Omega_t} \langle w, \ell_t + a_t \rangle + D_{\psi_t}(w, w'_t)$.

Second, our regularizer $\psi_t(w) = \sum_{i=1}^d \frac{1}{\eta_{t,i}} w_i \ln w_i$ is negative entropy with individual and time-varying learning rate $\eta_{t,i}$ for each expert i. For most applications, $\eta_{t,i}$ is the same for all t, in which case our regularizer is the same as that used in the MSMW algorithm of (Bubeck et al., 2017).

Finally, we adopt a second-order correction term a_t added to the loss vector ℓ_t in the update of w'_{t+1} (Line 5), which is the most important difference compared to MSMW (Bubeck et al., 2017). Similar correction terms have been used in prior works such as (Hazan and Kale, 2010; Steinhardt and Liang, 2014; Wei and Luo, 2018) and are known to be important to achieving a regret bound that depends on quantities only related to the expert being compared to.

One can see that essentially all ingredients of MSMWC appear before in the literature. However, the specific combination of these ingredients (which has not been studied before) and a careful analysis enable us to resolve the impossible tuning issue as well as developing other new results.

We present a general lemma on the regret guarantee of MSMWC below, which holds under a condition on the magnitude of $\eta_{t,i}|\ell_{t,i} - m_{t,i}|$; see Appendix B for the proof. We also note that the last negative term in the regret bound is particularly important for some of the applications.

Lemma 1 *Define $f_{\mathrm{KL}}(a, b) = a \ln \frac{a}{b} - a + b$ for $a, b \in [0, 1]$.*[2] *Suppose that for all $t \in [T]$, $32\eta_{t,i}|\ell_{t,i} - m_{t,i}| \leq 1$ holds for all i such that $w_{t,i} > 0$. Then MSMWC ensures for any $u \in \bigcap_{t=1}^T \Omega_t$,*

$$\begin{aligned} \mathrm{REG}(u) \leq \sum_{i=1}^d \frac{1}{\eta_{1,i}} f_{\mathrm{KL}}(u_i, w'_{1,i}) + \sum_{t=2}^T \sum_{i=1}^d \left(\frac{1}{\eta_{t,i}} - \frac{1}{\eta_{t-1,i}} \right) f_{\mathrm{KL}}(u_i, w'_{t,i}) \\ + 32 \sum_{t=1}^T \sum_{i=1}^d \eta_{t,i} u_i (\ell_{t,i} - m_{t,i})^2 - 16 \sum_{t=1}^T \sum_{i=1}^d \eta_{t,i} w_{t,i} (\ell_{t,i} - m_{t,i})^2. \end{aligned} \tag{1}$$

2.1. Impossible tuning made possible

To resolve the impossible tuning issue, we instantiate MSMWC in the following way with the decision sets fixed to a truncated simplex and the learning rates tuned using data observed so far.

Theorem 2 *Suppose $|\ell_{t,i}|$ and $|m_{t,i}|$ are bounded by 1 for all $t \in [T]$ and $i \in [d]$. Then MSMWC with $w'_1 = \frac{1}{d}\mathbf{1}$, $\Omega_1 = \cdots = \Omega_T = \{w \in \Delta_d : w_i \geq \frac{1}{dT}\}$, and $\eta_{t,i} = \min\left\{ \sqrt{\frac{\ln(dT)}{\sum_{s<t}(\ell_{s,i} - m_{s,i})^2}}, \frac{1}{64} \right\}$ ensures for all $i_\star \in [d]$, $\mathrm{REG}(e_{i_\star}) = \mathcal{O}\left(\ln(dT) + \sqrt{\ln(dT) \sum_{t=1}^T (\ell_{t,i_\star} - m_{t,i_\star})^2} \right)$.*

2. Define $f_{\mathrm{KL}}(0, b) = b$ for all $b \in [0, 1]$.

Proof [sketch] We apply Eq. (1) with $u = (1 - \frac{1}{T})e_{i_\star} + \frac{1}{T}w'_1 \in \bigcap_{t=1}^{T} \Omega_t$, so that $\textsc{Reg}(e_{i_\star}) \leq \textsc{Reg}(u) + 2$. Most calculation is straightforward, and the most important part is to realize that $\left(\frac{1}{\eta_{t,i}} - \frac{1}{\eta_{t-1,i}}\right) w'_{t,i}$, a term from $\left(\frac{1}{\eta_{t,i}} - \frac{1}{\eta_{t-1,i}}\right) f_{\mathrm{KL}}(u_i, w'_{t,i})$, can be bounded as:

$$\left(\frac{1}{\eta_{t,i}} - \frac{1}{\eta_{t-1,i}}\right) w'_{t,i} = \frac{\frac{1}{\eta_{t,i}^2} - \frac{1}{\eta_{t-1,i}^2}}{\frac{1}{\eta_{t,i}} + \frac{1}{\eta_{t-1,i}}} w'_{t,i} \leq \eta_{t-1,i} w'_{t,i} \left(\frac{1}{\eta_{t,i}^2} - \frac{1}{\eta_{t-1,i}^2}\right),$$

which is further bounded by $\frac{1}{\ln(dT)} \eta_{t-1,i} w'_{t,i} (\ell_{t-1,i} - m_{t-1,i})^2$ using the definition of $\eta_{t,i}$, and thus can be canceled by the last negative term in Eq. (1) (since $w'_{t,i}$ and $w_{t-1,i}$ are close). The complete proof can be found in Appendix B. ■

When $m_t = 0$, our bound exactly resolves the original impossible tuning issue (up to a $\ln T$ term). Below we discuss more implications of our bound by choosing different m_t.

Implication 1: improved variance or path-length bounds. Similarly to (Steinhardt and Liang, 2014), by setting m_t to be the running average of the loss vectors $\frac{1}{t-1}\sum_{s<t} \ell_s$, we obtain a bound that depends only on the variance of expert $i_\star$: $\mathcal{O}\left(\sqrt{\ln(dT)\sum_{t=1}^{T}(\ell_{t,i_\star} - \mu_{i_\star})^2}\right)$ where $\mu_{i_\star} = \frac{1}{T}\sum_{t=1}^{T} \ell_{t,i_\star}$. On the other hand, by setting $m_t = \ell_{t-1}$ (define $\ell_0 = \mathbf{0}$), we obtain a bound that depends only on the "path-length" of expert $i_\star$: $\mathcal{O}\left(\sqrt{\ln(dT)\sum_{t=1}^{T}(\ell_{t,i_\star} - \ell_{t-1,i_\star})^2}\right)$. The algorithm of (Steinhardt and Liang, 2014) uses a fixed learning rate and only achieves these bounds with an oracle tuning of the fixed learning rate, while our algorithm is completely adaptive and parameter-free.

In the next few implications, we make use of a trick similar to (Wei and Luo, 2018): if all coordinates of m_t are the same, then $\langle w, m_t \rangle$ is a constant independent of $w \in \Delta_d$ and thus $w_t = \operatorname{argmin}_{w \in \Omega_t} \langle w, m_t \rangle + D_{\psi_t}(w, w'_t) = \operatorname{argmin}_{w \in \Omega_t} D_{\psi_t}(w, w'_t)$, meaning that the algorithm and its guarantee are valid *even if* m_t *is set in terms of* ℓ_t which is unknown at the beginning of round t.

Implication 2: recovering $(\mathcal{A}, \mathcal{B})$-PROD guarantee. If we set $m_t = \ell_{t,1}\mathbf{1}$, then the regret against expert 1 becomes a constant $\mathcal{O}(\ln(dT))$ (while the regret against others remains $\mathcal{O}(\sqrt{T\ln(dT)})$). This is exactly the guarantee of the $(\mathcal{A}, \mathcal{B})$-PROD algorithm (Sani et al., 2014), useful for combining a set of base algorithms where one of them enjoys a regret bound significantly better than $\sqrt{T}$.

Implication 3: recovering ADAPT-ML-PROD guarantee. Next, we set $m_t = \langle w_t, \ell_t \rangle \mathbf{1}$ (again, valid even if unknown at the beginning of round t), leading to a bound $\mathcal{O}\left(\sqrt{\ln(dT)\sum_{t=1}^{T} r_{t,i_\star}^2}\right)$ where $r_{t,i} = \langle w_t - e_i, \ell_t \rangle$ is the instantaneous regret to expert i. A regret bound in terms of $\sqrt{\sum_{t=1}^{T} r_{t,i_\star}^2}$ is first achieved by the ADAPT-ML-PROD algorithm (Gaillard et al., 2014) (and later improved in (Koolen and Van Erven, 2015; Wintenberger, 2017)), and it has important consequences in achieving fast rates in stochastic settings; see (Koolen et al., 2016) for in-depth discussions.

Implication 4: recovering OPTIMISTIC-ADAPT-ML-PROD guarantee. By the same reason, it is also valid to set $m_t = m'_t + \langle w_t, \ell_t - m'_t \rangle \mathbf{1}$ for some prediction $m'_t \in [-1, +1]^d$ received at the beginning of round t.[3] Doing so leads to a bound $\mathcal{O}\left(\sqrt{\ln(dT)\sum_{t=1}^T r'^2_{t,i_\star}}\right)$ where $r'_{t,i} = \langle w_t - e_i, \ell_t - m'_t \rangle$ is the instantaneous regret to expert i measured with respect to the prediction difference $\ell_t - m'_t$. This bound first appears in OPTIMISTIC-ADAPT-ML-PROD (Wei et al., 2016) under the special choice of $m'_t = \ell_{t-1}$. In the following, we show that this bound preserves the fast rate consequences of the vanilla ADAPT-ML-PROD guarantee (Gaillard et al., 2014) (especially when m'_t is set to ℓ_{t-1}) in stochastic settings, while improving upon it whenever the predictions are accurate.

Theorem 3 *Suppose that $\ell_1, \ldots, \ell_T$ are generated randomly, and let $\mathbb{E}_t$ denote the conditional expectation given $\ell_1, \ldots, \ell_{t-1}$. Then the algorithm described in Implication 4 satisfies the following:*

- *If there exist $\Delta > 0$ and $i_\star$ such that $\mathbb{E}_t[\ell_{t,i} - \ell_{t,i_\star}] \geq \Delta$ for all t and $i \neq i_\star$, then with any $m'_t \in [-1,+1]^d$, $\mathrm{REG}(e_{i_\star}) = \mathcal{O}\left(\frac{\ln(dT)}{\Delta}\right)$ holds both in expectation and with high probability.*
- *If there exist $\kappa \in [0,1]$, $\Delta > 0$ and $i_\star$ such that $\mathbb{E}_t[\ell_{t,i} - \ell_{t,i_\star}]^\kappa \geq \Delta \mathbb{E}_t[(\ell_{t,i} - \ell_{t,i_\star})^2]$ for all t and $i \neq i_\star$, then with $m'_t = \ell_{t-1}$, $\mathrm{REG}(e_{i_\star}) = \mathcal{O}\left(\left(\frac{\ln(dT)}{\Delta}\right)^{\frac{1}{2-\kappa}} T^{\frac{1-\kappa}{2-\kappa}}\right)$ holds both in expectation and with high probability.*

The second condition in Theorem 3 is called the Bernstein condition and covers many interesting scenarios (Koolen et al., 2016). Note that in this case with $m'_t = \ell_{t-1}$, the algorithm *simultaneously* ensures a path-length bound $\mathcal{O}\left(\sqrt{\ln(dT)\sum_{t=1}^T \|\ell_t - \ell_{t-1}\|_\infty}\right)$ (since $r'_{t,i} \leq 2\|\ell_t - \ell_{t-1}\|_\infty$), which is useful for slowly changing environments such as some game playing settings (Rakhlin and Sridharan, 2013b; Syrgkanis et al., 2015). In Section 4, we also give an application for OLO.

We close this subsection with the following two remarks.

Differences in algorithms. We note that most existing algorithms discussed above are variants of either PROD (Sani et al., 2014; Gaillard et al., 2014) or "tilted exponential weight" (Koolen and Van Erven, 2015; Wintenberger, 2017),[4] which are somewhat similar to OMD with entropy regularizer. However, even if some of them adopt individual time-varying learning rates as well, they are different from our algorithm, as evidenced by the fact that these algorithm all take a closed "proportional" form, while our algorithm does not even when $\Omega_t = \Delta_d$ (see (Bubeck et al., 2017)). We are also only able to obtain our guarantee with a general m_t using this OMD framework but not the other methods (even though they achieve the bound for some special m_t as discussed). We conjecture that there are some subtle but fundamental differences between these algorithms.

Indeed impossible for bandits. It is natural to ask if the similar impossible tuning is in fact also possible for the more challenging multi-armed bandit problem (Auer et al., 2002), where the minimax regret is $\mathcal{O}(\sqrt{dT})$. In other words, is it possible to achieve $\mathrm{REG}(e_i) = \tilde{\mathcal{O}}\left(\sqrt{d\sum_{t=1}^T \ell_{t,i}^2}\right)$ for all i in multi-armed bandits? It turns out that this is *indeed impossible*, as a bound in this form would violate the multi-scale lower bound shown in (Bubeck et al., 2017, Theorem 23).

3. This is because $w_t = \mathrm{argmin}_{w\in\Omega_t} \langle w, m_t\rangle + D_{\psi_t}(w, w'_t) = \mathrm{argmin}_{w\in\Omega_t} \langle w, m'_t\rangle + D_{\psi_t}(w, w'_t)$. One caveat is that $m_{t,i}$ is now in the range of $[-3,+3]$, breaking the condition of Theorem 2, but this can be simply addressed by changing the constant 64 in the definition of $\eta_{t,i}$ to 128 so that the condition of Lemma 1 still holds.
4. The name "tilted exponential weight" is taken from (van Erven and Koolen, 2016).

Algorithm 2 MSMWC-MASTER

Input: a set of (learning rate, base algorithm) pairs $\mathcal{E}$.

Initialize: $p'_1 \in \Delta_{\mathcal{E}}$ such that $p'_{1,k} \propto \eta_k^2$ for each $k \in \mathcal{E}$.

for $t = 1, \ldots, T$ **do**

  Receive prediction $m_t \in \mathbb{R}^d$ and feed it to all base algorithms.

  For each $k \in \mathcal{E}$, receive decision $w_t^k \in \mathcal{K}$ from the base algorithm and define $h_{t,k} = \langle w_t^k, m_t \rangle$.

  Decide a compact convex decision subset $\Lambda_t \subseteq \Delta_{\mathcal{E}}$.

  Compute $p_t = \operatorname{argmin}_{p \in \Lambda_t} \langle p, h_t \rangle + D_\psi(p, p'_t)$ where $\psi(p) = \sum_{k \in \mathcal{E}} \frac{1}{\eta_k} p_k \ln p_k$.

  Play $w_t = \sum_{k \in \mathcal{E}} p_{t,k} w_t^k \in \mathcal{K}$, receive ℓ_t and feed it to all base algorithms.

  For each $k \in \mathcal{E}$, define $g_{t,k} = \langle w_t^k, \ell_t \rangle$ and $b_{t,k} = 32\eta_k (g_{t,k} - h_{t,k})^2$.

  Compute $p'_{t+1} = \operatorname{argmin}_{p \in \Lambda_t} \langle p, g_t + b_t \rangle + D_\psi(p, p'_t)$.

2.2. A new master algorithm

Next, we instantiate MSMWC differently to obtain a master algorithm MSMWC-MASTER that combines a set of base algorithms and adaptively learns the best one (see Algorithm 2). We will apply this master to both the expert problem (Section 3) and more generally the OLO problem (Section 4) where the decision set generalizes from Δ_d to an arbitrary closed convex set $\mathcal{K}$.

The instantiation still leaves the choices of Ω_t open for now and simply fixes the learning rate for each expert to be the same value over the T rounds. Since we will use this master, which itself deals with an expert problem with different base algorithms as experts, to deal with another expert/OLO problem, we adopt a different set of notations for the master. Specifically, the set of expert is denoted by $\mathcal{E}$, which consists of pairs in the form $(\eta, \mathcal{A})$ where η is the learning rate for this expert and $\mathcal{A}$ is a base algorithm. For each expert $k = (\eta, \mathcal{A}) \in \mathcal{E}$, we use η_k to denote the corresponding learning rate η.

MSMWC-MASTER maintains two sequences of distributions $p_1, \ldots, p_T$ and $p'_1, \ldots, p'_T$ over the set of experts. We use $\Delta_{\mathcal{E}}$ to denote the set of such distributions and $p_{t,k}$ to denote the weight assigned to expert k by p_t. We fix a specific initial distribution p'_1 such that $p'_{1,k} \propto \eta_k^2$. Upon receiving the prediction $m_t \in \mathbb{R}^d$ for the expert/OLO problem we are trying to solve, we feed it to all base algorithms, receive their decisions $\{w_t^k\}_{k \in \mathcal{E}}$, and then define the prediction $h_t \in \mathbb{R}^{\mathcal{E}}$ for the master expert problem with $h_{t,k} = \langle w_t^k, m_t \rangle$, that is, the predicted loss of the decision w_t^k. Next, MSMWC-MASTER decides a subset $\Lambda_t \in \Delta_{\mathcal{E}}$ and performs the OMD update with the regularizer $\psi(p) = \sum_{k \in \mathcal{E}} \frac{1}{\eta_k} p_k \ln p_k$ to compute p_t; note that the regularizer is now fixed over time.

With p_t, MSMWC-MASTER aggregates the decisions of all base algorithms by playing the convex combination $\sum_{k \in \mathcal{E}} p_{t,k} w_t^k$. After seeing the loss vector ℓ_t and feeding it to all base algorithms, MSMWC-MASTER naturally defines the loss vector $g_t \in \mathbb{R}^{\mathcal{E}}$ for its own expert problem with $g_{t,k} = \langle w_t^k, \ell_t \rangle$ and the corresponding correction term b_t with $b_{t,k} = 32\eta_k (g_{t,k} - h_{t,k})^2$. Finally, p'_{t+1} is calculated according to the OMD update rule using $g_t + b_t$.

To use MSMWC-MASTER, one simply designs a set of base algorithms with corresponding learning rates (and decides the subset Λ_t which is usually the set of distributions over some or all of the experts). These base algorithms are usually different instances of the same algorithm with different parameters such as a different learning rate, which usually coincides with the learning rate η_k for this expert. The point of having this construction is that MSMWC-MASTER can then learn the best parameter setting of the base algorithm automatically. Indeed, with $\text{REG}_{\mathcal{A}}$ being the regret of base algorithm $\mathcal{A}$, we have the following guarantee that is a direct corollary of Lemma 1.

Theorem 4 *Suppose that for all t, $32\eta_k |\left\langle w_t^k, \ell_t - m_t \right\rangle| \le 1$ holds for all $k \in \mathcal{E}$ with $p_{t,k} > 0$. Then for any $k_\star = (\eta_\star, \mathcal{A}_\star) \in \mathcal{E}$ such that $e_{k_\star} \in \bigcap_{t=1}^T \Lambda_t$,* MSMWC-MASTER *ensures*

$$\forall u \in \mathcal{K}, \;\; \text{REG}(u) \le \text{REG}_{\mathcal{A}_\star}(u) + \frac{1}{\eta_\star} \ln\left(\frac{\sum_k \eta_k^2}{\eta_\star^2}\right) + \frac{\sum_k \eta_k}{\sum_k \eta_k^2} + 32\eta_\star \sum_{t=1}^T \left\langle w_t^{k_\star}, \ell_t - m_t \right\rangle^2. \quad (2)$$

The proof is deferred to Appendix B. In all our applications, the learning rates are chosen from an exponential grid such that $\sum_k \eta_k$ and $\sum_k \eta_k^2$ are both constants. Moreover, the term $32\eta_\star \sum_{t=1}^T \langle w_t^{k_\star}, \ell_t - m_t\rangle^2$ can usually be *canceled* by the a negative term from $\text{REG}_{\mathcal{A}_\star}(u)$, making the overhead of the master simply be $\mathcal{O}(\frac{1}{\eta_\star} \ln \frac{1}{\eta_\star})$, which is rather small. We remark that the idea of combining a set of base algorithms or more specifically "learning the learning rate" has appeared in many prior works such as (Koolen et al., 2014; van Erven and Koolen, 2016; Foster et al., 2017; Cutkosky, 2019b; Bhaskara et al., 2020). However, the special regret guarantee of MSMWC that does not exist before allows us to derive new applications as shown in the next two sections.

3. Applications to the Expert Problem

In this section, we apply MSMWC-MASTER to derive yet another four new results for the expert problem (thus $\mathcal{K} = \Delta_d$ throughout this section). These results improve over the guarantee of Theorem 2 by respectively adapting to an arbitrary competitor and a prior, the scale of each expert, a switching sequence of competitors, and unknown loss ranges.[5]

3.1. Adapting to an arbitrary competitor

Typical regret bounds for the expert problem compete with an individual expert and pay for a $\sqrt{\ln d}$ factor. Several works generalize this by replacing $\ln d$ with $\text{KL}(u, \pi)$ when competing with an arbitrary competitor $u \in \Delta_d$, where π is a fixed prior distribution over the experts (Luo and Schapire, 2015; Koolen and Van Erven, 2015). Importantly, the bound holds simultaneously for all u. Inspired by these works, our goal here is to make the same generalization for Theorem 2. To do so, we again instantiate MSMWC differently to create a set of base algorithms, each with a fixed learning rate across all i and t (so both the master and the base algorithms are instances of MSMWC). Specifically, consider the following set of $\mathcal{O}(\ln T)$ experts:

$$\begin{aligned}\mathcal{E}_{\text{KL}} = \Big\{ (\eta_k, \mathcal{A}_k) : \forall k = 1, \dots, \lceil \log_2 T \rceil, \eta_k = \tfrac{1}{32 \cdot 2^k}, \mathcal{A}_k \text{ is MSMWC with } w_1' = \pi, \\ \Omega_t = \Delta_d, \text{ and } \eta_{t,i} = 2\eta_k \text{ for all } t \text{ and } i \Big\}. \end{aligned} \quad (3)$$

By Lemma 1, we know that $\mathcal{A}_k$ guarantees for all $u \in \Delta_d$:

$$\text{REG}_{\mathcal{A}_k}(u) \le \frac{\text{KL}(u, \pi)}{2\eta_k} + 64\eta_k \sum_{t=1}^T \sum_{i=1}^d u_i (\ell_{t,i} - m_{t,i})^2 - 32\eta_k \sum_{t=1}^T \sum_{i=1}^d w_{t,i}^k (\ell_{t,i} - m_{t,i})^2. \quad (4)$$

5. While we present all results using the master with appropriate base algorithms, it is actually possible to "flatten" this two-layer structure to just one layer by duplicating each expert and assigning each copy a different learning rate. We omit the details since this approach does not generalize to OLO.

MsMwC-Master can then learn the best η_k to achieve the optimal tuning. Indeed, directly combining the guarantee of MsMwC-Master from Theorem 4 and noting that, importantly, the last term in Eq. (2) can be canceled by the last negative term in Eq. (4) by Cauchy-Schwarz inequality, we obtain the following result (full proof deferred to Appendix C).

Theorem 5 *Suppose* $\|\ell_t - m_t\|_\infty \le 1, \forall t$. *Then for any* $\pi \in \Delta_d$, MsMwC-Master *with expert set* $\mathcal{E}_{\mathrm{KL}}$ *and* $\Lambda_t = \Delta_{\mathcal{E}_{\mathrm{KL}}}$ *ensures* $\mathrm{Reg}(u) = \mathcal{O}\left(\mathrm{KL}(u,\pi) + \ln V(u) + \sqrt{(\mathrm{KL}(u,\pi) + \ln V(u))V(u)}\right)$ *for all* $u \in \Delta_d$, *where* $V(u) = \max\left\{3, \sum_{t=1}^T \sum_{i=1}^d u_i(\ell_{t,i} - m_{t,i})^2\right\}$.

This result recovers the guarantee in Theorem 2 when $u = e_{i_\star}$ and π is uniform (in fact, it also improves the $\ln T$ factor to $\ln V(e_{i_\star})$).[6] Note that the implications discussed in Section 2.1 by selecting different m_t still apply here with the same improvement (from $\ln(dT)$ to $\mathrm{KL}(u,\pi) + \ln V(u)$). In particular, this means that our results recover and improve those of (Luo and Schapire, 2015; Koolen and Van Erven, 2015) (which only cover the case with $m_t = \langle w_t, \ell_t \rangle \mathbf{1}$).

3.2. Adapting to Multiple Scales

Consider the "multi-scale" expert problem (Bubeck et al., 2017; Foster et al., 2017; Cutkosky and Orabona, 2018) where each expert i has a different loss range $c_i > 0$ such that $|\ell_{t,i}| \le c_i$ (and naturally $|m_{t,i}| \le c_i$) for all t. Previous works all achieve a bound $\mathrm{Reg}(e_{i_\star}) = \tilde{\mathcal{O}}(c_{i_\star}\sqrt{T \ln d})$, scaling only in terms of $c_{i_\star}$. The main term of our bound in Theorem 2 is already *strictly better* since the term $\sqrt{\sum_{t=1}^T (\ell_{t,i_\star} - m_{t,i_\star})^2} \le 2c_{i_\star}\sqrt{T}$ inherently only scales with $c_{i_\star}$. The issue is that the lower-order term in the bound is in fact in terms of $\max_i c_i$. To improve it to $c_{i_\star}$, we apply similar ideas of Section 3.1 and again use MsMwC-Master to learn the best learning rate for the base algorithm MsMwC. To this end, first define a set $\mathcal{S} = \left\{k \in \mathbb{Z} : \exists i \in [d], c_i \le 2^{k-2} \le c_i\sqrt{T}\right\}$ so that $\{\frac{1}{32 \cdot 2^k}\}_{k \in \mathcal{S}}$ contains all the learning rates we want to search over. Then define expert set:

$$\begin{aligned} \mathcal{E}_{\mathrm{MS}} = \Big\{ &(\eta_k, \mathcal{A}_k) : \forall k \in \mathcal{S}, \eta_k = \tfrac{1}{32 \cdot 2^k}, \mathcal{A}_k \text{ is MsMwC with } w_1' \text{ being uniform over } \mathcal{Z}(k), \\ &\Omega_t = \{w \in \Delta_d : w_i = 0, \forall i \notin \mathcal{Z}(k)\}, \text{ and } \eta_{t,i} = 2\eta_k \text{ for all } t \text{ and } i \Big\}, \end{aligned} \tag{5}$$

where $\mathcal{Z}(k) = \{i \in [d] : c_i \le 2^{k-2}\}$. Compared to Eq. (3), another difference is that we restrict each base algorithm $\mathcal{A}_k$ to work with only a subset $\mathcal{Z}(k)$ of arms, which ensures the condition $32\eta_{t,i}|\ell_{t,i} - m_{t,i}| \le 128\eta_k c_i \le 1$ (for i with $w_{t,i}^k > 0$) of Lemma 1 and similarly the condition of Theorem 4. With this construction, we can then automatically learn the best instance and achieve the following multi-scale bound that is a strict improvement of aforementioned previous works.

Theorem 6 *Suppose for all* t, $|\ell_{t,i}| \le c_i$ *and* $|m_{t,i}| \le c_i$ *for some* $c_i > 0$. *Define* $c_{\min} = \min_i c_i$ *and* $\Gamma_i = \ln(\frac{dTc_i}{c_{\min}})$. *Then* MsMwC-Master *with expert set* $\mathcal{E}_{\mathrm{MS}}$ *defined in Eq. (5) and* $\Lambda_t = \Delta_{\mathcal{S}_{\mathrm{MS}}}$ *ensures:* $\mathrm{Reg}(e_{i_\star}) = \tilde{\mathcal{O}}\left(c_{i_\star}\Gamma_{i_\star} + \sqrt{\Gamma_{i_\star}\sum_{t=1}^T(\ell_{t,i_\star} - m_{t,i_\star})^2}\right)$ *for all* $i_\star \in [d]$.

6. However, we believe that the result of Theorem 2 is still valuable since the algorithm does not require maintaining multiple base algorithms and is more computationally efficient and practical.

3.3. Adapting to a switching sequence

So far, the regret measure we have considered compares with a fixed competitor across all T rounds. A more challenging notion of regret, called *switching regret*, compares with a sequence of changing competitors with a certain number of switches, which is a much more appropriate measure for non-stationary environments. Specifically, we use $\mathcal{I}$ to denote an interval of rounds (that is, a subset of $[T]$ in the form of $\{s, s+1, \ldots, t-1, t\}$) and $\textsc{Reg}^{\mathcal{I}}(u) = \sum_{t\in\mathcal{I}} \langle w_t - u, \ell_t\rangle$ to denote the regret against u on this interval. For a partition $\mathcal{I}_1, \ldots, \mathcal{I}_S$ of $[T]$ and competitors $u_1, \ldots, u_S \in \Delta_d$, the corresponding switching regret is then $\sum_{j=1}^{S} \textsc{Reg}^{\mathcal{I}_j}(u_j)$.

Now, we show that almost the same construction as in Section 3.1 generalizes our result in Theorem 2 to switching regret as well. Specifically, we deploy the following expert set:

$$\begin{aligned}\mathcal{E}_{\text{switch}} = \Big\{(\eta_k, \mathcal{A}_k) : \forall k = 1, \ldots, \lceil \log_2 T\rceil, \eta_k = \tfrac{1}{32\cdot 2^k}, \mathcal{A}_k \text{ is } \textsc{MsMwC} \text{ with } w_1' = \tfrac{1}{d}\mathbf{1}, \\ \Omega_t = \big\{w \in \Delta_d : w_i \geq \tfrac{1}{dT}\big\}, \text{ and } \eta_{t,i} = 2\eta_k \text{ for all } t \text{ and } i\Big\},\end{aligned} \tag{6}$$

where the only essential difference compared to $\mathcal{E}_{\text{KL}}$ is the the use of a truncated simplex for Ω_t. We then have the following new switching regret guarantee.

Theorem 7 *If $\|\ell_t - m_t\|_\infty \leq 1$ holds for all $t \in [T]$, then* MsMwC-Master *with expert set $\mathcal{E}_{\text{switch}}$ defined in Eq. (6) and $\Lambda_t = \big\{p \in \Delta_{\mathcal{E}_{\text{KL}}} : p_k \geq \frac{1}{T}\big\}$ ensures for any partition $\mathcal{I}_1, \ldots, \mathcal{I}_S$ of $[T]$ and competitors $u_1, \ldots, u_S \in \Delta_d$,*

$$\sum_{j=1}^{S} \textsc{Reg}^{\mathcal{I}_j}(u_j) = \mathcal{O}\left(S\ln(dT) + \sum_{j=1}^{S}\sqrt{\ln(dT)\sum_{t\in\mathcal{I}_j}\sum_{i=1}^{d} u_{j,i}(\ell_{t,i} - m_{t,i})^2}\right). \tag{7}$$

Our bound is never worse than the typical one $\mathcal{O}(\sqrt{ST\ln(dT)})$ (due to Cauchy-Schwarz inequality) and significantly improves over previous works such as (Cesa-Bianchi et al., 2012; Luo and Schapire, 2015) by again choosing different m_t according to the discussions in Section 2.1. It also resolves an open problem raised by Lu and Zhang (2019) on the possibility of making the switching regret bound adapt to the path length of the comparator sequence. The proof of Theorem 7 requires a more general version of Lemma 1 and is deferred to Appendix C.

Impossibility for interval regret. Looking at Eq. (7), one might wonder whether the natural bound $\textsc{Reg}^{\mathcal{I}_j}(u_j) = \mathcal{O}\big(\ln(dT) + \sqrt{\ln(dT)\sum_{t\in\mathcal{I}_j}\sum_{i=1}^{d} u_{j,i}(\ell_{t,i} - m_{t,i})^2}\big)$ holds for each interval $\mathcal{I}_j$ separately. Indeed, Eq. (7) could be derived from this (by summing over S intervals). It turns out that, even if its special case with $m_t = \langle w_t, \ell_t\rangle$ is achievable (Luo and Schapire, 2015), this cannot hold in general as shown in Appendix C.4. We find this intriguing (given that Eq. (7) is achievable) and reminiscent of the impossibility result for interval regret in bandits (Daniely et al., 2015).

3.4. Adapting to unknown loss ranges

The recent work of (Mhammedi et al., 2019) improves (Koolen and Van Erven, 2015) by adapting to the unknown loss range $\|\ell_t\|_\infty$. Here, we show that MsMwC-Master is readily capable of dealing with such cases as well. The high-level idea is to have each base algorithm to deal with a different possible loss range — a larger loss range is handled by a smaller learning rate. Once the loss becomes larger than what a base algorithm can handle, we remove this algorithm from the expert set, simply implemented by defining Λ_t to be a subset of distributions that put zero weight on this base algorithm. The removal of these base algorithms is necessary to ensure that the condition $32\eta_k|\left\langle w_t^k, \ell_t - m_t\right\rangle| \le 1$ of Theorem 4 always holds. We defer the details to Appendix C.5, which include some additional techniques similar to those of (Mhammedi et al., 2019) such as feeding the algorithm with truncated fake losses and a restarting scheme. Our final result is summarized below.

Theorem 8 *Let $\max_t \|\ell_t - m_t\|_\infty$ be unknown. For any prior $\pi \in \Delta_d$, Algorithm 3 (with input Eq. (19) and B_0) ensures* $\textsc{Reg}(u) = \mathcal{O}\left(B(\mathrm{KL}(u,\pi) + \ln T) + \sqrt{(\mathrm{KL}(u,\pi) + \ln T)V(u)}\right)$, $\forall u \in \Delta_d$, *where* $V(u) = \max\left\{3, \sum_{t=1}^T \sum_{i=1}^d u_i(\ell_{t,i} - m_{t,i})^2\right\}$ *and* $B = \max\{B_0, \max_t \|\ell_t - m_t\|_\infty\}$.

Note that B is in terms of the maximum range of the predicted error as opposed to $\max_t \|\ell_t\|_\infty$ used in (Mhammedi et al., 2019), and could be much smaller when the prediction is accurate. Besides, (Mhammedi et al., 2019) only achieves the bound with $m_t = \langle w_t, \ell_t\rangle \mathbf{1}$ in $V(u)$.

4. Applications to Online Linear Optimization

We next discuss applications of MsMwC-Master to general OLO. For simplicity, we assume that $\mathcal{K}$ is a compact convex set such that $\|w\| \le D$ for all $w \in \mathcal{K}$, and also $\max_t \|\ell_t - m_t\| \le 1$, where $\|\cdot\|$ is L_2 norm (extensions to general primal-dual norm are straightforward). In Appendix D.5, we show that all our results can be generalized to the unconstrained setting where $\mathcal{K}$ is unbounded and also the unknown Lipschitzness setting where $\max_t \|\ell_t - m_t\|$ is unknown ahead of time.

Application 1: combining Online Newton Step It is a folklore that one can reduce OLO to the expert problem by discretizing the decision set $\mathcal{K}$ into $\mathcal{O}\left(T^d\right)$ points and treating each point as an expert. With this reduction, our result in Theorem 2 immediately implies a bound $\textsc{Reg}(u) = \tilde{\mathcal{O}}(\sqrt{d\sum_t \langle u, \ell_t - m_t\rangle^2})$ for OLO. Of course, the caveat is that the reduction is computationally inefficient.[7] Below, we show that the same (or even better) bound can be achieved efficiently by using MsMwC-Master with a variant of Online Newton Step (ONS) (Hazan et al., 2007) as the base algorithm. Specifically, the ONS variant (denoted by $\mathcal{A}_k$ and parameterized by a fixed learning rate η) can be presented in the OMD framework again using an auxiliary cost function $c_t(w) = \langle w, \ell_t\rangle + 32\eta \langle w, \ell_t - m_t\rangle^2$ and a time-varying regularizer $\psi_t(w) = \frac{1}{2}\|w\|_{A_t}^2$ where $A_t = \eta\left(2I + \sum_{s<t}(\nabla_s - m_s)(\nabla_s - m_s)^\top\right)$ and $\nabla_s = \nabla c_s(w_s^k)$. This variant is similar to that in (Cutkosky and Orabona, 2018), but incorporates the prediction m_t as well. We defer the details to Appendix D.1, which shows: $\mathcal{A}_k$ ensures (with r being the rank of $\mathcal{L}_T = \sum_{t=1}^T(\ell_t - m_t)(\ell_t - m_t)^\top$)

$$\textsc{Reg}(u) \le \tilde{\mathcal{O}}\left(\frac{r}{\eta} + \eta\sum_{t=1}^T \langle u, \ell_t - m_t\rangle^2\right) - 16\eta\sum_{t=1}^T \left\langle w_t^k, \ell_t - m_t\right\rangle^2. \tag{8}$$

Therefore, using MsMwC-Master to learn the best learning rate and noting that the last negative term in Eq. (8) cancels the last term in Eq. (2), we obtain the following result.

7. The reduction is efficient when $d = 1$ though. This gives an alternative algorithm with the same guarantee as (Cutkosky and Orabona, 2018, Theorem 1) and is useful already with their reduction from general d to $d = 1$.

Theorem 9 *Let $r \le d$ be the rank of $\mathcal{L}_T = \sum_{t=1}^T (\ell_t - m_t)(\ell_t - m_t)^\top$.* MsMwC-MASTER *with expert set $\mathcal{E}_{\text{ONS}}$ defined in Eq. (21) and $\Lambda_t = \Delta_{\mathcal{E}_{\text{ONS}}}$ ensures*

$$\forall u \in \mathcal{K}, \quad \text{REG}(u) = \tilde{\mathcal{O}}\left(r\,\|u\| + \sqrt{r \sum_{t=1}^T \langle u, \ell_t - m_t \rangle^2} \right). \tag{9}$$

Similar bounds appear before but only with $m_t = 0$ (Cutkosky and Orabona, 2018; Cutkosky, 2020), and we are not able to incorporate general m_t into their algorithms. Our bound has no explicit dependence on D at all, and its dependence on $\ell_t - m_t$ is only through its projection on u.

Application 2: combining Gradient Descent Another natural choice of base algorithm is Optimistic Gradient Descent, which guarantees $\text{REG}(u) = \mathcal{O}\big(\frac{\|u\|^2}{\eta} + \eta \sum_{t=1}^T \|\ell_t - m_t\|^2\big)$ (see Appendix D.2). Combining instances with different learning rates that operate over subsets of $\mathcal{K}$ of different sizes (necessary to ensure $32\eta_k |\langle w_t^k, \ell_t - m_t \rangle| \le 1$ for Theorem 4), we obtain:

Theorem 10 MsMwC-MASTER *with expert set $\mathcal{E}_{\text{GD}}$ defined in Eq. (22) and $\Lambda_t = \Delta_{\mathcal{E}_{\text{GD}}}$ ensures*

$$\forall u \in \mathcal{K}, \quad \text{REG}(u) = \tilde{\mathcal{O}}\left(\|u\| + \|u\| \sqrt{\sum_{t=1}^T \|\ell_t - m_t\|^2} \right). \tag{10}$$

This bound appears before first with $m_t = 0$ in (Cutkosky and Orabona, 2018) and later with general m_t in (Cutkosky, 2019b). We recover the bound easily with our framework. Similar to Eq. (9), this bound adapts to the size of the competitor u (with no dependence on D). An advantage of Eq. (10) is that it is dimension-free, while Eq. (9) is potentially large for high-dimensional data.

Application 3: combining AdaGrad Inspired by the recent work of (Cutkosky, 2020) that provides an improved guarantee of the full-matrix version of AdaGrad (Duchi et al., 2011), we next design an optimistic version of AdaGrad and combine instances with different parameters to obtain the following new result.

Theorem 11 MsMwC-MASTER *with expert set $\mathcal{E}_{\text{AG}}$ defined in Eq. (23) and $\Lambda_t = \Delta_{\mathcal{E}_{\text{AG}}}$ ensures*

$$\forall u \in \mathcal{K}, \quad \text{REG}(u) = \tilde{\mathcal{O}}\left(\|u\| + \sqrt{\left(u^\top (I + \mathcal{L}_T)^{1/2} u\right) \operatorname{tr}\left(\mathcal{L}_T^{1/2}\right)} \right). \tag{11}$$

All details are deferred to Appendix D.3. Cutkosky (2020) achieves Eq. (11) for $m_t = 0$ (again, we are not able to extend their algorithm to deal with general m_t). The three types of bounds we have shown in Eq. (9), Eq. (10), and Eq. (11) are *incomparable*, that is, there are cases for each one to be the smallest; see (Cutkosky, 2020) for in-depth discussions with $m_t = 0$. However, since the configuration of MsMwC-MASTER is the same in all these three results (other than the expert set), we can in fact achieve *the best of three worlds* by feeding the union of these three expert set to MsMwC-MASTER, summarized in the following corollary.

Corollary 12 (Best-of-three-worlds) MsMwC-MASTER *with expert set $\mathcal{E} = \mathcal{E}_{\text{ONS}} \cup \mathcal{E}_{\text{GD}} \cup \mathcal{E}_{\text{AG}}$ and $\Lambda_t = \Delta_{\mathcal{E}}$ ensures regret bounds Eq. (9), Eq. (10), and Eq. (11) simultaneously.*

We remark that the technique proposed in (Cutkosky, 2019b) can similarly combine algorithm's guarantees with little overhead, but it only works for the unconstrained setting. It is tempting to apply the unconstrained-to-constrained reduction from (Cutkosky and Orabona, 2018) to lift this restriction, but that does not work generally as discussed in (Cutkosky, 2020, Section 4). All in all, we are not aware of any other methods capable of achieving this best-of-three-worlds result.

Application 4: combining MetaGrad's base algorithm Finally, we discuss how to recover and generalize the regret bound of MetaGrad (van Erven and Koolen, 2016) which depends on the sum of squared instantaneous regret and is the analogue of the ADAPT-ML-PROD guarantee for the expert problem. Our base algorithm is yet another variant of ONS that uses a different auxiliary cost function $c_t(w) = \langle w, \ell_t \rangle + 32\eta \langle w - w_t, \ell_t - m_t \rangle^2$ with an extra offset in terms of w_t (the decision of the master). When $m_t = 0$ this is the same base algorithm used in (van Erven and Koolen, 2016). Compared to Eq. (8), this variant ensures the following

$$\text{REG}(u) \leq \tilde{\mathcal{O}}\left(\frac{r}{\eta} + \eta \sum_{t=1}^{T} \langle u - w_t, \ell_t - m_t \rangle^2\right) - 16\eta \sum_{t=1}^{T} \left\langle w_t^k - w_t, \ell_t - m_t \right\rangle^2. \tag{12}$$

Note that the last negative term is now slightly different from the last term in Eq. (2). To make them match, we need to change the definition of $h_{t,k}$ in MSMWC-MASTER from $h'_{t,k} \stackrel{\text{def}}{=} \langle w_t^k, m_t \rangle$ to $h'_{t,k} + \langle p_t, g_t - h'_t \rangle$, the same trick used in Implication 4 of Section 2.1 (this is also the reason why we cannot include this result in Corollary 12 as well). We defer the details to Appendix D.4 and show the final bound below.

Theorem 13 *Let $r \leq d$ be the rank of $\sum_{t=1}^{T} (\ell_t - m_t)(\ell_t - m_t)^\top$. MSMWC-MASTER with the new definition of h_t described above, expert set $\mathcal{E}_{\text{MG}}$ defined in Eq. (24), and $\Lambda_t = \Delta_{\mathcal{E}_{\text{MG}}}$ ensures $\forall u \in \mathcal{K}, \text{REG}(u) = \left(rD + \sqrt{r \sum_{t=1}^{T} \langle u - w_t, \ell_t - m_t \rangle^2}\right)$.*

This bound generalizes the MetaGrad's guarantee from $m_t = 0$ to general m_t and is the analogue of the bound discussed in Implication 4 of Section 2.1 for the expert problem. Similarly to Theorem 3, when using $m_t = \ell_{t-1}$, our bound preserves all the fast rate consequences discussed in (van Erven and Koolen, 2016; Koolen et al., 2016), while ensuring a bound in terms of only the variation of the loss vectors $\sum_t \|\ell_t - \ell_{t-1}\|^2$. We remark that MetaGrad also uses a master algorithm to combine similar ONS variants, but the master is "tilted exponential weight" and cannot incorporate general m_t.

5. Discussions and Open Problems

We mention two open questions for the expert problem. First, in the case when we are required to select one expert i_t randomly in each round t, and the regret against i is measured by $\sum_{t=1}^{T} \ell_{t,i_t} - \ell_{t,i}$, it is unclear how to achieve our bounds such as $\tilde{\mathcal{O}}\left(\sqrt{(\ln d) \sum_{t=1}^{T} \ell_{t,i}^2}\right)$ with high probability (even though our results clearly imply this in expectation). The difficulty lies in handling the deviation between $\sum_{t=1}^{T} \langle w_t, \ell_t \rangle$ and $\sum_{t=1}^{T} \ell_{t,i_t}$ and bounding it in terms of only $\sum_{t=1}^{T} \ell_{t,i}^2$. We conjecture that impossible tuning might indeed be impossible in this case.

Second, note that even though we only focus on having one prediction sequence $\{m_t\}_{t \in [T]}$, we can in fact also deal with multiple sequences and learn the best via another expert algorithm, similarly to (Rakhlin and Sridharan, 2013a). One caveat is that the trick we apply in Implications 2-4 of Section 2.1 (that m_t can depend on ℓ_t even though it is unknown) does not work anymore, since different experts might be using different sources of predictions and thus the calculation of w_t does require knowing all predictions at the beginning of round t. Due to this issue, we for example cannot achieve a bound in the form of

$$\forall i \in [d], \quad \textsc{Reg}(e_i) = \tilde{\mathcal{O}}\left(\sqrt{(\ln d) \min\left\{\sum_{t=1}^{T} \ell_{t,i}^2, \sum_{t=1}^{T} (\ell_{t,i} - \langle w_t, \ell_t \rangle)^2\right\}}\right).$$

We leave the possibility of achieving such a bound as an open problem.

Acknowledgments

This work is supported by NSF Award IIS-1943607.

References

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multi-armed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.

Aditya Bhaskara, Ashok Cutkosky, Ravi Kumar, and Manish Purohit. Online linear optimization with many hints. *Advances in neural information processing systems*, 2020.

Sébastien Bubeck, Nikhil R Devanur, Zhiyi Huang, and Rad Niazadeh. Online auctions and multi-scale online learning. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 497–514, 2017.

Nicolo Cesa-Bianchi, Yishay Mansour, and Gilles Stoltz. Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66(2-3):321–352, 2007.

Nicolo Cesa-Bianchi, Pierre Gaillard, Gábor Lugosi, and Gilles Stoltz. Mirror descent meets fixed share (and feels no regret). *Advances in Neural Information Processing Systems*, 25:980–988, 2012.

Liyu Chen, Haipeng Luo, and Chen-Yu Wei. Minimax regret for stochastic shortest path with adversarial costs and known transition. *Conference on Learning Theory*, 2021.

Ashok Cutkosky. Artificial constraints and hints for unbounded online learning. In *Conference on Learning Theory*, pages 874–894, 2019a.

Ashok Cutkosky. Combining online learning guarantees. *Conference on Learning Theory*, 2019b.

Ashok Cutkosky. Better full-matrix regret via parameter-free online learning. *Advances in Neural Information Processing Systems*, 33, 2020.

Ashok Cutkosky and Francesco Orabona. Black-box reductions for parameter-free online learning in banach spaces. In *Conference on Learning Theory (COLT)*, pages 1493–1529, 2018.

Amit Daniely, Alon Gonen, and Shai Shalev-Shwartz. Strongly adaptive online learning. In *International Conference on Machine Learning*, pages 1405–1411, 2015.

John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

Dylan J Foster, Satyen Kale, Mehryar Mohri, and Karthik Sridharan. Parameter-free online learning via model selection. In *Advances in Neural Information Processing Systems*, pages 6020–6030, 2017.

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, 1997.

Pierre Gaillard, Gilles Stoltz, and Tim Van Erven. A second-order bound with excess losses. In *Conference on Learning Theory*, pages 176–196, 2014.

Elad Hazan and Satyen Kale. Extracting certainty from uncertainty: Regret bounded by variation in costs. *Machine learning*, 80(2-3):165–188, 2010.

Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

Wouter M Koolen and Tim Van Erven. Second-order quantile methods for experts and combinatorial games. In *Conference on Learning Theory*, pages 1155–1175, 2015.

Wouter M Koolen, Tim Van Erven, and Peter Grünwald. Learning the learning rate for prediction with expert advice. *Advances in neural information processing systems*, 27:2294–2302, 2014.

Wouter M Koolen, Peter Grünwald, and Tim van Erven. Combining adversarial guarantees and stochastic fast rates in online learning. *Advances in Neural Information Processing Systems*, 29: 4457–4465, 2016.

Tomer Koren and Roi Livni. Affine-invariant online optimization and the low-rank experts problem. In *Advances in Neural Information Processing Systems*, pages 4747–4755, 2017.

Shiyin Lu and Lijun Zhang. Adaptive and efficient algorithms for tracking the best expert. *arXiv preprint arXiv:1909.02187*, 2019.

Haipeng Luo and Robert E Schapire. Achieving all with no parameters: Adanormalhedge. In *Conference on Learning Theory*, pages 1286–1304, 2015.

Zakaria Mhammedi and Wouter M Koolen. Lipschitz and comparator-norm adaptivity in online learning. *Conference on Learning Theory*, 2020.

Zakaria Mhammedi, Wouter M Koolen, and Tim Van Erven. Lipschitz adaptivity with multiple learning rates in online learning. *Conference on Learning Theory*, 2019.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. *Conference on Learning Theory*, 2013a.

Alexander Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. *Advances in Neural Information Processing Systems*, 26:3066–3074, 2013b.

Amir Sani, Gergely Neu, and Alessandro Lazaric. Exploiting easy data in online optimization. *Advances in Neural Information Processing Systems*, 27:810–818, 2014.

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

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. *Advances in Neural Information Processing Systems*, 28:2989–2997, 2015.

Tim van Erven and Wouter M Koolen. Metagrad: Multiple learning rates in online learning. *Advances in Neural Information Processing Systems*, 2016.

Chen-Yu Wei and Haipeng Luo. More adaptive algorithms for adversarial bandits. In *Conference On Learning Theory*, pages 1263–1291, 2018.

Chen-Yu Wei, Yi-Te Hong, and Chi-Jen Lu. Tracking the best expert in non-stationary stochastic environments. *Advances in neural information processing systems*, 29:3972–3980, 2016.

Olivier Wintenberger. Optimal learning with bernstein online aggregation. *Machine Learning*, 106 (1):119–141, 2017.

Appendix A. Useful Lemmas Related to OMD

Lemma 14 *Define $w^\star = \operatorname{argmin}_{w\in\mathcal{K}} \langle w, x\rangle + D_\psi(w, w')$ for some compact convex set $\mathcal{K} \subset \mathbb{R}^d$, convex function ψ, an arbitrary point $x \in \mathbb{R}^d$, and a point $w' \in \mathcal{K}$. Then for any $u \in \mathcal{K}$:*

$$\langle w^\star - u, x\rangle \le D_\psi(u, w') - D_\psi(u, w^\star) - D_\psi(w^\star, w').$$

Proof This is shown for example in the proof of (Wei and Luo, 2018, Lemma 1), and is by direct calculations plus the first-order optimality condition of $w^\star$. ∎

Lemma 15 *Let $w_t = \operatorname{argmin}_{w\in\mathcal{K}} \langle w, m_t\rangle + D_{\psi_t}(w, w'_t)$ and $w'_{t+1} = \operatorname{argmin}_{w\in\mathcal{K}} \langle w, \ell_t\rangle + D_{\psi_t}(w, w'_t)$ for some compact convex set $\mathcal{K} \subset \mathbb{R}^d$, convex function ψ_t, arbitrary points $\ell_t, m_t \in \mathbb{R}^d$, and a point $w'_t \in \mathcal{K}$. Then, for any $u \in \mathcal{K}$ we have*

$$\langle w_t - u, \ell_t\rangle \le \left\langle w_t - w'_{t+1}, \ell_t - m_t\right\rangle + D_{\psi_t}(u, w'_t) - D_{\psi_t}(u, w'_{t+1}) - D_{\psi_t}(w'_{t+1}, w_t) - D_{\psi_t}(w_t, w'_t).$$

Proof We apply Lemma 14 with $w^\star = w_t, u = w'_{t+1}$ to obtain

$$\left\langle w_t - w'_{t+1}, m_t\right\rangle \le D_{\psi_t}(w'_{t+1}, w'_t) - D_{\psi_t}(w'_{t+1}, w_t) - D_{\psi_t}(w_t, w'_t),$$

and then with $w^\star = w'_{t+1}$ to obtain:

$$\left\langle w'_{t+1} - u, \ell_t\right\rangle \le D_{\psi_t}(u, w'_t) - D_{\psi_t}(u, w'_{t+1}) - D_{\psi_t}(w'_{t+1}, w'_t).$$

Summing the two inequalities above, we have:

$$\left\langle w_t - w'_{t+1}, m_t\right\rangle + \left\langle w'_{t+1} - u, \ell_t\right\rangle \le D_{\psi_t}(u, w'_t) - D_{\psi_t}(u, w'_{t+1}) - D_{\psi_t}(w'_{t+1}, w_t) - D_{\psi_t}(w_t, w'_t).$$

Also note that the left-hand side is equal to:

$$\begin{aligned}\left\langle w_t - w'_{t+1}, m_t\right\rangle + \left\langle w'_{t+1} - u, \ell_t\right\rangle &= \left\langle w_t - w'_{t+1}, m_t - \ell_t\right\rangle + \left\langle w_t - w'_{t+1}, \ell_t\right\rangle + \left\langle w'_{t+1} - u, \ell_t\right\rangle \\ &= \left\langle w_t - w'_{t+1}, m_t - \ell_t\right\rangle + \langle w_t - u, \ell_t\rangle .\end{aligned}$$

Combining and reorganizing terms, we get the desired result. ∎

Lemma 16 *For any convex function ψ defined on convex set $\mathcal{K} \subset \mathbb{R}^d$ and a point $x \in \mathbb{R}^d$, define $F_x(w) = \langle w, x\rangle + \psi(w)$ and $w_x = \operatorname{argmin}_{w\in\mathcal{K}} F_x(w)$. Suppose that for some $x, x' \in \mathbb{R}^d$, there is a constant c such that for all ξ on the segment connecting w_x and $w_{x'}$, $\nabla^2\psi(\xi) \succcurlyeq c\nabla^2\psi(w_x)$ holds (which means $\nabla^2\psi(\xi) - c\nabla^2\psi(w_x)$ is positive semi-definite). Then, we have $\langle w_x - w_{x'}, x' - x\rangle \ge 0$ and $\|w_x - w_{x'}\|_{\nabla^2\psi(w_x)} \le \frac{2}{c}\|x - x'\|_{\nabla^{-2}\psi(w_x)}$.*

Proof Note that

$$\begin{aligned}&F_{x'}(w_x) - F_{x'}(w_{x'}) \\ &= \left\langle w_x - w_{x'}, x' - x\right\rangle + F_x(w_x) - F_x(w_{x'}) && \text{(definition of } F) \\ &\le \left\langle w_x - w_{x'}, x' - x\right\rangle && \text{(optimality of } w_x)\end{aligned}$$

$$\leq \|w_x - w_{x'}\|_{\nabla^2\psi(w_x)} \left\|x' - x\right\|_{\nabla^{-2}\psi(w_x)}. \qquad \text{(Hölder's inequality)}$$

Using Taylor expansion, for some ξ on the segment connecting w_x and $w_{x'}$, we have

$$\begin{aligned} F_{x'}(w_x) - F_{x'}(w_{x'}) &= \left\langle w_x - w'_x, \nabla F_{x'}(w_{x'})\right\rangle + \frac{1}{2}\|w_x - w_{x'}\|^2_{\nabla^2\psi(\xi)} \\ &\geq \frac{1}{2}\|w_x - w_{x'}\|^2_{\nabla^2\psi(\xi)} && \text{(first-order optimality of } w_{x'}) \\ &\geq \frac{c}{2}\|w_x - w_{x'}\|^2_{\nabla^2\psi(w_x)}. && \text{(condition of the lemma)} \end{aligned}$$

Combining we have, $\langle w_x - w_{x'}, x' - x\rangle \geq F_{x'}(w_x) - F_{x'}(w_{x'}) \geq c\|w_x - w_{x'}\|^2_{\nabla^2\psi(w_x)} \geq 0$, and also $\frac{c}{2}\|w_x - w_{x'}\|^2_{\nabla^2\psi(w_x)} \leq \|w_x - w_{x'}\|_{\nabla^2\psi(w_x)}\|x' - x\|_{\nabla^{-2}\psi(w_x)}$, which implies

$$\|w_x - w_{x'}\|_{\nabla^2\psi(w_x)} \leq \frac{2}{c}\left\|x - x'\right\|_{\nabla^{-2}\psi(w_x)}$$

and finishes the proof. ■

Lemma 17 (Multiplicative Stability) *Let $\Omega = \{w \in \Delta_d :\ w_i \geq b_i,\ \forall i \in [d]\}$ for some $b_i \in [0,1]$, $w' \in \Omega$ be such that $w'_i > 0$ for all $i \in [d]$, $w = \operatorname{argmin}_{w\in\Omega}\{\langle w, \ell\rangle + D_\psi(w, w')\}$ where $\psi(w) = \sum_{i=1}^d \frac{1}{\eta_i} w_i \ln w_i$, $|\ell_i| \leq c_{\max}$, and $\eta_i c_{\max} \leq \frac{1}{32}$ for all i and some $c_{\max} > 0$. Then $w_i \in [\frac{1}{\sqrt{2}}w'_i, \sqrt{2}w'_i]$.*

Proof Recall that $D_\psi(w, w') = \sum_i \frac{1}{\eta_i}\left(w_i \ln\frac{w_i}{w'_i} - w_i + w'_i\right)$. By the KKT condition of the optimization problem, we have for some λ and $\mu_i \geq 0$,

$$\ell_i + \frac{1}{\eta_i}\ln\frac{w_i}{w'_i} - \lambda - \mu_i = 0$$

and $\mu_i(w_i - b_i) = 0$ for all i. The above gives $w_i = w'_i \exp\left(\eta_i\left(-\ell_i + \lambda + \mu_i\right)\right)$. We now separately discuss two cases.

Case 1: $\min_i(\ell_i - \mu_i) \neq \max_i(\ell_i - \mu_i)$**.** In this case, we claim that $\min_i(\ell_i - \mu_i) < \lambda < \max_i(\ell_i - \mu_i)$. We prove it by contradiction: If $\lambda \geq \max_i(\ell_i - \mu_i)$, then

$$\sum_i w_i = \sum_i w'_i \exp\left(\eta_i\left(-\ell_i + \lambda + \mu_i\right)\right) > \sum_i w'_i = 1$$

contradicting with $w \in \Delta_d$ (the strict inequality is because there exists some j such that $\max_i(\ell_i - \mu_i) > (\ell_j - \mu_j)$ and $w'_j > 0$). We can derive a similar contradiction if $\lambda \leq \min_i(\ell_i - \mu_i)$. Thus, we conclude $\min_i(\ell_i - \mu_i) < \lambda < \max_i(\ell_i - \mu_i)$.

Our second claim is that for all i with $\mu_i \neq 0$, $\ell_i - \mu_i \geq \lambda$. Indeed, when $\mu_i \neq 0$, we have $b_i = w_i = w'_i \exp\left(\eta_i\left(-\ell_i + \lambda + \mu_i\right)\right)$. Clearly, $\exp\left(\eta_i\left(-\ell_i + \lambda + \mu_i\right)\right) \leq 1$ must hold; otherwise we have $w'_i < b_i$ which is a contradiction with $w' \in \Omega$. Therefore, $-\ell_i + \lambda + \mu_i \leq 0$.

Combining the above two claims, we see that $\min_i(\ell_i - \mu_i)$ must be equal to $\min_i \ell_i$; otherwise, we have $\min_i(\ell_i - \mu_i) < \min_i \ell_i$, which implies that there exists an j such that $\min_i(\ell_i - \mu_i) = \ell_j - \mu_j$ and $\mu_j > 0$. By the first claim, $\lambda > \ell_j - \mu_j$, and this contradicts with the second claim.

Thus, $\max_i(\ell_i - \mu_i) - \min_i(\ell_i - \mu_i) = \max_i(\ell_i - \mu_i) - \min_i \ell_i \leq \max_i \ell_i - \min_i \ell_i \leq 2c_{\max}$ (the inequality is by $\mu_i \geq 0$). Since both λ and $\ell_i - \mu_i$ are in the range $[\min_i(\ell_i - \mu_i), \max_i(\ell_i - \mu_i)]$, we have $|-\ell_i + \lambda + \mu_i| \leq \max_i(\ell_i - \mu_i) - \min_i(\ell_i - \mu_i) \leq 2c_{\max}$. By the condition on η_i, we then have $w_i \in \left[\exp(-\frac{1}{16})w_i', \exp(\frac{1}{16})w_i'\right] \subset \left[\frac{1}{\sqrt{2}}w_i', \sqrt{2}w_i'\right]$.

Case 2: $\min_i(\ell_i - \mu_i) = \max_i(\ell_i - \mu_i)$. In this case, it is clear that $\lambda = \ell_i - \mu_i$ must hold for all i to make w and w' both distributions. Thus, $w_{t,i} = w_{t,i}'$ for all i. ■

Appendix B. Omitted Details for Section 2

In this section, we provide the omitted proofs for Section 2.

B.1. Proof of Lemma 1

Proof By Lemma 15, we have (dropping one non-positive term)

$$\begin{aligned}
&\sum_{t=1}^{T} \langle w_t - u, \ell_t + a_t \rangle \\
&\leq \sum_{t=1}^{T} \left(D_{\psi_t}(u, w_t') - D_{\psi_t}(u, w_{t+1}') \right) + \sum_{t=1}^{T} \left(\langle w_t - w_{t+1}', \ell_t - m_t + a_t \rangle - D_{\psi_t}(w_{t+1}', w_t) \right).
\end{aligned} \tag{13}$$

For the first term, we reorder it and use $D_{\psi_t}(u, v) = \sum_{i=1}^{d} \frac{1}{\eta_{t,i}} f_{\mathrm{KL}}(u_i, v_i)$:

$$\begin{aligned}
&\sum_{t=1}^{T} \left(D_{\psi_t}(u, w_t') - D_{\psi_t}(u, w_{t+1}') \right) = D_{\psi_1}(u, w_1') + \sum_{t=2}^{T} \left(D_{\psi_t}(u, w_t') - D_{\psi_{t-1}}(u, w_t') \right) - D_{\psi_T}(u, w_{T+1}') \\
&\leq \sum_{i=1}^{d} \frac{1}{\eta_{1,i}} f_{\mathrm{KL}}(u_i, w_{1,i}') + \sum_{t=2}^{T} \sum_{i=1}^{d} \left(\frac{1}{\eta_{t,i}} - \frac{1}{\eta_{t-1,i}} \right) f_{\mathrm{KL}}(u_i, w_{t,i}').
\end{aligned}$$

For the second term, fix a particular t and define $w^\star = \operatorname{argmax}_{w \in \mathbb{R}_+^d} \langle w_t - w, \ell_t - m_t + a_t \rangle - D_{\psi_t}(w, w_t)$. By the optimality of $w^\star$, we have: $\ell_t - m_t + a_t = \nabla \psi_t(w_t) - \nabla \psi_t(w^\star)$ and thus $w_i^\star = w_{t,i} e^{-\eta_{t,i}(\ell_{t,i} - m_{t,i} + a_{t,i})}$. Therefore, we have

$$\begin{aligned}
&\left\langle w_t - w_{t+1}', \ell_t - m_t + a_t \right\rangle - D_{\psi_t}(w_{t+1}', w_t) \\
&\leq \langle w_t - w^\star, \ell_t - m_t + a_t \rangle - D_{\psi_t}(w^\star, w_t) \\
&= \langle w_t - w^\star, \nabla \psi_t(w_t) - \nabla \psi_t(w^\star) \rangle - D_{\psi_t}(w^\star, w_t) \\
&= D_{\psi_t}(w_t, w^\star) = \sum_{i=1}^{d} \frac{1}{\eta_{t,i}} \left(w_{t,i} \ln \frac{w_{t,i}}{w_i^\star} - w_{t,i} + w_i^\star \right) \\
&= \sum_{i=1}^{d} \frac{w_{t,i}}{\eta_{t,i}} \left(\eta_{t,i}(\ell_{t,i} - m_{t,i} + a_{t,i}) - 1 + e^{-\eta_{t,i}(\ell_{t,i} - m_{t,i} + a_{t,i})} \right)
\end{aligned}$$

$$\leq \sum_{i=1}^{d} \eta_{t,i} w_{t,i} \left(\ell_{t,i} - m_{t,i} + a_{t,i}\right)^2,$$

where in the last inequality we apply $e^{-x} - 1 + x \leq x^2$ for $x \geq -1$ and the condition of the lemma $\eta_{t,i}|\ell_{t,i} - m_{t,i}| \leq \frac{1}{32}$ such that $\eta_{t,i}|\ell_{t,i} - m_{t,i} + a_{t,i}| \leq \eta_{t,i}|\ell_{t,i} - m_{t,i}| + 32\eta_{t,i}^2(\ell_{t,i} - m_{t,i})^2 \leq \frac{1}{32} + \frac{32}{32^2} \leq \frac{1}{16}$. Using the definition of a_t and the condition $\eta_{t,i}|\ell_{t,i} - m_{t,i}| \leq \frac{1}{32}$ again, we also continue with

$$\begin{aligned}\left\langle w_t - w'_{t+1}, \ell_t - m_t + a_t \right\rangle - D_{\psi_t}(w'_{t+1}, w_t) &\leq \sum_{i=1}^{d} \eta_{t,i} w_{t,i} \left(\ell_{t,i} - m_{t,i} + 32\eta_{t,i}\left(\ell_{t,i} - m_{t,i}\right)^2\right)^2 \\ &\leq 4 \sum_{i=1}^{d} \eta_{t,i} w_{t,i} \left(\ell_{t,i} - m_{t,i}\right)^2.\end{aligned}$$

To sum up, combining everything, we have,

$$\begin{aligned}&\sum_{t=1}^{T} \langle w_t - u, \ell_t + a_t \rangle \\ &\leq \sum_{i=1}^{d} \frac{1}{\eta_{1,i}} f_{\text{KL}}(u_i, w'_{1,i}) + \sum_{t=2}^{T}\sum_{i=1}^{d} \left(\frac{1}{\eta_{t,i}} - \frac{1}{\eta_{t-1,i}}\right) f_{\text{KL}}(u_i, w'_{t,i}) + 4\sum_{t=1}^{T}\sum_{i=1}^{d} \eta_{t,i} w_{t,i}(\ell_{t,i} - m_{t,i})^2.\end{aligned}$$

Finally, moving $\sum_{t=1}^{T} \langle w_t - u, a_t \rangle$ to the right-hand side of the inequality and using the definition of a_t again finishes the proof. ■

B.2. Proof of Theorem 2

Proof To apply Lemma 1, we notice that the condition $32\eta_{t,i}|\ell_{t,i} - m_{t,i}| \leq 1$ of Lemma 1 holds trivially by the definition of $\eta_{t,i}$. Therefore, applying Eq. (1) with $u = (1-\frac{1}{T})e_{i_\star} + \frac{1}{T}w'_1 \in \bigcap_{t=1}^{T} \Omega_t$, we have:

$$\begin{aligned}\textsc{Reg}(e_{i_\star}) &= \textsc{Reg}(u) + \sum_{t=1}^{T} \langle u - e_{i_\star}, \ell_t \rangle \\ &= \textsc{Reg}(u) + \frac{1}{T}\sum_{t=1}^{T} \langle w'_1 - e_{i_\star}, \ell_t \rangle \\ &\leq \textsc{Reg}(u) + 2 \\ &\leq \sum_{i=1}^{d} \frac{1}{\eta_{1,i}} f_{\text{KL}}(u_i, w'_{1,i}) + \sum_{t=2}^{T}\sum_{i=1}^{d} \left(\frac{1}{\eta_{t,i}} - \frac{1}{\eta_{t-1,i}}\right) f_{\text{KL}}(u_i, w'_{t,i}) \\ &\qquad + 32\sum_{t=1}^{T}\sum_{i=1}^{d} \eta_{t,i} u_i(\ell_{t,i} - m_{t,i})^2 - 16\sum_{t=1}^{T}\sum_{i=1}^{d} \eta_{t,i} w_{t,i}(\ell_{t,i} - m_{t,i})^2 + 2. \quad (14)\end{aligned}$$

For the first term, note that $u_i \le w'_{1,i}$ when $i \ne i_\star$, and $\eta_{1,i} = \frac{1}{64}$. Thus,

$$\sum_{i=1}^{d} \frac{1}{\eta_{1,i}} f_{\mathrm{KL}}(u_i, w'_{1,i}) = \sum_{i=1}^{d} \frac{1}{\eta_{1,i}} \left(u_i \ln \frac{u_i}{w'_{1,i}} - u_i + w'_{1,i} \right) \le 64 u_{i_\star} \ln \frac{u_{i_\star}}{w'_{1,i_\star}} + \sum_{i=1}^{d} 64 \cdot \frac{1}{d} = \mathcal{O}(\ln d).$$

For the second term, we proceed as

$$\begin{aligned}
&\sum_{t=2}^{T} \sum_{i=1}^{d} \left(\frac{1}{\eta_{t,i}} - \frac{1}{\eta_{t-1,i}} \right) f_{\mathrm{KL}}(u_i, w'_{t,i}) \\
&= \sum_{t=2}^{T} \sum_{i=1}^{d} \left(\frac{1}{\eta_{t,i}} - \frac{1}{\eta_{t-1,i}} \right) \left(u_i \ln \frac{u_i}{w'_{t,i}} - u_i + w'_{t,i} \right) \\
&\le \sum_{t=2}^{T} \left(\frac{1}{\eta_{t,i_\star}} - \frac{1}{\eta_{t-1,i_\star}} \right) \left(u_{i_\star} \ln \frac{u_{i_\star}}{w'_{t,i_\star}} \right) + \sum_{t=2}^{T} \sum_{i=1}^{d} \left(\frac{1}{\eta_{t,i}} - \frac{1}{\eta_{t-1,i}} \right) w'_{t,i} \\
&\hspace{20em} \left(u_i = \tfrac{1}{dT} \le w'_{t,i} \text{ for } i \ne i_\star\right) \\
&= \sum_{t=2}^{T} \left(\frac{1}{\eta_{t,i_\star}} - \frac{1}{\eta_{t-1,i_\star}} \right) \left(u_{i_\star} \ln \frac{u_{i_\star}}{w'_{t,i_\star}} \right) + \sum_{t=2}^{T} \sum_{i=1}^{d} \frac{\frac{1}{\eta_{t,i}^2} - \frac{1}{\eta_{t-1,i}^2}}{\frac{1}{\eta_{t,i}} + \frac{1}{\eta_{t-1,i}}} w'_{t,i} \\
&\le \frac{\ln(dT)}{\eta_{T,i_\star}} + \sum_{t=2}^{T} \sum_{i=1}^{d} \eta_{t-1,i} \left(\frac{1}{\eta_{t,i}^2} - \frac{1}{\eta_{t-1,i}^2} \right) w'_{t,i} \qquad \left(u_{i_\star} \ln \tfrac{u_{i_\star}}{w'_{t,i_\star}} \le \ln(dT)\right) \\
&\le 64 \ln(dT) + \sqrt{\ln(dT) \sum_{t=1}^{T} (\ell_{t,i_\star} - m_{t,i_\star})^2} + \sum_{t=2}^{T} \sum_{i=1}^{d} \frac{1}{\ln(dT)} \eta_{t-1,i} w'_{t,i} (\ell_{t-1,i} - m_{t-1,i})^2 \\
&\hspace{20em} (\text{by the definition of } \eta_{t,i}) \\
&\le 64 \ln(dT) + \sqrt{\ln(dT) \sum_{t=1}^{T} (\ell_{t,i_\star} - m_{t,i_\star})^2} + \sum_{t=2}^{T} \sum_{i=1}^{d} 2 \eta_{t-1,i} w_{t-1,i} (\ell_{t-1,i} - m_{t-1,i})^2,
\end{aligned}$$

where the last step uses the fact $w'_{t,i} \le 2 w_{t-1,i}$ according to the multiplicative stability lemma Lemma 17 (which asserts $w'_{t,i} \in [\frac{1}{\sqrt{2}} w'_{t-1,i}, \sqrt{2} w'_{t-1,i}]$ and $w_{t-1,i} \in [\frac{1}{\sqrt{2}} w'_{t-1,i}, \sqrt{2} w'_{t-1,i}]$). Note that the last term is then canceled by the fourth term of Eq. (14). For the third term of Eq. (14), we have

$$\begin{aligned}
&\sum_{t=1}^{T} \sum_{i=1}^{d} \eta_{t,i} u_i (\ell_{t,i} - m_{t,i})^2 \\
&\le \sum_{t=1}^{T} \eta_{t,i_\star} (\ell_{t,i_\star} - m_{t,i_\star})^2 + \frac{1}{dT} \sum_{t=1}^{T} \sum_{i=1}^{d} \eta_{t,i} (\ell_{t,i} - m_{t,i})^2 \\
&\le \sum_{t=1}^{T} \sqrt{\frac{\ln(dT)}{\sum_{s<t} (\ell_{s,i_\star} - m_{s,i_\star})^2}} \cdot (\ell_{t,i_\star} - m_{t,i_\star})^2 + 1
\end{aligned}$$

$$\leq \mathcal{O}\left(\sqrt{\ln(dT)\sum_{t=1}^{T}(\ell_{t,i_\star} - m_{t,i_\star})^2} + 1\right).$$

Combining everything then proves

$$\textsc{Reg}(e_{i_\star}) = \mathcal{O}\left(\ln(dT) + \sqrt{\ln(dT)\sum_{t=1}^{T}(\ell_{t,i_\star} - m_{t,i_\star})^2}\right).$$

■

B.3. Proof of Theorem 3

The proof largely follows (Koolen et al., 2016), and thus for simplicity we only show it for the expectation results. We start from the regret bound:

$$\textsc{Reg}(e_{i_\star}) = \mathcal{O}\left(\ln(dT) + \sqrt{\ln(dT)\sum_{t=1}^{T}\langle w_t - e_{i_\star}, \ell_t - m'_t\rangle^2}\right).$$

For the first result, by the condition, we have

$$\mathbb{E}[\textsc{Reg}(e_{i_\star})] = \mathbb{E}\left[\sum_{t=1}^{T}\langle w_t - e_{i_\star}, \ell_t\rangle\right] = \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} w_{t,i}\mathbb{E}_t[\ell_{t,i} - \ell_{t,i_\star}]\right] \geq \Delta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\neq i_\star} w_{t,i}\right].$$

On the other hand,

$$\begin{aligned}\sum_{t=1}^{T}\langle w_t - e_{i_\star}, \ell_t - m'_t\rangle^2 &= \sum_{t=1}^{T}\left(\sum_{i\neq i_\star} w_{t,i}(\ell_{t,i} - m'_{t,i} - \ell_{t,i_\star} + m'_{t,i_\star})\right)^2 \\ &\leq \sum_{t=1}^{T}\sum_{i\neq i_\star} w_{t,i}\left(\ell_{t,i} - m'_{t,i} - \ell_{t,i_\star} + m'_{t,i_\star}\right)^2 \leq 16\sum_{t=1}^{T}\sum_{i\neq i_\star} w_{t,i}.\end{aligned}$$

Therefore,

$$\Delta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\neq i_\star} w_{t,i}\right] \leq \mathbb{E}[\textsc{Reg}(e_{i_\star})] = \mathcal{O}\left(\sqrt{\ln(dT)\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\neq i_\star} w_{t,i}\right]} + \ln(dT)\right).$$

Treating $\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\neq i_\star} w_{t,i}\right]$ as a variable and solving the inequality, we get $\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\neq i_\star} w_{t,i}\right] = \mathcal{O}\left(\ln(dT)/\Delta^2\right)$. Plugging this back we get $\mathbb{E}[\textsc{Reg}(e_{i_\star})] = \mathcal{O}\left(\frac{\ln(dT)}{\Delta}\right)$.

For the second result, first note that by Lemma 17 we have $w_{t,i} \in [\frac{1}{\sqrt{2}}w'_{t,i}, \sqrt{2}w'_{t,i}]$, $w'_{t,i} \in [\frac{1}{\sqrt{2}}w'_{t-1,i}, \sqrt{2}w'_{t-1,i}]$ and $w_{t-1,i} \in [\frac{1}{\sqrt{2}}w'_{t-1,i}, \sqrt{2}w'_{t-1,i}]$, which implies $w_{t,i} \in [\frac{1}{2\sqrt{2}}w_{t-1,i}, 2\sqrt{2}w_{t-1,i}]$. We then proceed as follows:

$$
\begin{aligned}
\mathbb{E}[\textsc{Reg}(e_{i_\star})] &= \mathcal{O}\left(\mathbb{E}\left[\sqrt{\ln(dT)\sum_{t=1}^{T}\langle w_t - e_{i_\star}, \ell_t - \ell_{t-1}\rangle^2}\right]\right) \\
&= \mathcal{O}\left(\mathbb{E}\left[\sqrt{\ln(dT)\sum_{t=1}^{T}\langle w_t - e_{i_\star}, \ell_t\rangle^2 + \langle w_t - e_{i_\star}, \ell_{t-1}\rangle^2}\right]\right) \\
&= \mathcal{O}\left(\mathbb{E}\left[\sqrt{\ln(dT)\sum_{t=1}^{T}\left(\sum_{i=1}^{d} w_{t,i}(\ell_{t,i} - \ell_{t,i_\star})\right)^2 + \left(\sum_{i=1}^{d} w_{t,i}(\ell_{t-1,i} - \ell_{t-1,i_\star})\right)^2}\right]\right) \\
&= \mathcal{O}\left(\mathbb{E}\left[\sqrt{\ln(dT)\sum_{t=1}^{T}\left(\sum_{i=1}^{d} w_{t,i}(\ell_{t,i} - \ell_{t,i_\star})\right)^2}\right]\right) && (w_{t,i} \le 2\sqrt{2}w_{t-1,i}) \\
&= \mathcal{O}\left(\sqrt{\ln(dT)\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} w_{t,i}(\ell_{t,i} - \ell_{t,i_\star})^2\right]}\right) \\
& && \text{(Jensen's and Cauchy-Schwarz inequality)} \\
&= \mathcal{O}\left(\sqrt{\frac{\ln(dT)}{\Delta}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} w_{t,i}\mathbb{E}_t[\ell_{t,i} - \ell_{t,i_\star}]^\kappa\right]}\right) && \text{(by the assumption)} \\
&= \mathcal{O}\left(\sqrt{\frac{\ln(dT)}{\Delta}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{d} w_{t,i}^{1-\kappa}\mathbb{E}_t[w_{t,i}(\ell_{t,i} - \ell_{t,i_\star})]^\kappa\right]}\right) \\
&= \mathcal{O}\left(\sqrt{\frac{\ln(dT)}{\Delta}\mathbb{E}\left[\left(\sum_{t=1}^{T}\sum_{i=1}^{d} w_{t,i}\right)^{1-\kappa}\left(\sum_{t=1}^{T}\sum_{i=1}^{d}\mathbb{E}_t[w_{t,i}(\ell_{t,i} - \ell_{t,i_\star})]\right)^\kappa\right]}\right) \\
& && \text{(Hölder's inequality)} \\
&= \mathcal{O}\left(\sqrt{\frac{\ln(dT)}{\Delta}T^{1-\kappa}\mathbb{E}[\textsc{Reg}(e_{i_\star})]^\kappa}\right). && \text{(Jensen's inequality)}
\end{aligned}
$$

Therefore, $\mathbb{E}[\textsc{Reg}(e_{i_\star})]^{1-\kappa/2} = \mathcal{O}\left(\sqrt{\frac{\ln(dT)}{\Delta}T^{1-\kappa}}\right)$, and $\mathbb{E}[\textsc{Reg}(e_{i_\star})] = \mathcal{O}\left(\left(\frac{\ln(dT)}{\Delta}\right)^{\frac{1}{2-\kappa}} T^{\frac{1-\kappa}{2-\kappa}}\right)$.

B.4. Proof of Theorem 4

Proof The regret $\textsc{Reg}(u) = \sum_{t=1}^{T}\langle w_t - u, \ell_t\rangle$ can be decomposed as the regret of base algorithm $k_\star$: $\textsc{Reg}_{\mathcal{A}_\star}(u) = \sum_{t=1}^{T}\langle w_t^{k_\star} - u, \ell_t\rangle$, plus the regret of the master to this base algorithm: $\sum_{t=1}^{T}\langle p_t - e_{k_\star}, g_t\rangle = \sum_{t=1}^{T}\langle w_t - w_t^{k_\star}, \ell_t\rangle$ (by the definition of w_t and g_t). It thus remains to apply the regret guarantee of MsMwC from Lemma 1 (with u in that lemma set to $e_{k_\star}$), since the conditions of the lemma hold by the fact $g_{t,k} - h_{t,k} = \langle w_t^k, \ell_t - m_t \rangle$. The first term in Eq. (1) becomes $\frac{1}{\eta_\star} \ln \frac{1}{p'_{1,k_\star}} + \sum_k \frac{p'_{1,k}}{\eta_k}$, which is $\frac{1}{\eta_\star} \ln \frac{\sum_k \eta_k^2}{\eta_\star^2} + \frac{\sum_k \eta_k}{\sum_k \eta_k^2}$ by the definition of p'_1. The second term in Eq. (1) is simply zero since the learning rate stays the same over time. The third term equals $32\eta_\star \sum_{t=1}^T \langle w_t^{k_\star}, \ell_t - m_t \rangle^2$. Dropping the last negative term then finishes the proof. ■

Appendix C. Omitted Details for Section 3

C.1. Proof of Theorem 5

Proof By the construction, for any u there exists $k_\star$ such that $\eta_{k_\star} \le \min\left\{\frac{1}{64}, \sqrt{\frac{\mathrm{KL}(u,\pi)+\ln V(u)}{V(u)}}\right\} \le 2\eta_{k_\star}$. Therefore, from Eq. (4) we have:

$$\begin{aligned}
\mathrm{REG}_{\mathcal{A}_{k_\star}}(u) &\le \frac{\mathrm{KL}(u,\pi)}{2\eta_{k_\star}} + 64\eta_{k_\star} \sum_{t=1}^T \sum_{i=1}^d u_i(\ell_{t,i} - m_{t,i})^2 - 32\eta_{k_\star} \sum_{t=1}^T \sum_{i=1}^d w_{t,i}^{k_\star}(\ell_{t,i} - m_{t,i})^2 \\
&\le \mathcal{O}\left(\mathrm{KL}(u,\pi) + \sqrt{(\mathrm{KL}(u,\pi) + \ln V(u))V(u)}\right) - 32\eta_{k_\star} \sum_{t=1}^T \sum_{i=1}^d w_{t,i}^{k_\star}(\ell_{t,i} - m_{t,i})^2. \qquad (15)
\end{aligned}$$

Further note that $32\eta_k \left|\langle w_t^k, \ell_t - m_t \rangle\right| \le 32\eta_k \|\ell_t - m_t\|_\infty \le 1$. Hence, we apply Theorem 4 with $\sum_k \eta_k = \Theta(1), \sum_k \eta_k^2 = \Theta(1), \frac{\sum_k \eta_k^2}{\eta_{k_\star}^2} = \mathcal{O}\left(1/\eta_{k_\star}^2\right) = \mathcal{O}\left(\frac{V(u)}{\mathrm{KL}(u,\pi)+\ln V(u)}\right) = \mathcal{O}(V(u))$, and cancel the last term in Eq. (2) by the last negative term in Eq. (15) via Cauchy-Schwarz inequality, arriving at

$$\begin{aligned}
\mathrm{REG}(u) &\le \mathcal{O}\left(\mathrm{KL}(u,\pi) + \sqrt{(\mathrm{KL}(u,\pi) + \ln V(u))V(u)}\right) + \frac{1}{\eta_{k_\star}} \ln \frac{\sum_k \eta_k^2}{\eta_{k_\star}^2} \\
&= \mathcal{O}\left(\mathrm{KL}(u,\pi) + \ln V(u) + \sqrt{(\mathrm{KL}(u,\pi) + \ln V(u))V(u)}\right)
\end{aligned}$$

and finishing the proof. ■

C.2. Proof of Theorem 6

Proof By the definition of $\mathcal{S}$, it is clear that $|\mathcal{S}|$ is at most $\mathcal{O}(d \log_2 T)$ so our algorithm is efficient. For any $i_\star \in [d]$, there exists a $k_\star$ such that $\eta_{k_\star} \le \min\left\{\frac{1}{128 c_{i_\star}}, \sqrt{\frac{\Gamma_{i_\star}}{\sum_{t=1}^T (\ell_{t,i_\star} - m_{t,i_\star})^2}}\right\} \le 2\eta_{k_\star}$. Moreover, $32 \cdot \eta_{t,i}|\ell_{t,i} - m_{t,i}| \le 128\eta_k c_i \le 1$ for all $i \in \mathcal{Z}(k)$. Hence, the conditions of Lemma 1 hold, and with $|\mathcal{Z}(k)| \le d$ we have

$$\begin{aligned}
\mathrm{REG}_{\mathcal{A}_{k_\star}}(e_{i_\star}) &\le \mathcal{O}\left(\frac{\ln d}{\eta_{k_\star}}\right) + 64\eta_{k_\star} \sum_{t=1}^T (\ell_{t,i_\star} - m_{t,i_\star})^2 - 32\eta_{k_\star} \sum_{t=1}^T \sum_{i=1}^d w_{t,i}(\ell_{t,i} - m_{t,i})^2 \\
&= \mathcal{O}\left(c_{i_\star}\Gamma_{i_\star} + \sqrt{\Gamma_{i_\star} \sum_{t=1}^T (\ell_{t,i_\star} - m_{t,i_\star})^2}\right) - 32\eta_{k_\star} \sum_{t=1}^T \sum_{i=1}^d w_{t,i}(\ell_{t,i} - m_{t,i})^2.
\end{aligned}$$

Next, also note that the conditions of Theorem 4 hold since

$$32\eta_k|\left\langle w_t^k, \ell_t - m_t \right\rangle| \le 64\eta_k \max_{i \in \mathcal{Z}(k)} c_i \le 1.$$

Thus, with the last negative term from the bound for $\textsc{Reg}_{\mathcal{A}_{k_\star}}(e_{i_\star})$ above canceling the last term of Eq. (2), and $\sum_k \eta_k = \Theta(1/c_{\min})$, $\sum_k \eta_k^2 = \Theta(1/c_{\min}^2)$, and $\frac{\sum_k \eta_k^2}{\eta_{k_\star}^2} = \mathcal{O}\left(\frac{c_{i_\star}^2 T}{c_{\min}^2}\right)$, we obtain:

$$\begin{aligned}
\textsc{Reg}(e_{i_\star}) &= \mathcal{O}\left(c_{i_\star}\Gamma_{i_\star} + \sqrt{\Gamma_{i_\star}\sum_{t=1}^T (\ell_{t,i_\star} - m_{t,i_\star})^2 + \frac{1}{\eta_{k_\star}}\Gamma_{i_\star} + c_{\min}}\right) \\
&= \mathcal{O}\left(c_{i_\star}\Gamma_{i_\star} + \sqrt{\Gamma_{i_\star}\sum_{t=1}^T (\ell_{t,i_\star} - m_{t,i_\star})^2}\right),
\end{aligned}$$

which completes the proof. ■

C.3. Proof of Theorem 7

Proof We first focus on a specific j and bound the regret within $\mathcal{I}_j$. The regret in this interval can be decomposed as

$$\begin{aligned}
\sum_{t \in \mathcal{I}_j} \langle w_t - u_j, \ell_t \rangle &= \sum_{t \in \mathcal{I}_j} \langle w_t - w_t^r, \ell_t \rangle + \sum_{t \in \mathcal{I}_j} \langle w_t^r - u_j, \ell_t \rangle \\
&= \sum_{t \in \mathcal{I}_j} \langle p_t - e_r, g_t \rangle + \sum_{t \in \mathcal{I}_j} \langle w_t^r - u_j, \ell_t \rangle \\
&\le \sum_{t \in \mathcal{I}_j} \langle p_t - \overline{e}_r, g_t \rangle + \sum_{t \in \mathcal{I}_j} \langle w_t^r - u_j, \ell_t \rangle + \mathcal{O}(1) \\
&\qquad\qquad \left(\text{define } \overline{e}_r = (1 - \tfrac{1}{T})e_r + \tfrac{\mathbf{1}}{\lceil \log_2 T \rceil T}\right)
\end{aligned}$$

for any $r \in [\lceil \log_2 T \rceil]$.

The term $\sum_{t \in \mathcal{I}_j} \langle w_t^r - u_j, \ell_t \rangle$ corresponds to the regret of the r-th base algorithm in the interval $\mathcal{I}_j$. Let s_j be the first time index in $\mathcal{I}_j$, and recall that the r-th expert is an MsMwC with a fixed learning rate $2\eta_r$, and a feasible set $\Omega_t = \{w \in \Delta_d : w_i \ge \frac{1}{dT}\}$. To upper bound it, we follow the exact same arguments as in the proof of Lemma 1, except for replacing the summation range $[1, T]$ with $\mathcal{I}_j$. This leads to:

$$\begin{aligned}
&\sum_{t \in \mathcal{I}_j} \langle w_t^r - u_j, \ell_t \rangle \\
&\le \frac{1}{2\eta_r}\sum_{i=1}^d f_{\mathrm{KL}}(u_{j,i}, w_{s_j,i}^{r\prime}) + 32\sum_{t \in \mathcal{I}_j}\sum_{i=1}^d 2\eta_r u_{j,i}(\ell_{t,i} - m_{t,i})^2 - 16\sum_{t \in \mathcal{I}_j}\sum_{i=1}^d 2\eta_r w_{t,i}^r(\ell_{t,i} - m_{t,i})^2 \\
&= \frac{1}{2\eta_r}\sum_{i=1}^d u_{j,i}\ln\frac{u_{j,i}}{w_{s_j,i}^{r\prime}} + 32\sum_{t \in \mathcal{I}_j}\sum_{i=1}^d 2\eta_r u_{j,i}(\ell_{t,i} - m_{t,i})^2 - 16\sum_{t \in \mathcal{I}_j}\sum_{i=1}^d 2\eta_r w_{t,i}^r(\ell_{t,i} - m_{t,i})^2
\end{aligned}$$

$$\leq \frac{1}{2\eta_r}\ln(dT) + 32\sum_{t\in\mathcal{I}_j}\sum_{i=1}^{d} 2\eta_r u_{j,i}(\ell_{t,i}-m_{t,i})^2 - 16\sum_{t\in\mathcal{I}_j}\sum_{i=1}^{d} 2\eta_r w^r_{t,i}(\ell_{t,i}-m_{t,i})^2.$$

Next, we deal with $\sum_{t\in\mathcal{I}_j}\langle p_t - \bar{e}_r, g_t\rangle$. Recall that MSMWC-MASTER uses a regularizer $\psi(p) = \sum_{k=1}^{\lceil \log_2 T\rceil} \frac{1}{\eta_k} p_k \ln p_k$. Again, similarly to the proof of Lemma 1, considering the regret only in $\mathcal{I}_j$ and dropping the negative term, we have

$$\begin{aligned}\sum_{t\in\mathcal{I}_j}\langle p_t - \bar{e}_r, g_t\rangle &\leq \sum_{t\in\mathcal{I}_j}\left(D_\psi(\bar{e}_r, p'_t) - D_\psi(\bar{e}_r, p'_{t+1})\right) + 32\sum_{t\in\mathcal{I}_j}\sum_{k=1}^{\lceil \log_2 T\rceil}\eta_k \bar{e}_{r,k}(g_{t,k}-h_{t,k})^2 \\ &\leq D_\psi(\bar{e}_r, p'_{s_j}) - D_\psi(\bar{e}_r, p'_{s_{j+1}}) + 32\eta_r\sum_{t\in\mathcal{I}_j}\langle w^r_t, \ell_t - m_t\rangle^2 + \mathcal{O}(1),\end{aligned}$$

where s_{j+1} is defined as $T+1$ if j is the last interval. We further deal with the first term above:

$$\begin{aligned}&D_\psi(\bar{e}_r, p'_{s_j}) - D_\psi(\bar{e}_r, p'_{s_{j+1}}) \\ &= \sum_{k=1}^{\lceil \log_2 T\rceil}\frac{1}{\eta_k}\left(\bar{e}_{r,k}\ln\frac{p'_{s_{j+1},k}}{p'_{s_j,k}} + p'_{s_j,k} - p'_{s_{j+1},k}\right) \\ &\leq \frac{\ln(\lceil \log_2 T\rceil T)}{\eta_r} + \sum_{k=1}^{\lceil \log_2 T\rceil}\frac{1}{\eta_k}\left(p'_{s_j,k} - p'_{s_{j+1},k}\right) + \mathcal{O}(\ln(dT)).\end{aligned}$$

Combining all bounds above, we get that for any $r \in \lceil \log_2 T\rceil$:

$$\begin{aligned}&\sum_{t\in\mathcal{I}_j}\langle w_t - u_j, \ell_t\rangle \\ &\leq \frac{1}{2\eta_r}\ln(dT) + 32\sum_{t\in\mathcal{I}_j}\sum_{i=1}^{d} 2\eta_r u_{j,i}(\ell_{t,i}-m_{t,i})^2 - 16\sum_{t\in\mathcal{I}_j}\sum_{i=1}^{d} 2\eta_r w^r_{t,i}(\ell_{t,i}-m_{t,i})^2 \\ &\qquad + \frac{\ln(\lceil \log_2 T\rceil T)}{\eta_r} + \sum_{k=1}^{\lceil \log_2 T\rceil}\frac{1}{\eta_k}\left(p'_{s_j,k} - p'_{s_{j+1},k}\right) + 32\eta_r\sum_{t\in\mathcal{I}_j}\langle w^r_t, \ell_t - m_t\rangle^2 + \mathcal{O}(\ln(dT)) \\ &\leq \mathcal{O}\left(\frac{\ln(dT)}{\eta_r} + \eta_r\sum_{t\in\mathcal{I}_j}\sum_{i=1}^{d} u_{j,i}(\ell_{t,i}-m_{t,i})^2\right) + \mathcal{O}(\ln(dT)) + \sum_{k=1}^{\lceil \log_2 T\rceil}\frac{1}{\eta_k}\left(p'_{s_j,k} - p'_{s_{j+1},k}\right)\end{aligned}$$

where we use Jenson's inequality: $\langle w^r_t, \ell_t - m_t\rangle^2 \leq \sum_{i=1}^d w^r_{t,i}(\ell_{t,i}-m_{t,i})^2$. Specifically, applying the above bound with the r such that

$$\eta_r \leq \min\left\{\frac{1}{64}, \sqrt{\frac{\ln(dT)}{\sum_{t\in\mathcal{I}_j}\sum_{i=1}^d u_{j,i}(\ell_{t,i}-m_{t,i})^2}}\right\} \leq 2\eta_r,$$

we get

$$\sum_{t\in\mathcal{I}_j}\langle w_t - u_j, \ell_t\rangle = \mathcal{O}\left(\sqrt{\ln(dT)\sum_{t\in\mathcal{I}_j}\sum_{i=1}^{d} u_{j,i}(\ell_{t,i}-m_{t,i})^2} + \ln(dT)\right) + \sum_{k=1}^{\lceil \log_2 T\rceil}\frac{1}{\eta_k}\left(p'_{s_j,k} - p'_{s_{j+1},k}\right). \tag{16}$$

Finally, summing the above bound over $j = 1, 2, \ldots, S$ and telescoping, we get

$$\sum_{j=1}^{S} \sum_{t \in \mathcal{I}_j} \langle w_t - u_j, \ell_t \rangle = \mathcal{O}\left(\sum_{j=1}^{S} \sqrt{\ln(dT) \sum_{t \in \mathcal{I}_j} \sum_{i=1}^{d} u_{j,i}(\ell_{t,i} - m_{t,i})^2 + S \ln(dT)} \right) + \sum_{k=1}^{\lceil \log_2 T \rceil} \frac{p'_{1,k}}{\eta_k}$$
$$= \mathcal{O}\left(\sum_{j=1}^{S} \sqrt{\ln(dT) \sum_{t \in \mathcal{I}_j} \sum_{i=1}^{d} u_{j,i}(\ell_{t,i} - m_{t,i})^2 + S \ln(dT)} \right),$$

finishing the proof. ■

Note that importantly, the last term in Eq. (16) only disappears (mostly) after summed over all intervals. As mentioned, getting an interval regret bound like Eq. (16) but without the last term is impossible, proven in the next section.

C.4. Impossible results for interval regret

Theorem 18 *For a two-expert problem with loss range $[-1, 1]$, it is impossible to achieve the following regret bound for all interval $\mathcal{I} \subseteq [1, T]$ and all comparators $i \in \{1, 2\}$ simultaneously:*

$$\sum_{t \in \mathcal{I}} \langle p_t - e_i, \ell_t \rangle = \tilde{\mathcal{O}}\left(\sqrt{\sum_{t \in \mathcal{I}} |\ell_{t,i}| + 1} \right).$$

Proof Consider an envinronment where the losses of Expert 1 is a deterministic value $\ell_{t,1} = 0$, and the losses of Expert 2 are i.i.d. chosen in each round according to the following:

$$\ell_{t,2} = \begin{cases} 1 & \text{with probability } \frac{1}{2} - \epsilon \\ -1 & \text{with probability } \frac{1}{2} + \epsilon \end{cases}$$

where $\epsilon = T^{-\frac{1}{5}}$. We assume that $\epsilon \leq \frac{1}{4}$ (which is equivalent to assuming $T \geq 4^5$). For simplicity, we call this distribution $\mathcal{D}$. Note that the expected loss of Expert 2 is -2ϵ, smaller than that of Expert 1. Therefore, in this environment, the expected regret of the learner during $[1, T]$ would be

$$\mathbb{E}[\text{REG}^{[1,T]}(e_2)] = 2\epsilon \mathbb{E}\left[\sum_{t=1}^{T} p_{t,1} \right].$$

Define $L = T^{\frac{3}{10}}$, and divide the whole horizon into $\frac{T}{L} = T^{\frac{7}{10}}$ intervals. Denote them as $\mathcal{I}_k = \{(k-1)L + 1, \ldots, kL\}$ for $k = 1, 2, \ldots, \frac{T}{L}$. Let

$$k^\star = \operatorname*{argmin}_k \mathbb{E}\left[\sum_{t \in \mathcal{I}_k} p_{t,1} \right].$$

That is, $k^\star$ is the interval where the learner would put least weight on Expert 1 in expectation. We then create another environment, where the loss of Expert 2 is same as the previous environment in interval $1, 2, \ldots, k^\star - 1$, but change to the following starting from interval $k^\star$:

$$\ell_{t,2} = \begin{cases} 1 & \text{with probability } \frac{1}{2} + \epsilon \\ -1 & \text{with probability } \frac{1}{2} - \epsilon \end{cases}$$

We call this distribution $\mathcal{D}'$. In this alternative environment, starting from interval $k^\star$, the best expert becomes Expert 1, and the expected interval regret of the learner is

$$\mathbb{E}'[\textsc{Reg}^{\mathcal{I}_{k^\star}}(e_1)] = 2\epsilon\mathbb{E}'\left[\sum_{t\in\mathcal{I}_{k^\star}} p_{t,2}\right] = 2\epsilon L - 2\epsilon\mathbb{E}'\left[\sum_{t\in\mathcal{I}_{k^\star}} p_{t,1}\right]. \tag{17}$$

where we use $\mathbb{E}'[\cdot]$ to denote the expectation under this alternative environment.

Below we denote the probability measure under the two environments as $\mathcal{P}$ and $\mathcal{P}'$ respectively. Since $p_{t,1}$ is a function of $\{\ell_\tau\}_{\tau=1}^{t-1}$, by standard arguments,

$$\begin{aligned}
&\mathbb{E}'\left[\sum_{t\in\mathcal{I}_{k^\star}} p_{t,1}\right] - \mathbb{E}\left[\sum_{t\in\mathcal{I}_{k^\star}} p_{t,1}\right] \\
&\leq L\left\|\mathcal{P}(\{\ell_\tau\}_{\tau=1,\ldots,k^\star L}) - \mathcal{P}'(\{\ell_\tau\}_{\tau=1,\ldots,k^\star L})\right\|_{\text{TV}} && (\|\cdot\|_{\text{TV}} \text{ is the total variation}) \\
&\leq \frac{L}{2}\sqrt{\text{KL}(\mathcal{P}(\{\ell_\tau\}_{\tau=1,\ldots,k^\star L}),\ \mathcal{P}'(\{\ell_\tau\}_{\tau=1,\ldots,k^\star L}))} && (\text{Pinsker's inequality}) \\
&= \frac{L}{2}\sqrt{L\text{KL}(\mathcal{D},\ \mathcal{D}')} \\
&= \frac{L^{\frac{3}{2}}}{2}\sqrt{\left(\frac{1}{2}+\epsilon\right)\ln\frac{\frac{1}{2}+\epsilon}{\frac{1}{2}-\epsilon} + \left(\frac{1}{2}-\epsilon\right)\ln\frac{\frac{1}{2}-\epsilon}{\frac{1}{2}+\epsilon}} \\
&\leq \frac{L^{\frac{3}{2}}}{2}\sqrt{2\epsilon\ln\frac{\frac{1}{2}+\epsilon}{\frac{1}{2}-\epsilon}} \leq \frac{L^{\frac{3}{2}}}{2}\sqrt{\frac{4\epsilon^2}{\frac{1}{2}-\epsilon}} \leq 2L^{\frac{3}{2}}\epsilon,
\end{aligned}$$

where we use $\ln(1+\alpha) \leq \alpha$ and $\epsilon \leq \frac{1}{4}$. Notice that $\frac{L}{T}\frac{1}{2\epsilon}\mathbb{E}[\textsc{Reg}^{[1,T]}(e_2)] = \frac{L}{T}\mathbb{E}\left[\sum_{t=1}^{T} p_{t,1}\right] \geq \mathbb{E}\left[\sum_{t\in\mathcal{I}_{k^\star}} p_{t,1}\right]$ by the definition of $k^\star$, and $\mathbb{E}'\left[\sum_{t\in\mathcal{I}_{k^\star}} p_{t,1}\right] = L - \frac{\mathbb{E}'[\textsc{Reg}^{\mathcal{I}_{k^\star}}(e_1)]}{2\epsilon}$ by Eq. (17). Using them in the above inequality, we get

$$L - \frac{\mathbb{E}'[\textsc{Reg}^{\mathcal{I}_{k^\star}}(e_1)]}{2\epsilon} - \frac{L}{2\epsilon T}\mathbb{E}[\textsc{Reg}^{[1,T]}(e_2)] \leq 2L^{\frac{3}{2}}\epsilon.$$

Using the values we choose, this is equivalent to

$$T^{\frac{3}{10}} - \frac{T^{\frac{2}{10}}}{2}\mathbb{E}'[\textsc{Reg}^{\mathcal{I}_{k^\star}}(e_1)] - \frac{1}{2T^{\frac{5}{10}}}\mathbb{E}[\textsc{Reg}^{[1,T]}(e_2)] \leq 2T^{\frac{1}{4}}.$$

When T is large enough, we see that either $\mathbb{E}[\textsc{Reg}^{[1,T]}(e_2)] \geq \Omega(T^{\frac{8}{10}})$ or $\mathbb{E}'[\textsc{Reg}^{\mathcal{I}_{k^\star}}(e_1)] \geq \Omega(T^{\frac{1}{10}})$. However, the desired bound $\sqrt{\sum_{t\in\mathcal{I}}|\ell_{t,i}|}$ is $\mathcal{O}(\sqrt{T})$ and $\mathcal{O}(1)$ in the two cases respectively. One of them must be violated, thus the desired bound is impossible. ■

Algorithm 3 MSMWC-MASTER with unknown loss range

Input: An expert set generator $\mathcal{E}$, initial scale B_0.
Initialize: $\widetilde{B} = B_0$, $\mathcal{A}$ as an instance of Algorithm 2 with input $\mathcal{E}(\widetilde{B})$.
for $t = 1, \ldots, T$ **do**
  Obtain decision w_t from $\mathcal{A}$, play w_t.
  Receive loss ℓ_t, and feed $\bar{\ell}_t = m_t + \frac{B_{t-1}}{B_t}(\ell_t - m_t)$ to $\mathcal{A}$, where $B_t = \max_{0 \le s \le t} \|\ell_s - m_s\|_\star$.
  if $B_t/\widetilde{B} > T$ **then**
    $\widetilde{B} = B_t$.
    Reset $\mathcal{A}$ as a new instance of Algorithm 2 with input $\mathcal{E}(\widetilde{B})$.

C.5. Omitted details for Section 3.4

Since some of the results in this section will be used later for OLO as well, we use $\|\cdot\|_\star$ to denote L_∞ norm in the context of an expert problem and L_2 norm in the context of an OLO problem.

We apply a variant of the techniques introduced in Cutkosky (2019a) to deal with unknown loss range. We start with an initial guess B_0 on the range of $\max_t \|\ell_t - m_t\|_\star$. Denote by $B_t = \max_{0 \le s \le t} \|\ell_s - m_s\|_\star$ the range of predicted error up to episode t, and $B = B_T$. We feed the following truncated loss to the algorithm in each episode:

$$\bar{\ell}_t = m_t + \frac{B_{t-1}}{B_t}(\ell_t - m_t). \tag{18}$$

Note that $\left\|\bar{\ell}_t - m_t\right\|_\star \le B_{t-1}$. Thus, the truncated loss allows the learner to assume that the range of predicted error in episode t is known at the beginning of this episode. Doing so already gives an algorithm that can deal with unknown loss range when B_T/B_0 is not too big. To further deal with arbitrary ratio B_t/B_0, we also incorporate a restarting scheme which is a simplified version of that in Mhammedi et al. (2019). The restarting scheme makes sure the learning rate can always be properly tuned and replace the potential $\ln(B_T/B_0)$ dependency by $\ln T$. We summarize ideas above as a new master algorithm in Algorithm 3, which requires an expert set generator $\mathcal{E}$ as input. The expert set generator $\mathcal{E}$ is a function that maps any initial guess B_0 to a set of (learning rate, base algorithm) pairs $\mathcal{E}(B_0)$.

To obtain data-dependent bound in expert problem with unknown range, it suffices to run Algorithm 3 with the following expert set generator:

$$\begin{aligned}\mathcal{E}_{\text{UR}}(B_0) = \Big\{(\eta_k, \mathcal{A}_k) : \forall k = 1, \ldots, N, \eta_k = \tfrac{1}{32B_0 \cdot 2^k}, \mathcal{A}_k \text{ is MSMWC with } w_1' = \pi,\\ \Omega_t = \Delta_d, \text{ and } \eta_{t,i} = 2\eta_k \text{ for all } t \text{ and } i\Big\},\end{aligned} \tag{19}$$

and $\Lambda_t = \Delta_{\mathcal{S}_{\text{UR}}(t)}$, where $\mathcal{S}_{\text{UR}}(t) = \{k : \frac{1}{32B_0 \cdot 2^k} \le \frac{1}{64B_{t-1}}\}$, and $N = \lceil \log_2(2T^2) \rceil$.

Proof [of Theorem 8] Define $\bar{V}(u) = \max\left\{3, \sum_{t=1}^T \sum_{i=1}^d u_i(\bar{\ell}_{t,i} - m_{t,i})^2\right\}$. We first show that the desired bound holds when there is no restart before episode T, that is, $\frac{B_{T-1}}{B_0} \le T$. In this case, $\frac{1}{\max\{64,T\}B_{T-1}} \ge \frac{1}{32B_0 \cdot 2^N}$. Hence, there exists $k_\star \in \mathcal{E}_{\text{UR}}(B_0)$ such that

$$\eta_{k_\star} \le \min\left\{\frac{1}{64B_{T-1}}, \sqrt{\frac{\text{KL}(u,\pi) + \ln T}{\bar{V}(u)}}\right\} \le 2\eta_{k_\star}.$$

The conditions of Lemma 1 hold since $32 \cdot 2\eta_{k_\star} \left\|\bar{\ell}_t - m_t\right\|_\infty \le 64\eta_{k_\star} B_{T-1} \le 1$ for any $t \le T$. We thus have

$$\sum_{t=1}^{T}\left\langle w_t^{k_\star} - u, \bar{\ell}_t\right\rangle = \frac{\mathrm{KL}(u,\pi)}{2\eta_{k_\star}} + 64\eta_{k_\star}\sum_{t=1}^{T}\sum_{i=1}^{d} u_i(\bar{\ell}_{t,i} - m_{t,i})^2 - 32\eta_{k_\star}\sum_{t=1}^{T}\sum_{i=1}^{d} w_{t,i}^{k_\star}(\bar{\ell}_{t,i} - m_{t,i})^2$$
$$= \mathcal{O}\left(\sqrt{(\mathrm{KL}(u,\pi) + \ln T)\bar{V}(u)} + B\mathrm{KL}(u,\pi)\right) - 32\eta_{k_\star}\sum_{t=1}^{T}\sum_{i=1}^{d} w_{t,i}^{k_\star}(\bar{\ell}_{t,i} - m_{t,i})^2.$$

Note that $k_\star \in \mathcal{S}_{\mathrm{UR}}(T)$, and for any $k \in \mathcal{S}_{\mathrm{UR}}(t)$,

$$32\eta_k\left|\left\langle w_t^k, \bar{\ell}_t - m_t\right\rangle\right| \le 32\eta_k\left\|\bar{\ell}_t - m_t\right\|_\infty \le 1.$$

Hence, the conditions of Theorem 4 also hold, and with $\sum_k \eta_k = \Theta(\frac{1}{B_0}), \sum_k \eta_k^2 = \Theta(\frac{1}{B_0^2})$, $\frac{\sum_k \eta_k^2}{\eta_{k_\star}^2} = \mathcal{O}\left((\eta_1/\eta_{k_\star})^2\right) = \mathcal{O}\left(\ln T\right)$ by $2^N = \mathcal{O}(T^2)$ and $\left\langle w_t^{k_\star}, \bar{\ell}_t - m_t\right\rangle^2 \le \sum_{i=1}^{d} w_{t,i}^{k_\star}(\bar{\ell}_{t,i} - m_{t,i})^2$, we have:

$$\sum_{t=1}^{T}\left\langle w_t - u, \bar{\ell}_t\right\rangle \le \sum_{t=1}^{T}\left\langle w_t^{k_\star} - u, \bar{\ell}_t\right\rangle + \mathcal{O}\left(\frac{1}{\eta_{k_\star}}\ln T + B_0\right) + 32\eta_{k_\star}\sum_{t=1}^{T}\left\langle w_t^{k_\star}, \bar{\ell}_t - m_t\right\rangle^2$$
$$\le \mathcal{O}\left(\sqrt{(\mathrm{KL}(u,\pi) + \ln T)\bar{V}(u)}\right) + B(\mathrm{KL}(u,\pi) + \ln T).$$

Moreover, since $\ell_t - \bar{\ell}_t = \frac{B_t - B_{t-1}}{B_t}(\ell_t - m_t)$, the difference between the regret measured with ℓ_t and that with $\bar{\ell}_t$ is

$$\sum_{t=1}^{T}\left\langle w_t - u, \ell_t - \bar{\ell}_t\right\rangle \le 2\sum_{t=1}^{T}\left\|\ell_t - \bar{\ell}_t\right\|_\infty \le 2\sum_{t=1}^{T}\frac{B_t - B_{t-1}}{B_t}\left\|\ell_t - m_t\right\|_\infty$$
$$\le 2\sum_{t=1}^{T}(B_t - B_{t-1}) \le 2B.$$

Therefore, noticing $\bar{V}(u) \le V(u)$, we prove the desired result when there is no restart:

$$\sum_{t=1}^{T}\left\langle w_t - u, \ell_t\right\rangle = \sum_{t=1}^{T}\left\langle w_t - u, \bar{\ell}_t\right\rangle + \sum_{t=1}^{T}\left\langle w_t - u, \ell_t - \bar{\ell}_t\right\rangle$$
$$= \mathcal{O}\left(\sqrt{(\mathrm{KL}(u,\pi) + \ln T)\bar{V}(u)} + B(\mathrm{KL}(u,\pi) + \ln T)\right).$$

Next, we show that the desired bound also holds when there is at least one restart before episode T. Denote by τ_2 the episode of the last restart, and by τ_1 the episode of the second last restart ($\tau_1 = 0$ if the algorithm only restarts once). We consider the regret in the following three intervals: $[1,\tau_1]$, $(\tau_1,\tau_2]$, $(\tau_2,T]$. Denote $y_t = \left\|\ell_t - m_t\right\|_\infty$. For regret in $[1,\tau_1]$, we have:

$$\text{REG}^{[1,\tau_1]}(u) \le \sum_{t=1}^{\tau_1} y_t = B_{\tau_1}\sum_{t=1}^{\tau_1}\frac{y_t}{B_{\tau_1}} \le B_{\tau_1}T < B_{\tau_1}\frac{B_{\tau_2}}{B_{\tau_1}} \le B_T, \tag{20}$$

Algorithm 4 A Variant of Online Newton Step

Parameters: learning rate $\eta > 0$, $w'_1 = \mathbf{0}$.

Define: $c_t(w) = \langle w, \ell_t \rangle + 16\eta \langle w, \ell_t - m_t \rangle^2$ and $\nabla_t = \nabla c_t(w_t) = \ell_t + 32\eta \langle w_t, \ell_t - m_t \rangle (\ell_t - m_t)$.

for $t = 1, \ldots, T$ **do**

Receive prediction m_t and range hint z_t.

Update $w_t = \text{argmin}_{w \in \Omega} \{\langle w, m_t \rangle + D_{\psi_t}(w, w'_t)\}$ where $\psi_t(w) = \frac{1}{2} \|w\|^2_{A_t}$ and

$$A_t = \eta \left(4z_1^2 \cdot I + \sum_{s=1}^{t-1} (\nabla_s - m_s)(\nabla_s - m_s)^\top + 4z_t^2 \cdot I \right).$$

Receive ℓ_t.

Update $w'_{t+1} = \text{argmin}_{w \in \Omega} \{\langle w, \nabla_t \rangle + D_{\psi_t}(w, w'_t)\}$.

where we apply $T < B_{\tau_2}/B_{\tau_1}$ due to the restart condition. Within intervals $(\tau_1, \tau_2]$ and $(\tau_2, T]$, we have $B_{\tau_2 - 1}/B_{\tau_1} \leq T, B_{T-1}/B_{\tau_2} \leq T$. Thus, by the regret guarantee with restart only at the end of an interval,

$$\begin{aligned}
\text{REG}^{(\tau_1, \tau_2]}(u) &= \mathcal{O}\left(B_{\tau_2}(\text{KL}(u, \pi) + \ln T) + \sqrt{(\text{KL}(u, \pi) + \ln T) V^{(\tau_1, \tau_2]}(u)} \right) \\
\text{REG}^{(\tau_2, T]}(u) &= \mathcal{O}\left(B_T(\text{KL}(u, \pi) + \ln T) + \sqrt{(\text{KL}(u, \pi) + \ln T) V^{(\tau_2, T]}(u)} \right),
\end{aligned}$$

where $V^{\mathcal{I}}(u) = \max\left\{3, \sum_{t \in \mathcal{I}} \sum_{i=1}^d u_i (\ell_{t,i} - m_{t,i})^2 \right\}$. Summing all three regret bounds together and applying the Cauchy-Schwarz inequality, we get the desired result. ■

Appendix D. Omitted Details for Section 4

In this section, when presenting the base algorithms, we sometimes use Ω as its decision set, which should be seen as a subset of $\mathcal{K}$ (thus its size is bounded by D as well) and will be set appropriately by the master.

D.1. Combining Online Newton Step

We first introduce a variant of the ONS algorithm (Algorithm 4) and present its regret guarantee. To make the algorithm general enough to deal with unknown loss range in Appendix D.5, we consider a slightly more general setup where the algorithm receives a *range hint* z_t at the beginning of round t, which is guaranteed to satisfy $\|\ell_t - m_t\| \leq z_t$. For this section and the result of Theorem 9, it suffices to set $z_t = 1$ for all t. The guarantee of this ONS variant is as follows.

Lemma 19 *Suppose $\|\ell_t - m_t\|_2 \leq z_t$, z_t is non-decreasing in t, and $64\eta D z_T \leq 1$. Then Algorithm 4 ensures for any $u \in \Omega$ (with r being the rank of $\mathcal{L}_T = \sum_{t=1}^T (\ell_t - m_t)(\ell_t - m_t)^\top$)*

$$\sum_{t=1}^T \langle w_t - u, \ell_t \rangle$$

$$\leq \mathcal{O}\left(\frac{r\ln(Tz_T/z_1)}{\eta} + z_1\|u\|_2 + D(z_T - z_1) + \eta\sum_{t=1}^{T}\langle u, \ell_t - m_t\rangle^2\right) - 11\eta\sum_{t=1}^{T}\langle w_t, \ell_t - m_t\rangle^2.$$

Proof By Lemma 15 and Lemma 20, we have:

$$\begin{aligned}\sum_{t=1}^{T}\langle w_t - u, \nabla_t\rangle &\leq \sum_{t=1}^{T}\left\langle w_t - w'_{t+1}, \nabla_t - m_t\right\rangle + D_{\psi_t}(u, w'_t) - D_{\psi_t}(u, w'_{t+1}) \\ &\leq 2\sum_{t=1}^{T}\|\nabla_t - m_t\|^2_{A_t^{-1}} + D_{\psi_1}(u, w'_1) + \sum_{t=1}^{T-1}D_{\psi_{t+1}}(u, w'_{t+1}) - D_{\psi_t}(u, w'_{t+1}) \\ &\leq \mathcal{O}\left(\frac{r\ln(Tz_T/z_1)}{\eta} + \eta z_1^2\|u\|_2^2\right) + \sum_{t=1}^{T-1}D_{\psi_{t+1}}(u, w'_{t+1}) - D_{\psi_t}(u, w'_{t+1}).\end{aligned}$$

Note that $\eta z_1^2\|u\|_2^2 \leq \eta D z_T z_1\|u\|_2 = \mathcal{O}(z_1\|u\|_2)$. Moreover,

$$\begin{aligned}&\sum_{t=1}^{T-1}D_{\psi_{t+1}}(u, w'_{t+1}) - D_{\psi_t}(u, w'_{t+1}) \\ &\leq \frac{\eta}{2}\sum_{t=1}^{T-1}\left\langle u - w'_{t+1}, \nabla_t - m_t\right\rangle^2 + \mathcal{O}\left(\eta D^2\sum_{t=1}^{T-1}(z_{t+1}^2 - z_t^2)\right) \\ &\leq \eta\sum_{t=1}^{T-1}\langle u - w_t, \nabla_t - m_t\rangle^2 + \eta\sum_{t=1}^{T-1}\left\langle w_t - w'_{t+1}, \nabla_t - m_t\right\rangle^2 + \mathcal{O}\left(\eta D^2 z_T\sum_{t=1}^{T-1}(z_{t+1} - z_t)\right) \\ &\leq 2\eta\sum_{t=1}^{T}\langle u, \nabla_t - m_t\rangle^2 + 2\eta\sum_{t=1}^{T}\langle w_t, \nabla_t - m_t\rangle^2 + \mathcal{O}\left(\frac{r\ln(Tz_T/z_1)}{\eta}\right) + \mathcal{O}\left(D(z_T - z_1)\right) \\ &\qquad\qquad \text{(by } 0 \leq \eta\left\langle w_t - w'_{t+1}, \nabla_t - m_t\right\rangle \leq 3\eta D z_t = \mathcal{O}(1) \text{ and Lemma 20)} \\ &\leq 5\eta\sum_{t=1}^{T}\langle u, \ell_t - m_t\rangle^2 + 5\eta\sum_{t=1}^{T}\langle w_t, \ell_t - m_t\rangle^2 + \mathcal{O}\left(\frac{r\ln(Tz_T/z_1)}{\eta}\right) + \mathcal{O}\left(D(z_T - z_1)\right). \\ &\qquad\qquad \text{(by definition of } \nabla_t \text{ and } 32\eta|\langle w_t, \ell_t - m_t\rangle| \leq 32\eta D z_t \leq \tfrac{1}{2})\end{aligned}$$

Since $c_t(w)$ is convex in w, we have

$$\sum_{t=1}^{T}c_t(w_t) - c_t(u) = \sum_{t=1}^{T}\langle w_t - u, \ell_t\rangle + 16\eta\sum_{t=1}^{T}\langle w_t, \ell_t - m_t\rangle^2 - \langle u, \ell_t - m_t\rangle^2 \leq \sum_{t=1}^{T}\langle w_t - u, \nabla_t\rangle.$$

Reorganizing terms then finishes the proof. ■

Lemma 20 *In Algorithm 4, we have* $0 \leq \left\langle w_t - w'_{t+1}, \nabla_t - m_t\right\rangle \leq 2\|\nabla_t - m_t\|^2_{A_t^{-1}}$ *and also* $\sum_{t=1}^{T}\|\nabla_t - m_t\|^2_{A_t^{-1}} = \mathcal{O}\left(\frac{r\ln(Tz_T/z_1)}{\eta}\right)$.

Proof For any t, define $F_x(w) = \langle w, x\rangle + D_{\psi_t}(w, w'_t)$. Then, we have

$$w_t = \operatorname*{argmin}_{w\in\mathcal{K}} F_{m_t}(w) \quad \text{and} \quad w'_{t+1} = \operatorname*{argmin}_{w\in\mathcal{K}} F_{\nabla_t}(w).$$

Moreover, $\nabla^2_w D_{\psi_t}(w, w'_t) = A_t$ is a constant matrix. Hence, by Lemma 16 with $c = 1$, $0 \le \left\langle w_t - w'_{t+1}, \nabla_t - m_t\right\rangle \le 2\left\|\nabla_t - m_t\right\|^2_{A_t^{-1}}$.

Define $A'_t = \eta\left(4z_1^2 \cdot I + \sum_{s=1}^t (\nabla_s - m_s)(\nabla_s - m_s)^\top\right)$. Note that $\left\|\nabla_t - m_t\right\|_2^2 \le 4\left\|\ell_t - m_t\right\|_2^2 \le 4z_t^2$. Thus, $A_t \succcurlyeq A'_t$. By similar arguments in (Koren and Livni, 2017, Lemma 6), we have

$$\begin{aligned}
\sum_{t=1}^T \left\|\nabla_t - m_t\right\|^2_{A_t^{-1}} &= \frac{1}{\eta}\sum_{t=1}^T \operatorname{tr}\left(A_t^{-1}(A'_t - A'_{t-1})\right) \le \frac{1}{\eta}\sum_{t=1}^T \operatorname{tr}\left((A'_t)^{-1}(A'_t - A'_{t-1})\right) \\
&\le \frac{1}{\eta}\ln\frac{|A'_T|}{|A'_0|} = \frac{1}{\eta}\ln\left|I + \sum_{t=1}^T \frac{(\nabla_t - m_t)(\nabla_t - m_t)^\top}{4z_1^2}\right| \\
&= \mathcal{O}\left(\frac{r\ln\left(1 + \sum_{t=1}^T \frac{\|\ell_t - m_t\|_2^2}{rz_1^2}\right)}{\eta}\right) = \mathcal{O}\left(\frac{r\ln(Tz_T/z_1)}{\eta}\right),
\end{aligned}$$

where r is rank of $\sum_{t=1}^T (\ell_t - m_t)(\ell_t - m_t)^\top$. ■

To obtain the regret bound in Theorem 9, we instantiate MsMwC-Master with the following set of experts:

$$\begin{aligned}
\mathcal{E}_{\text{ONS}} = \Big\{(\eta_k, \mathcal{A}_k) : \forall k = (d_k, s_k) \in \{-\lceil\log_2(dT)\rceil, \ldots, \lceil\log_2 D\rceil\} \times [\lceil\log_2 T\rceil], \eta_k = \tfrac{1}{64\cdot 2^{d_k + s_k}}, \\
\mathcal{A}_k \text{ is Algorithm 4 with } z_t = 1 \text{ for all } t, \Omega = \mathcal{K}\cap\{w : \|w\|_2 \le 2^{d_k}\}, \text{ and } \eta = 3\eta_k\Big\}
\end{aligned} \tag{21}$$

Proof [of Theorem 9] We first assume $\|u\|_2 > \frac{1}{dT}$, and thus there exists $k_\star$ such that $\eta_{k_\star} \le \min\left\{\frac{1}{192\cdot 2^{d_{k_\star}}}, \sqrt{\frac{r\ln T}{\sum_{t=1}^T \langle u, \ell_t - m_t\rangle^2}}\right\} \le 2\eta_{k_\star}$, and $2^{d_{k_\star}-1} \le \|u\|_2 \le 2^{d_{k_\star}}$. Then by Lemma 19 with $64\cdot 3\eta_{k_\star}\cdot 2^{d_{k_\star}} \le 1$:

$$\begin{aligned}
\sum_{t=1}^T \left\langle w_t^{k_\star} - u, \ell_t\right\rangle &\le \mathcal{O}\left(\frac{r\ln T}{\eta_{k_\star}} + \|u\|_2 + \eta_{k_\star}\sum_{t=1}^T \langle u, \ell_t - m_t\rangle^2\right) - 33\eta_{k_\star}\sum_{t=1}^T \left\langle w_t^{k_\star}, \ell_t - m_t\right\rangle^2 \\
&= \tilde{\mathcal{O}}\left(r\|u\|_2 + \sqrt{r\sum_{t=1}^T \langle u, \ell_t - m_t\rangle^2}\right) - 33\eta_{k_\star}\sum_{t=1}^T \left\langle w_t^{k_\star}, \ell_t - m_t\right\rangle^2.
\end{aligned}$$

Next, by Theorem 4 with $32\eta_k \left|\left\langle w_t^k, \ell_t - m_t \right\rangle\right| \leq 32\eta_k \left\|w_t^k\right\|_2 \leq 1$, $\sum_k \eta_k = \Theta(dT)$, $\sum_k \eta_k^2 = \Theta(d^2T^2)$, and $\frac{\sum_k \eta_k^2}{\eta_{k_\star}^2} = \mathcal{O}(d^2T^2/\eta_{k_\star}^2) = \mathcal{O}(d^2D^2T^4)$, we have:

$$\sum_{t=1}^{T} \langle w_t - u, \ell_t \rangle = \tilde{\mathcal{O}}\left(r\left\|u\right\|_2 + \sqrt{r\sum_{t=1}^{T} \langle u, \ell_t - m_t \rangle^2 + \frac{1}{\eta_{k_\star}}} \right)$$
$$= \tilde{\mathcal{O}}\left(r\left\|u\right\|_2 + \sqrt{r\sum_{t=1}^{T} \langle u, \ell_t - m_t \rangle^2} \right).$$

When $\|u\|_2 \leq \frac{1}{dT} \leq D$ (if $D < \frac{1}{dT}$, we achieve constant regret by picking w_t arbitrarily), pick any $u' \in \mathcal{K}$ such that $\|u'\|_2 = \frac{1}{dT}$ (this is possible since $\mathbf{0} \in \mathcal{K}$). Then:

$$\sum_{t=1}^{T} \langle w_t - u, \ell_t \rangle = \sum_{t=1}^{T} \left\langle w_t - u', \ell_t \right\rangle + \sum_{t=1}^{T} \left\langle u' - u, \ell_t \right\rangle$$
$$\leq \tilde{\mathcal{O}}\left(r\left\|u'\right\|_2 + \sqrt{r\sum_{t=1}^{T} \langle u', \ell_t - m_t \rangle^2} + \left\|u'\right\|_2 T \right) = \tilde{\mathcal{O}}(1).$$

This finishes the proof. ■

D.2. Combining Gradient Descent

For gradient descent type of bound, we use the optimistic gradient descent algorithm (OptGD) as the base algorithm, which achieves the following regret bound with learning rate η (see (Rakhlin and Sridharan, 2013b, Lemma 3)):

$$\sum_{t=1}^{T} \langle w_t - u, \ell_t \rangle \leq \frac{\|u\|_2^2}{\eta} + \eta \sum_{t=1}^{T} \|\ell_t - m_t\|_2^2 .$$

To obtain the regret bound in Theorem 10, it suffices to instantiate MSMWC-MASTER with the following set of experts:

$$\mathcal{E}_{\text{GD}} = \Big\{ (\eta_k, \mathcal{A}_k) : \forall k = (d_k, s_k) \in \{-\lceil \log_2 T \rceil, \dots, \lceil \log_2 D \rceil\} \times [\lceil \log_2 T \rceil], \eta_k = \tfrac{1}{32 \cdot 2^{d_k + s_k}},$$
$$\mathcal{A}_k \text{ is OptGD with decision set } \Omega = \mathcal{K} \cap \{w : \|w\|_2 \leq 2^{d_k}\}, \text{ and } \eta = 4^{d_k}\eta_k \Big\}. \tag{22}$$

Proof [of Theorem 10] We first assume $\|u\|_2 > \frac{1}{T}$, so that there exists $k_\star$ such that

$$\eta_{k_\star} \leq \min\left\{ \frac{1}{64 \cdot 2^{d_{k_\star}}}, \frac{1}{\|u\|_2 \sqrt{\sum_{t=1}^{T} \|\ell_t - m_t\|_2^2}} \right\} \leq 2\eta_{k_\star},$$

and $2^{d_{k_\star}-1} \le \|u\|_2 \le 2^{d_{k_\star}}$. By the regret guarantee of OptGD, we have:

$$\sum_{t=1}^{T}\left\langle w_t^{k_\star}-u,\ell_t\right\rangle \le \frac{\|u\|_2^2}{4^{d_{k_\star}}\eta_{k_\star}}+4^{d_{k_\star}}\eta_{k_\star}\sum_{t=1}^{T}\|\ell_t-m_t\|_2^2=\mathcal{O}\left(\|u\|_2+\|u\|_2\sqrt{\sum_{t=1}^{T}\|\ell_t-m_t\|_2^2}\right).$$

Next, by Theorem 4 with $32\eta_k\left|\langle w_t^k,\ell_t-m_t\rangle\right| \le 32\eta_k\left\|w_t^k\right\|_2 \le 1$, $\sum_k\eta_k=\Theta(T)$, $\sum_k\eta_k^2=\Theta(T^2)$, and $\frac{\sum_k\eta_k^2}{\eta_{k_\star}^2}=\mathcal{O}(D^2T^3)$, we have:

$$\begin{aligned}\sum_{t=1}^{T}\langle w_t-u,\ell_t\rangle &= \tilde{\mathcal{O}}\left(\|u\|_2+\|u\|_2\sqrt{\sum_{t=1}^{T}\|\ell_t-m_t\|_2^2+\eta_{k_\star}\sum_{t=1}^{T}\left\langle w_t^{k_\star},\ell_t-m_t\right\rangle^2}\right)\\ &= \tilde{\mathcal{O}}\left(\|u\|_2+\|u\|_2\sqrt{\sum_{t=1}^{T}\|\ell_t-m_t\|_2^2}\right).\end{aligned}$$

When $\|u\|_2 \le \frac{1}{T}$, pick any $u' \in \mathcal{K}$ such that $\|u'\|_2=\frac{1}{T}$, then:

$$\begin{aligned}\sum_{t=1}^{T}\langle w_t-u,\ell_t\rangle &= \sum_{t=1}^{T}\langle w_t-u',\ell_t\rangle+\sum_{t=1}^{T}\langle u'-u,\ell_t\rangle\\ &\le \tilde{\mathcal{O}}\left(\|u\|_2+\|u\|_2\sqrt{\sum_{t=1}^{T}\|\ell_t-m_t\|_2^2}+\|u\|_2\right)=\tilde{\mathcal{O}}(1).\end{aligned}$$

This finishes the proof. ∎

D.3. Combining AdaGrad

We first introduce the base algorithm Algorithm 5, which is a variant of the AdaGrad algorithm with predictor m_t incorporated. It guarantees the following.

Theorem 21 *Define $A'_t=(I+\sum_{s=1}^{t}(\ell_s-m_s)(\ell_s-m_s)^\top)^{1/2}$. Assume $64\eta'|\langle w_t,\ell_t-m_t\rangle|\le 1$ for all t, and $\eta'\le\eta/\|u\|_{A'_T}^2$. Algorithm 5 ensures for any $u\in\Omega$,*

$$\sum_{t=1}^{T}\langle w_t-u,\ell_t\rangle=\mathcal{O}\left(\eta\mathrm{tr}\left(\mathcal{L}_T^{1/2}\right)+\frac{u^\top(I+\mathcal{L}_T)^{1/2}u}{\eta}\right)-16\eta'\sum_{t=1}^{T}\langle w_t,\ell_t-m_t\rangle^2.$$

Proof For any t, define $F_x(w)=\langle w,x\rangle+\psi_{t-1}(w)$. Note that $w_t=\mathrm{argmin}_{w\in\mathcal{K}}F_{\sum_{s=1}^{t-1}\nabla_s+m_t}(w)$, and denote $w'_t=\mathrm{argmin}_{w\in\mathcal{K}}F_{\sum_{s=1}^{t}\nabla_s}(w)$. Moreover, $\nabla^2\psi_{t-1}(w)=A_{t-1}$ is a constant matrix. Hence, by Lemma 16 with $c=1$, $\langle w_t-w'_t,\nabla_t-m_t\rangle\le 2\|\nabla_t-m_t\|_{A_{t-1}^{-1}}^2$, and for any $u\in\Omega$ we have:

$$\sum_{t=1}^{T}\langle w_t-u,\nabla_t\rangle=\sum_{t=1}^{T}\langle w_t-w'_t,\nabla_t-m_t\rangle+\langle w_t-w'_t,m_t\rangle+\langle w'_t-u,\nabla_t\rangle$$

Algorithm 5 Optimistic AdaGrad

Parameters: learning rate $\eta, \eta' > 0$, $w'_1 = \mathbf{0}$.
Define:

$$
\begin{aligned}
c_t(w) &= \langle w, \ell_t \rangle + 16\eta' \langle w, \ell_t - m_t \rangle^2 \\
\nabla_t &= \nabla c_t(w_t) = \ell_t + 32\eta' \langle w_t, \ell_t - m_t \rangle (\ell_t - m_t) \\
\psi_t(w) &= \frac{1}{2} \|w\|_{A_t}^2, \qquad \text{where } A_t \triangleq \frac{1}{\eta}(I + \mathcal{G}_t)^{1/2}, \quad \mathcal{G}_t = \sum_{s=1}^{t} (\nabla_t - m_t)(\nabla_t - m_t)^\top.
\end{aligned}
$$

for $t = 1, \ldots, T$ **do**
Receive prediction m_t.
Compute $w_t = \operatorname{argmin}_{w \in \Omega} \left\{ \left\langle w, \sum_{s=1}^{t-1} \nabla_s + m_t \right\rangle + \psi_{t-1}(w) \right\}$.
Play w_t and receive ℓ_t.

$$
\leq 2 \sum_{t=1}^{T} \|\nabla_t - m_t\|_{A_{t-1}^{-1}}^2 + \sum_{t=1}^{T} \left\langle w_t - w'_t, m_t \right\rangle + \left\langle w'_t - u, \nabla_t \right\rangle.
$$

We prove by induction that for any $\tau, u \in \Omega$:

$$
\sum_{t=1}^{\tau} \left\langle w_t - w'_t, m_t \right\rangle + \left\langle w'_t, \nabla_t \right\rangle \leq \sum_{t=1}^{\tau} \langle u, \nabla_t \rangle + \psi_{\tau-1}(u).
$$

When $\tau = 1$, it suffices to show:

$$
\left\langle w_1 - w'_1, m_1 \right\rangle + \left\langle w'_1, \nabla_1 \right\rangle \leq \left\langle w'_1, \nabla_1 \right\rangle + \psi_0(w'_1).
$$

This is clearly true since $\langle w_1, m_1 \rangle \leq \langle w_1, m_1 \rangle + \psi_0(w_1) \leq \langle w'_1, m_1 \rangle + \psi_0(w'_1)$. Now suppose the result is true for $\tau = T$, then for $\tau = T + 1$:

$$
\begin{aligned}
&\sum_{t=1}^{T+1} \left\langle w_t - w'_t, m_t \right\rangle + \left\langle w'_t, \nabla_t \right\rangle \\
&\leq \left\langle w_{T+1}, \sum_{t=1}^{T} \nabla_t \right\rangle + \psi_{T-1}(w_{T+1}) + \left\langle w_{T+1} - w'_{T+1}, m_{T+1} \right\rangle + \left\langle w'_{T+1}, \nabla_{T+1} \right\rangle \\
&\hspace{20em} \text{(induction step for } \tau = T \text{ with } u = w_{T+1}) \\
&\leq \left\langle w'_{T+1}, \sum_{t=1}^{T} \nabla_t + m_{T+1} \right\rangle + \psi_T(w'_{T+1}) - \left\langle w'_{T+1}, m_{T+1} \right\rangle + \left\langle w'_{T+1}, \nabla_{T+1} \right\rangle \\
&\hspace{6em} \text{(by } \psi_{T-1}(w) \leq \psi_T(w), \text{ and } F_{\sum_{t=1}^{T} \nabla_t + m_{T+1}}(w_{T+1}) \leq F_{\sum_{t=1}^{T} \nabla_t + m_{T+1}}(w'_{T+1})) \\
&= \left\langle w'_{T+1}, \sum_{t=1}^{T+1} \nabla_t \right\rangle + \psi_T(w'_{T+1}) \leq \left\langle u, \sum_{t=1}^{T+1} \nabla_t \right\rangle + \psi_T(u),
\end{aligned}
$$

for any $u \in \Omega$ by the definition of w'_{T+1}. Therefore, by (Cutkosky, 2020, Theorem 7), we have:

$$
\sum_{t=1}^{T} \langle w_t - u, \nabla_t \rangle \leq 2 \sum_{t=1}^{T} \|\nabla_t - m_t\|_{A_{t-1}^{-1}}^2 + \psi_{T-1}(u) \leq 2 \sum_{t=1}^{T} \|\nabla_t - m_t\|_{A_{t-1}^{-1}}^2 + \frac{u^\top (I + \mathcal{G}_T)^{1/2} u}{\eta}
$$

$$= \mathcal{O}\left(\eta \mathrm{tr}\left(\mathcal{G}_T^{1/2}\right) + \frac{u^\top (I+\mathcal{G}_T)^{1/2} u}{\eta}\right) = \mathcal{O}\left(\eta \mathrm{tr}\left(\mathcal{L}_T^{1/2}\right) + \frac{u^\top (I+\mathcal{L}_T)^{1/2} u}{\eta}\right).$$

The reasoning of the last equality is as follows: note that $\nabla_t - m_t = (1 + 32\eta' \langle w_t, \ell_t - m_t\rangle)(\ell_t - m_t)$ has the same direction as $\ell_t - m_t$. Thus by assumption on η', $\mathcal{G}_t \preccurlyeq \frac{3}{2}\mathcal{L}_t$. Finally, note that c_t is a convex function. Therefore, $\sum_{t=1}^T c_t(w_t) - c_t(u) \le \sum_{t=1}^T \langle w_t - u, \nabla_t\rangle$. Reorganizing terms, we get:

$$\begin{aligned}
&\sum_{t=1}^T \langle w_t - u, \ell_t\rangle \\
&\le \mathcal{O}\left(\eta \mathrm{tr}\left(\mathcal{L}_T^{1/2}\right) + \frac{u^\top (I+\mathcal{L}_T)^{1/2} u}{\eta}\right) - 16\eta' \sum_{t=1}^T \langle w_t, \ell_t - m_t\rangle^2 + 16\eta' \sum_{t=1}^T \langle u, \ell_t - m_t\rangle^2 .
\end{aligned}$$

By $\eta' \le \eta / \|u\|_{A'_T}^2$ (note that $\|u\|_{A'_T}^2 = u^\top (I + \mathcal{L}_T)^{1/2} u$), we have:

$$\eta' \sum_{t=1}^T \langle u, \ell_t - m_t\rangle^2 \le \eta' \sum_{t=1}^T \|u\|_{A'_T}^2 \|\ell_t - m_t\|_{(A'_{t-1})^{-1}}^2 = \mathcal{O}\left(\eta \mathrm{tr}\left(\mathcal{L}_T^{1/2}\right)\right).$$

Therefore,

$$\sum_{t=1}^T \langle w_t - u, \ell_t\rangle = \mathcal{O}\left(\eta \mathrm{tr}\left(\mathcal{L}_T^{1/2}\right) + \frac{u^\top (I+\mathcal{L}_T)^{1/2} u}{\eta}\right) - 16\eta' \sum_{t=1}^T \langle w_t, \ell_t - m_t\rangle^2 .$$

■

Now we instantiate MSMWC-MASTER with the following set of experts to obtain the desired bound in Theorem 11.

$$\begin{aligned}
\mathcal{E}_{\text{AG}} = \Big\{&(\eta_k, \mathcal{A}_k) : \forall k = (d_k, t_k, l_k) \in \mathcal{S}_{\text{AG}}, \\
&\eta_k = \tfrac{1}{64 \cdot 2^{d_k + t_k}}, \mathcal{A}_k \text{ is Algorithm 5 with decision set } \Omega = \mathcal{K} \cap \{w : \|w\|_2 \le 2^{d_k}\}, \\
&\eta' = 2\eta_k \text{ and } \eta = 2^{l_k+1}\eta_k\Big\},
\end{aligned} \tag{23}$$

where $\mathcal{S}_{\text{AG}} = \{-\lceil \log_2 T\rceil, \ldots, \lceil \log_2 D\rceil\} \times [\lceil \log_2(dT)\rceil] \times \{-\lceil \log_2 T\rceil, \ldots, \lceil \log_2(2D^2T)\rceil\}$.

Proof [of Theorem 11] First assume $\|u\|_2 > \frac{1}{T}$, so that there exists $k_\star$ such that:

$$2^{d_{k_\star}-1} \le \|u\|_2 \le 2^{d_{k_\star}}, \eta_{k_\star} \le \min\left\{\frac{1}{128 \cdot 2^{d_{k_\star}}}, \frac{1}{\sqrt{\|u\|_{(I+\mathcal{L}_T)^{1/2}}^2 \mathrm{tr}\left(\mathcal{L}_T^{1/2}\right)}}\right\} \le 2\eta_{k_\star},$$

and $2^{l_k-1} \leq u^\top (I+\mathcal{L}_T)^{1/2} u \leq 2^{l_k}$. Note that $64\eta' |\left\langle w_t^{k_\star}, \ell_t - m_t \right\rangle| \leq 64\eta' \left\| w_t^{k_\star} \right\|_2 \leq 1$, and $\|u\|^2_{A'_T} \eta' \leq 2^{l_{k_\star}} \cdot 2\eta_{k_\star} = \eta$. Hence, by the regret guarantee of Algorithm 5, we have:

$$\begin{aligned}
\sum_{t=1}^T \left\langle w_t^{k_\star} - u, \ell_t \right\rangle &\leq \mathcal{O}\left(2^{l_k+1}\eta_{k_\star} \mathrm{tr}\left(\mathcal{L}_T^{1/2}\right) + \frac{u^\top (I+\mathcal{L}_T)^{1/2} u}{2^{l_k+1}\eta_{k_\star}} \right) - 32\eta_{k_\star} \sum_{t=1}^T \left\langle w_t^{k_\star}, \ell_t - m_t \right\rangle^2 \\
&\leq \mathcal{O}\left(\|u\| + \sqrt{(u^\top (I+\mathcal{L}_T)^{1/2} u)\mathrm{tr}\left(\mathcal{L}_T^{1/2}\right)} \right) - 32\eta_{k_\star} \sum_{t=1}^T \left\langle w_t^{k_\star}, \ell_t - m_t \right\rangle^2 .
\end{aligned}$$

Next, by Theorem 4 with $32\eta_k \left|\left\langle w_t^k, \ell_t - m_t \right\rangle\right| \leq 32\eta_k \left\| w_t^k \right\|_2 \leq 1$, $\sum_k \eta_k = \Theta(T)$, $\sum_k \eta_k^2 = \Theta(T^2)$, and $\frac{\sum_k \eta_k^2}{\eta_{k_\star}^2} = \mathcal{O}(d^2 D^2 T^4)$, we have:

$$\sum_{t=1}^T \langle w_t - u, \ell_t \rangle = \tilde{\mathcal{O}}\left(\|u\|_2 + \sqrt{(u^\top (I+\mathcal{L}_T)^{1/2} u)\mathrm{tr}\left(\mathcal{L}_T^{1/2}\right)} \right).$$

When $\|u\|_2 \leq \frac{1}{T}$, pick any $u' \in \mathcal{K}$ such that $\|u'\|_2 = \frac{1}{T}$, then:

$$\begin{aligned}
\sum_{t=1}^T \langle w_t - u, \ell_t \rangle &= \sum_{t=1}^T \langle w_t - u', \ell_t \rangle + \sum_{t=1}^T \langle u' - u, \ell_t \rangle \\
&\leq \tilde{\mathcal{O}}\left(\left\|u'\right\|_2 + \sqrt{(u'^\top (I+\mathcal{L}_T)^{1/2} u')\mathrm{tr}\left(\mathcal{L}_T^{1/2}\right)} + \left\|u'\right\|_2 \right) = \tilde{\mathcal{O}}\left(1\right).
\end{aligned}$$

This finishes the proof. ■

D.4. Combining MetaGrad's base algorithm

We first present the MetaGrad base algorithm (Algorithm 6) and its regret guarantee below (note that the algorithm receives $\bar{w}_t$ at the end of round t, which will eventually be set to the master's prediction in our construction).

Lemma 22 *Assume* $64\eta D \leq 1$. *Algorithm 6 ensures:*

$$\sum_{t=1}^T \langle w_t - u, \ell_t \rangle \leq \mathcal{O}\left(\|u\|_2 + \frac{r \ln T}{\eta} + \eta \sum_{t=1}^T \langle u - \bar{w}_t, \ell_t - m_t \rangle^2 \right) - 10\eta \sum_{t=1}^T \langle w_t - \bar{w}_t, \ell_t - m_t \rangle^2 .$$

Proof By Lemma 15 and Lemma 20 with $z_t = 1$ for all t, we have:

$$\begin{aligned}
&\sum_{t=1}^T \langle w_t - u, \nabla_t \rangle \\
&\leq \sum_{t=1}^T \left\langle w_t - w'_{t+1}, \nabla_t - m_t \right\rangle + D_{\psi_t}(u, w'_t) - D_{\psi_t}(u, w'_{t+1})
\end{aligned}$$

Algorithm 6 MetaGrad

Parameters: learning rate $\eta > 0$, $w'_1 = \mathbf{0}$.
Define:

$$c_t(w) = \langle w, \ell_t \rangle + 16\eta \langle w - \bar{w}_t, \ell_t - m_t \rangle^2$$
$$\nabla_t = \nabla c_t(w_t) = \ell_t + 32\eta \langle w_t - \bar{w}_t, \ell_t - m_t \rangle (\ell_t - m_t)$$
$$\psi_t(w) = \frac{1}{2} \|w\|_{A_t}^2, \qquad \text{where } A_t \triangleq \eta \left(8I + \sum_{s=1}^{t-1} (\nabla_s - m_s)(\nabla_s - m_s)^\top \right).$$

for $t = 1, \ldots, T$ **do**
- Receive prediction m_t.
- Play $w_t = \mathrm{argmin}_{w \in \mathcal{K}} \{\langle w, m_t \rangle + D_{\psi_t}(w, w'_t)\}$.
- Receive ℓ_t and $\bar{w}_t$.
- Compute $w'_{t+1} = \mathrm{argmin}_{w \in \mathcal{K}} \{\langle w, \nabla_t \rangle + D_{\psi_t}(w, w'_t)\}$.

$$\le 2 \sum_{t=1}^{T} \|\nabla_t - m_t\|_{A_t^{-1}}^2 + D_{\psi_1}(u, w'_1) + \sum_{t=1}^{T-1} D_{\psi_{t+1}}(u, w'_{t+1}) - D_{\psi_t}(u, w'_{t+1})$$
$$\le \mathcal{O}\left(\frac{r \ln T}{\eta} + \eta \|u\|_2^2 \right) + \sum_{t=1}^{T-1} D_{\psi_{t+1}}(u, w'_{t+1}) - D_{\psi_t}(u, w'_{t+1}).$$

Note that $\eta \|u\|_2^2 = \mathcal{O}(\|u\|_2)$. Moreover,

$$\begin{aligned}
&\sum_{t=1}^{T-1} D_{\psi_{t+1}}(u, w'_{t+1}) - D_{\psi_t}(u, w'_{t+1}) \\
&= \frac{\eta}{2} \sum_{t=1}^{T-1} \left\langle u - w'_{t+1}, \nabla_t - m_t \right\rangle^2 \\
&\le \eta \sum_{t=1}^{T-1} \langle u - w_t, \nabla_t - m_t \rangle^2 + \eta \sum_{t=1}^{T-1} \left\langle w_t - w'_{t+1}, \nabla_t - m_t \right\rangle^2 \\
&\le 3\eta \sum_{t=1}^{T-1} \langle u - w_t, \ell_t - m_t \rangle^2 + \mathcal{O}\left(\frac{r \ln T}{\eta} \right),
\end{aligned}$$

where the last step is by $0 \le \eta \left\langle w_t - w'_{t+1}, \nabla_t - m_t \right\rangle \le 3\eta D = \mathcal{O}(1)$ and Lemma 20. Since $c_t(w)$ is convex in w, we have $\sum_{t=1}^{T} c_t(w_t) - c_t(u) \le \sum_{t=1}^{T} \langle w_t - u, \nabla_t \rangle$. Re-organzing terms, we have:

$$\begin{aligned}
\sum_{t=1}^{T} \langle w_t - u, \ell_t \rangle &\le \mathcal{O}\left(\frac{r \ln T}{\eta} + \|u\|_2 \right) + 3\eta \sum_{t=1}^{T} \langle u - w_t, \ell_t - m_t \rangle^2 \\
&\quad + 16\eta \sum_{t=1}^{T} \langle u - \bar{w}_t, \ell_t - m_t \rangle^2 - 16\eta \sum_{t=1}^{T} \langle w_t - \bar{w}_t, \ell_t - m_t \rangle^2
\end{aligned}$$

$$\leq \mathcal{O}\left(\frac{r \ln T}{\eta} + \|u\|_2 + \eta \sum_{t=1}^{T} \langle u - \bar{w}_t, \ell_t - m_t \rangle^2\right) - 10\eta \sum_{t=1}^{T} \langle w_t - \bar{w}_t, \ell_t - m_t \rangle^2 .$$

■

Then, we instantiate MSMWC-MASTER with the following set of experts to obtain the desired bound in Theorem 13.

$$\mathcal{E}_{\text{MG}} = \Big\{(\eta_k, \mathcal{A}_k) : \forall k \in [\lceil \log_2(2DT) \rceil], \eta_k = \tfrac{1}{64D \cdot 2^k}, \\ \mathcal{A}_k \text{ is Algorithm 6 with } \bar{w}_t = w_t \text{ for all } t \text{ and } \eta = 4\eta_k \Big\}. \tag{24}$$

Proof [of Theorem 13] There exists $k_\star$ such that $\eta_{k_\star} \leq \min\left\{\frac{1}{256D}, \sqrt{\frac{r \ln T}{\sum_{t=1}^T \langle u - w_t, \ell_t - m_t \rangle^2}}\right\} \leq 2\eta_{k_\star}$. Then by Lemma 22 with $64 \cdot 4\eta_{k_\star} D \leq 1$:

$$\begin{aligned}
&\sum_{t=1}^{T} \left\langle w_t^{k_\star} - u, \ell_t \right\rangle \\
&\leq \mathcal{O}\left(\frac{r \ln T}{\eta_{k_\star}} + \|u\|_2 + \eta_{k_\star} \sum_{t=1}^{T} \langle u - w_t, \ell_t - m_t \rangle^2\right) - 40\eta_{k_\star} \sum_{t=1}^{T} \left\langle w_t^{k_\star} - w_t, \ell_t - m_t \right\rangle^2 \\
&= \tilde{\mathcal{O}}\left(rD + \sqrt{r \sum_{t=1}^{T} \langle u - w_t, \ell_t - m_t \rangle^2}\right) - 40\eta_{k_\star} \sum_{t=1}^{T} \left\langle w_t^{k_\star} - w_t, \ell_t - m_t \right\rangle^2 .
\end{aligned}$$

Next, by Theorem 4 with $32\eta_k |g_{t,k} - h_{t,k}| = 32\eta_k \left|\left\langle w_t^k - w_t, \ell_t - m_t \right\rangle\right| \leq 64\eta_k D \leq 1$, $\sum_k \eta_k = \Theta(1/D)$, $\sum_k \eta_k^2 = \Theta(1/D^2)$, and $\frac{\sum_k \eta_k^2}{\eta_{k_\star}^2} = \mathcal{O}(D^4 T^2)$, we have:

$$\begin{aligned}
\sum_{t=1}^{T} \langle w_t - u, \ell_t \rangle &= \tilde{\mathcal{O}}\left(rD + \sqrt{r \sum_{t=1}^{T} \langle u - w_t, \ell_t - m_t \rangle^2 + \frac{1}{\eta_{k_\star}}}\right) \\
&= \tilde{\mathcal{O}}\left(rD + \sqrt{r \sum_{t=1}^{T} \langle u - w_t, \ell_t - m_t \rangle^2}\right) .
\end{aligned}$$

This completes the proof. ■

D.5. Extensions to unconstrained learning and unknown Lipschitzness

In this subsection, we present general ideas on extending our OLO results to the setting with an unconstrained decision set, unknown Lipschitzness, or both. We focus on $\sqrt{r \sum_t \langle u, \ell_t - m_t \rangle^2}$ type of bound and omit the details for the others for simplicity.

D.5.1. UNCONSTRAINED LEARNING WITH KNOWN LIPSCHITZNESS

We first consider the case where $D = \infty$ and $\max_t \max\{\|\ell_t\|_2, \|\ell_t - m_t\|_2\} \le 1$. We argue that in this case we can simply assume that $\|u\|_2 \le 2^T$, so that we only need to maintain $\mathcal{O}(T)$ experts. Suppose the assumption does not hold and $T < \log_2 \|u\|_2$. Then, by constraining $\|w_t\|_2 \le T$, we have: $\sum_{t=1}^T \langle w_t - u, \ell_t\rangle \le T\|u\|_2 < \|u\|_2 \log_2 \|u\|_2 = \tilde{\mathcal{O}}(\|u\|_2)$. Therefore, running the algorithm in Theorem 9 assuming the diameter is T, we obtain the same bound as before:

$$\text{REG}(u) = \tilde{\mathcal{O}}\left(r\|u\|_2 + \sqrt{r\sum_{t=1}^T \langle u, \ell_t - m_t\rangle^2}\right).$$

Note that when $m_t = 0$, the bound we obtained has the same order as that in (Cutkosky and Orabona, 2018, Theorem 8).

D.5.2. CONSTRAINED LEARNING WITH UNKNOWN LIPSCHITZNESS

Next, we consider the case where $D < \infty$ and $\max_t \|\ell_t - m_t\|_2$ is unknown. We can handle this by simply applying our master with unknown loss range (Algorithm 3) with the following expert set generator:

$$\mathcal{E}_{\text{ONSUL}}(B_0) = \Big\{(\eta_k, \mathcal{A}_k) : \forall k \in [N], \eta_k = \tfrac{1}{192DB_0 2^k}, \mathcal{A}_k \text{ is Algorithm 4}$$
$$\text{with } z_t = B_{t-1} = \max_{0\le s<t}\|\ell_s - m_s\| \text{ for all } t, \Omega = \mathcal{K}, \text{ and } \eta = 3\eta_k\Big\}.$$

and $\Lambda_t = \Delta_{\mathcal{S}_{\text{ONSUL}}(t)}$, where $N = \lceil \log_2 T^2\rceil$, $\mathcal{S}_{\text{ONSUL}}(t) = \Big\{k \in [N] : \frac{1}{192DB_0 2^k} \le \frac{1}{192DB_{t-1}}\Big\}$.

Theorem 23 *Let $\max_t \|\ell_t - m_t\|_2$ be unknown, $r \le d$ be the rank of $\mathcal{L}_T = \sum_{t=1}^T (\ell_t - m_t)(\ell_t - m_t)^\top$. Algorithm 3 with expert set generator $\mathcal{E}_{\text{ONSUL}}$ and $\Lambda_t = \Delta_{\mathcal{S}_{\text{ONSUL}}(t)}$ ensures for all $u \in \Delta_d$,*

$$\forall u \in \mathcal{K},\quad \text{REG}(u) = \tilde{\mathcal{O}}\left(rDB + \sqrt{r\sum_{t=1}^T \langle u, \ell_t - m_t\rangle^2}\right).$$

Proof We first show that when there is no restart before episode t, we obtain the desired regret bound. The assumption implies that $\frac{B_{T-1}}{B_0} \le T$, and thus $\frac{1}{\max\{192,T\}DB_{T-1}} \ge \frac{1}{192DB_0 2^N}$. Therefore, there exists $k_\star$ such that

$$\eta_{k_\star} \le \min\left\{\frac{1}{192DB_{T-1}}, \sqrt{\frac{r\ln(TB_{T-1}/B_0)}{\sum_{t=1}^T \left\langle u, \bar{\ell}_t - m_t\right\rangle^2}}\right\} \le 2\eta_{k_\star}.$$

Hence, by Lemma 19 with $64 \cdot 3\eta_{k_\star} DB_{T-1} \le 1$, we have:

$$\sum_{t=1}^T \left\langle w_t^{k_\star} - u, \bar{\ell}_t\right\rangle = \mathcal{O}\left(\frac{r\ln T}{\eta_{k_\star}} + DB + \eta_{k_\star}\sum_{t=1}^T \left\langle u, \bar{\ell}_t - m_t\right\rangle^2\right) - 33\eta_{k_\star}\sum_{t=1}^T \left\langle w_t^{k_\star}, \ell_t - m_t\right\rangle^2$$

$$= \tilde{\mathcal{O}}\left(rDB + \sqrt{r\sum_{t=1}^{T}\left\langle u, \bar{\ell}_t - m_t\right\rangle^2}\right) - 33\eta_{k_\star}\sum_{t=1}^{T}\left\langle w_t^{k_\star}, \ell_t - m_t\right\rangle^2.$$

By Theorem 4 with $32\eta_k\left|\left\langle w_t^k, \bar{\ell}_t - m_t\right\rangle\right| \leq 32\eta_k DB_{t-1} \leq 1$ for any $k \in \mathcal{S}_{\text{ONSUL}}(t)$, $\sum_k \eta_k = \Theta(\frac{1}{DB_0})$, $\sum_k \eta_k^2 = \Theta(\frac{1}{D^2B_0^2})$, and $\frac{\sum_k \eta_k^2}{\eta_{k_\star}^2} = \mathcal{O}(\eta_1^2/\eta_{k_\star}^2) = \mathcal{O}(T^4)$, we have:

$$\sum_{t=1}^{T}\left\langle w_t - u, \bar{\ell}_t\right\rangle = \tilde{\mathcal{O}}\left(rDB + \sqrt{r\sum_{t=1}^{T}\left\langle u, \bar{\ell}_t - m_t\right\rangle^2}\right).$$

Moreover, note that,

$$\begin{aligned}\sum_{t=1}^{T}\left\langle w_t - u, \ell_t - \bar{\ell}_t\right\rangle &\leq 2\sum_{t=1}^{T} D\left\|\ell_t - \bar{\ell}_t\right\|_2 \leq 2\sum_{t=1}^{T} D\frac{B_t - B_{t-1}}{B_t}\left\|\ell_t - m_t\right\|_2 \\ &\leq 2D\sum_{t=1}^{T}(B_t - B_{t-1}) \leq 2DB.\end{aligned}$$

Therefore, by $\left\langle u, \bar{\ell}_t - m_t\right\rangle^2 = (1 - \frac{B_{t-1}}{B_t})^2\left\langle u, \ell_t - m_t\right\rangle^2 \leq \left\langle u, \ell_t - m_t\right\rangle^2$,

$$\begin{aligned}\sum_{t=1}^{T}\left\langle w_t - u, \ell_t\right\rangle &= \sum_{t=1}^{T}\left\langle w_t - u, \bar{\ell}_t\right\rangle + \sum_{t=1}^{T}\left\langle w_t - u, \ell_t - \bar{\ell}_t\right\rangle \\ &= \tilde{\mathcal{O}}\left(rDB + \sqrt{r\sum_{t=1}^{T}\left\langle u, \ell_t - m_t\right\rangle^2}\right).\end{aligned}$$

Finally, we assume there are at least one restarts. Following similar analysis in the proof of Theorem 8, we consider regret in the following three intervals: $[1, \tau_1]$, $(\tau_1, \tau_2]$, and $(\tau_2, T]$. The regret in $[1, \tau_1]$ is bounded by B according to Eq. (20). By $B_{\tau_2 - 1}/B_{\tau_1} \leq T, B_{T-1}/B_{\tau_2} \leq T$, we have:

$$\begin{aligned}\textsc{Reg}^{(\tau_1,\tau_2]}(u) &= \mathcal{O}\left(rDB + \sqrt{r\sum_{t\in(\tau_1,\tau_2]}\left\langle u, \bar{\ell}_t - m_t\right\rangle^2}\right) \\ \textsc{Reg}^{(\tau_2,T]}(u) &= \mathcal{O}\left(rDB + \sqrt{r\sum_{t\in(\tau_2,T]}\left\langle u, \bar{\ell}_t - m_t\right\rangle^2}\right).\end{aligned}$$

Summing the regret in three intervals and applying the Cauchy-Schwarz inequality, we get the desired result. ■

D.5.3. UNCONSTRAINED LEARNING WITH UNKNOWN LIPSCHITZNESS

Finally, we consider the case where $D = \infty$ and $\max_t \left\|\ell_t - m_t\right\|_2$ is unknown. Cutkosky (2019a); Mhammedi and Koolen (2020) show that to obtain $\tilde{\mathcal{O}}\left(\sqrt{T}\right)$ regret, it is sufficient to control the diameter of decision set to be of order $\tilde{\mathcal{O}}\left(\sqrt{T}\right)$. Specifically, they set the size of the decision set to be $\sqrt{\max_{s\le t}\sum_{s'=1}^{s}\|\ell_{s'}\|_2/G_s}$ in episode t, where $G_t = \max_{s\le t}\|\ell_s\|_2$. To bound the regret when the comparator is not in the decision set, they make use of a reduction to constrained domain (Cutkosky and Orabona, 2018). However, their reduction is not directly applicable in our case, since the reduction modifies the loss function and ruins the data-dependent bound. There is a follow up work (Cutkosky, 2020) achieving the bound $\tilde{\mathcal{O}}\left(\sqrt{r\sum_{t=1}^{T}\langle u,\ell_t\rangle^2}\right)$ under constrained domain by adapting to time-dependent norms. However, it is not obvious how to incorporate predictor m_t into their algorithm.

Here, we take a different route: we search over the appropriate constraint of the decision set with doubling trick: if in episode t we find that $\sqrt{\sum_{s=1}^{t}\|\ell_s\|_2/G_t} > D_t$, where D_t is the diameter of decision set in episode t, then, we let $D_{t+1} = 2\sqrt{\sum_{s=1}^{t}\|\ell_s\|_2/G_t}$, and restart the algorithm with the new decision set. Otherwise we let $D_{t+1} = D_t$. The number of restart is $\mathcal{O}(\log_2 T)$ since $\max_{s\le t}\sum_{s'=1}^{s}\|\ell_{s'}\|_2/G_s \le T$. We summarize our algorithm as a new variant of MSMWC-MASTER in Algorithm 7.

Algorithm 7 MSMWC-MASTER with unknown loss range and unbounded diameter

Input: An expert set generator $\mathcal{E}$ that takes diameter and initial scale as input, initial scale B_0.
Initialization: $D_1 = 1$. Initialize $\mathcal{A}$ as an instance of Algorithm 3 with input $\mathcal{E}(D_1,\cdot)$ and B_0.
for $t = 1,\ldots,T$ **do**
  Execute $\mathcal{A}$ for episode t.
  if $D_t < \sqrt{\sum_{s=1}^{t}\frac{\|\ell_s\|_2}{G_t}}$ **then**
    $D_{t+1} = 2\sqrt{\sum_{s=1}^{t}\frac{\|\ell_s\|_2}{G_t}}$.
    Initialize $\mathcal{A}$ as an instance of Algorithm 3 with input $\mathcal{E}(D_{t+1},\cdot)$ and B_t.
  else
    $D_{t+1} = D_t$.

Now we show how to extend the regret bound of ONS to the setting with unconstrained diameter and unknown Lipschitzness.

Theorem 24 *Define the expert set generator:*

$$\mathcal{E}_{\text{ONSULD}}(D,B_0) = \Big\{(\eta_k,\mathcal{A}_k) : \forall k\in[N], \eta_k = \tfrac{1}{192DB_02^k}, \mathcal{A}_k \text{ is Algorithm 4 with } z_t = B_{t-1} = \max_{0\le s<t}\|\ell_s - m_s\| \text{ for all } t,\ \Omega = \mathcal{K}\cap\{w : \|w\|_2\le D\}, \text{ and } \eta = 3\eta_k\Big\}.$$

Then, Algorithm 7 with input $\mathcal{E}_{\text{ONSULD}}$, B_0, ensures

$$\text{REG}(u) = \tilde{\mathcal{O}}\left(\sqrt{r\sum_{t=1}^{T}\langle u,\ell_t - m_t\rangle^2} + rB\sqrt{\max_{t\le T}\sum_{s=1}^{t}\|\ell_s\|_2/G_t} + G_T\|u\|_2^3\right),$$

where $G_T = \max_{t\le T}\|\ell_t\|$, $B = \max\{B_0, \max_{t\le T}\|\ell_t - m_t\|\}$.

Proof We split T episodes into M intervals $I_{1:M}$, where the last episode of I_m (denote by t_m) either equals to T or $D_{t_m+1} \neq D_{t_m}$. Define projection function $f(u, D) = \min\left\{1, \frac{D}{\|u\|_2}\right\} u$. Then, the regret is bounded as follows (note that $D_t = D_{t_m}$ for all $t \in I_m$):

$$\sum_{t=1}^{T} \langle w_t - u, \ell_t \rangle = \sum_{m=1}^{M} \sum_{t \in I_m} \langle w_t - f(u, D_{t_m}), \ell_t \rangle + \sum_{t=1}^{T} \langle f(u, D_t) - u, \ell_t \rangle .$$

For the first term, by Theorem 23 with $\langle f(u, D), \ell_t - m_t \rangle^2 \leq \langle u, \ell_t - m_t \rangle^2$ for any $D > 0$, and $M = \mathcal{O}\left(\log_2 T\right)$, we obtain

$$\begin{aligned}
\sum_{m=1}^{M} \sum_{t \in I_m} \langle w_t - f(u, D_{t_m}), \ell_t \rangle &= \mathcal{O}\left(\sum_{m=1}^{M} r D_{t_m} B + \sqrt{r \sum_{t \in I_m} \langle u, \ell_t - m_t \rangle^2}\right) \\
&= \tilde{\mathcal{O}}\left(r D_T B + \sqrt{r \sum_{t=1}^{T} \langle u, \ell_t - m_t \rangle^2}\right) .
\end{aligned}$$

For the second term, denote by $t_\star$ the last episode such that $u \neq f(u, D_{t_\star})$. Then, $\|u\|_2 \geq \sqrt{\sum_{t=1}^{t_\star - 1} \|\ell_t\|_2 / G_{t_\star - 1}}$, $\|u\|_2 \geq \|f(u, D_t)\|_2$ for $t \leq t_\star$, and

$$\begin{aligned}
\sum_{t=1}^{T} \langle f(u, D_t) - u, \ell_t \rangle = \sum_{t=1}^{t_\star} \langle f(u, D_t) - u, \ell_t \rangle &\leq 2 \sum_{t=1}^{t_\star - 1} \|u\|_2 \|\ell_t\|_2 + 2 \|u\|_2 G_T \\
&\leq 2 \|u\|_2 G_T \sum_{t=1}^{t_\star - 1} \frac{\|\ell_t\|_2}{G_{t_\star - 1}} + 2 \|u\|_2 G_T \leq 2 G_T \|u\|_2^3 + 2 \|u\|_2 G_T.
\end{aligned}$$

■