Proceedings of Machine Learning Research vol 313:1–20, 2026 37th International Conference on Algorithmic Learning Theory

Multi-distribution Learning: From Worst-Case Optimality to Lexicographic Min-Max Optimality

Guanghui Wang GWANG369@GATECH.EDU
College of Computing, Georgia Institute of Technology

Umar Syed USYED@GOOGLE.COM
Google Research

Robert E. Schapire SCHAPIRE@MICROSOFT.COM
Microsoft Research

Jacob Abernethy ABERNETHYJ@GOOGLE.COM
Google Research
College of Computing, Georgia Institute of Technology

Editors: Matus Telgarsky and Jonathan Ullman

Abstract

We study multi-distribution learning (MDL), where the goal is to train a model that performs well across a set of underlying groups or distributions. The predominant performance metric in MDL is *min-max* optimality, which guarantees minimum error on the "hardest" distribution presented. Optimizing this metric has the unfortunate side effect of producing models that potentially sacrifice performance gains on non-worst-case groups. In the present work we propose a natural alternative, the lexicographic min-max (lex-min-max) objective, which promotes balanced performance by sequentially minimizing the worst, second-worst, and subsequent group losses. Despite its non-convex nature, we show that obtaining an (approximate) lex-min-max solution can be *as easy as* achieving (approximate) min-max optimality. We develop an efficient algorithm that directly approximates lex-min-max optimality via implementing stochastic no-regret dynamics on a regularized variant of the classical min-max objective. Our method is efficient and easy to implement, and it advances the frontier of multi-distribution learning by providing stronger, hierarchy-aware performance guarantees.

Keywords: Multi-distribution Learning, Lexicographic Min-Max Optimization, Game Theory

1. Introduction

In classical statistical machine learning, a central assumption is that the training and test data are drawn from a fixed underlying distribution. However, this assumption often fails in real-world applications, where data naturally emerges from a variety of sources. For instance, in collaborative learning (Blum et al., 2017; Chen et al., 2018; Blum et al., 2021a,b), different parties possess data generated from distinct distributions and aim to jointly train a single model. In agnostic federated learning (McMahan et al., 2017; Mohri et al., 2019), client devices collect data under heterogeneous and often unknown conditions, necessitating models that generalize well across all clients. Similarly, in group distributionally robust optimization (Group DRO) (Soma et al., 2022; Sagawa et al., 2020; Zhang et al., 2023), the objective is to handle the learner's uncertainty about the true task by ensuring strong performance across a range of underlying distributions.

These scenarios motivate the framework of multi-distribution learning (MDL) (Haghtalab et al., 2022), where the goal is to train a single model that generalizes well across all input distributions.

In recent years, MDL has attracted significant attention, leading to a variety of MDL approaches (Haghtalab et al., 2022; Awasthi et al., 2023; Peng, 2024; Zhang et al., 2024; Lin et al., 2024). The prevalent performance metric in the MDL literature is *min-max optimality*, which emphasizes robustness by optimizing performance on the worst-case distribution (Haghtalab et al., 2022). This approach is particularly appealing in high-stakes settings, where it is crucial to avoid catastrophic failures for any group. By minimizing the maximum loss across all groups, the min-max objective offers strong performance guarantees for the most challenging subgroup. By formulating the objective as a zero-sum game, the target model can be efficiently optimized using standard no-regret dynamics (Awasthi et al., 2023).

However, focusing solely on the worst-case distribution introduces important limitations. A central drawback of the min-max objective is its indifference to performance on all *non-worst-case distributions*. Once the maximum loss is minimized, the objective provides no further incentive to improve outcomes for the remaining groups. Consequently, a model that is min-max optimal can still underperform across the majority of distributions—an issue that becomes particularly salient in two common scenarios. First, when there exists an extreme outlier distribution on which all models perform poorly, the worst-case risk becomes effectively fixed, making the min-max criterion insensitive to model differences. Second, in high-dimensional settings with many underlying distributions, the solution set minimizing the worst-case loss can be very large, encompassing models with vastly different behaviors on the remaining groups. These observations lead to a natural question: *among all models achieving optimal worst-case performance, how can we identify the best one?*

To address this challenge, we propose a natural extension to the standard min-max MDL framework: *lexicographic min-max (or lex-min-max) MDL*. In this formulation, the learner first minimizes the worst-case loss; subject to achieving that minimum, it then minimizes the second-worst loss; and so on. The resulting solution is lexicographically min-max optimal—that is, optimal with respect to a lexicographic ordering of the per-distribution losses. This framework yields more balanced performance across distributions, strictly generalizing the standard min-max criterion by providing performance guarantees across all distributions.

The lexicographic min-max objective has deep roots in other domains. It was originally introduced in computing the nucleolus in cooperative game theory (Kohlberg, 1972), and has been extensively studied in bandwidth allocation for computer networks (Hayden, 1981; Mo and Walrand, 2000; Le Boudec, 2005), as well as in optimization (Behringer, 1977; Ogryczak, 1997; Yager, 1997; Ogryczak and Śliwiński, 2006; Nacson et al., 2019; Hartman et al., 2023; Abernethy et al., 2024). Despite its intuitive appeal, the objective has received little attention in machine learning—likely due to its non-convex nature. To our knowledge, the only prior work in this direction is by Diana et al. (2021), who proposed a *convexified relaxation* of the lex-min-max objective. Rather than comparing the exact j-th largest risk at each level j, they consider the sum of the top-j risks. While computationally convenient due to its convex nature, this formulation has notable limitations. First, it can obscure distinctions—especially when the largest risk dominates the rest by a significant amount—making the top-j sum effectively equivalent to the top-1 loss. Additionally, the relationship between this relaxation and true lexicographic optimality is unclear when an approximation error appears. Finally, due to the nested structure of their performance metric, their method is purely offline and cannot be updated efficiently with newly observed examples.

In this paper, we depart from convex relaxations and focus on directly obtaining an approximately lex-min-max optimal model in an efficient manner. We show that, while standard min-max MDL approaches solve a zero-sum game using online learning dynamics, lex-min-max optimality can be approximated by solving a regularized variant of the same game, with a modified support. This simple yet powerful modification enables the learner to achieve lex-min-max optimality using standard no-regret dynamics, making the process *almost as easy* as attaining approximate min-max optimality. Our work thus introduces a principled generalization of robust learning that accounts for the full hierarchy of distributions. It provides strictly stronger performance guarantees across all distributions, advancing the goal of equitable and robust machine learning in multi-distribution environments. In summary, our main contributions are:

- We propose a lexicographic min-max objective for multi-distribution learning, which directly optimizes the lex-min-max criterion rather than relying on convex relaxations.
- We show that our approach is simple to implement using standard no-regret online learning dynamics on a modified min-max game. It requires only a lightweight modification of existing MDL algorithms and supports efficient online updates.
- We also conduct experiments on both synthetic and real-world data sets to demonstrate the effectiveness of the proposed methods.

2. Preliminaries

In this section, we introduce the basic setup and notation used throughout the paper. We then formally define the MDL problem and discuss several key performance metrics for evaluating algorithmic optimality.

Notation We use lower case bold face letters $\mathbf{x}, \mathbf{y}$ to denote vectors, lower case letters a, b to denote scalars. For a vector $\mathbf{r} \in \mathbb{R}^k$, we use r_i to denote the i-th component of $\mathbf{r}$. We also use $\sigma_j(\mathbf{r})$ to denote the j-th largest element in $\mathbf{r}$, for $j \in [k]$. For example, suppose $\mathbf{r} = [0, 0.5, 0.4] \in \mathbb{R}^3$. Then, $\sigma_1(\mathbf{r}) = 0.5$, $\sigma_2(\mathbf{r}) = 0.4$, and $\sigma_3(\mathbf{r}) = 0$. For any two vectors $\mathbf{p}, \mathbf{q}$ from some convex set $\mathcal{S} \subseteq \mathbb{R}^d$, and a strictly convex, differentiable function $\phi : \mathcal{S} \mapsto \mathbb{R}$, we use $B_\phi(\mathbf{p}, \mathbf{q})$ to denote the Bregman divergence between $\mathbf{p}$ and $\mathbf{q}$, defined as:

$$B_\phi(\mathbf{p}, \mathbf{q}) = \phi(\mathbf{p}) - \phi(\mathbf{q}) - \langle \nabla \phi(\mathbf{q}), \mathbf{p} - \mathbf{q} \rangle. \tag{1}$$

In particular, consider $\mathcal{S} = \Delta_d$. Then, if ϕ is the negative entropy function, i.e., $\phi(\mathbf{q}) = \sum_{i=1}^d q_i \log q_i$, the Bregman divergence leads to the generalized KL-divergence. More specifically, for any $\mathbf{p}, \mathbf{q} \in \Delta_d$, $\mathrm{KL}(\mathbf{p} \| \mathbf{q}) = \sum_{i=1}^d p_i \log \frac{p_i}{q_i} - \sum_i p_i + \sum_i q_i$.

Basic setting In this paper, we study the problem of multi-distribution learning (Blum et al., 2017; Haghtalab et al., 2022). We use $(\mathbf{x}, y)$ to denote a learning example, where $\mathbf{x}$ belongs to an instance space $\mathcal{X}$ and y lies in the label space $\mathcal{Y}$. The data is generated by a collection of k distinct data distributions, denoted by $\mathcal{D} = (\mathcal{D}^{(1)}, \mathcal{D}^{(2)}, \ldots, \mathcal{D}^{(k)})$, each supported over $\mathcal{X} \times \mathcal{Y}$. The learning algorithm chooses hypotheses from a finite[1] hypothesis class $\mathcal{H}$ with bounded cardinality, which consists of functions $h \in \mathcal{H}$ mapping instances in $\mathcal{X}$ to labels in $\mathcal{Y}$. Following previous work, we also allow for randomized selection of hypotheses via a distribution $\mathbf{q} \in \Delta(\mathcal{H})$. Model performance on a specific example is assessed using a loss function $\ell : \mathcal{H} \times \mathcal{X} \times \mathcal{Y} \to [0, 1]$, where $\ell(h, (\mathbf{x}, y))$

1. See Appendix B.1 for a generalization to infinite hypothesis classes

measures the error incurred when hypothesis h is applied to the input-label pair $(\mathbf{x}, y)$. For a policy $\mathbf{q} \in \Delta(\mathcal{H})$, the risk of $\mathbf{q}$ under distribution $\mathcal{D}^{(i)}$ is given by

$$R(\mathbf{q}, \mathcal{D}^{(i)}) = \mathbb{E}_{h\sim\mathbf{q},(\mathbf{x},y)\sim\mathcal{D}^{(i)}} \left[\ell(h, (\mathbf{x}, y))\right].$$

Finally, to further simplify the notation, we use the vector

$$\mathbf{r}_{\mathbf{q},\mathcal{D}} = [R(\mathbf{q}, \mathcal{D}^{(1)}), \ldots, R(\mathbf{q}, \mathcal{D}^{(k)})]^\top$$

to denote the risk of $\mathbf{q}$ over $\mathcal{D}$.

2.1. Performance Measures for MDL

Next, we discuss several performance metrics for evaluating the optimality of algorithms in the MDL setting.

The first performance metric is min-max optimality, which is predominantly used in the MDL literature (Haghtalab et al., 2022; Awasthi et al., 2023; Peng, 2024; Zhang et al., 2024).

Definition 1 (ε-min-max optimality) *Given a target accuracy level $\varepsilon \in (0, 1)$, $\mathbf{q} \in \Delta(\mathcal{H})$ is ε-min-max optimal if*

$$\sigma_1(\mathbf{r}_{\mathbf{q},\mathcal{D}}) \leq \min_{\mathbf{q}'\in\Delta(\mathcal{H})} \sigma_1\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right) + \varepsilon, \tag{2}$$

where $\sigma_1(\mathbf{r})$ denotes the maximum element of $\mathbf{r}$.

Min-max optimality primarily focuses on the worst-case risk, which may fail to identify truly optimal solutions in many cases. For example, there might exist a distribution under which all hypotheses incur extremely large errors, making the worst-case risk less informative. Furthermore, especially in settings with a large hypothesis class and many distributions, there may exist a large set of hypotheses that achieve the same minimal maximum error. Motivated by these limitations, we introduce the notion of lexicographic min-max (lex-min-max) optimality into the MDL setting—a natural and principled generalization of min-max optimality. Definition 2 below is equivalent to $(1, \varepsilon)$-leximin optimality in Hartman et al. (2023), as well as a solution to Algorithm 1 with error tolerance ε in Abernethy et al. (2024).

Definition 2 (ε-lex-min-max optimality) *Given any $\mathbf{q} \in \Delta(\mathcal{H})$, we define a sequence of constraint sets $\mathcal{Q}_1, \ldots, \mathcal{Q}_k$ as follows: let $\mathcal{Q}_1 := \Delta(\mathcal{H})$ and, for $j \in \{2, \ldots, k\}$, let*

$$\mathcal{Q}_j := \left\{\mathbf{q}'' \in \Delta(\mathcal{H}) \middle| \sigma_m\left(\mathbf{r}_{\mathbf{q}'',\mathcal{D}}\right) \leq \sigma_m\left(\mathbf{r}_{\mathbf{q},\mathcal{D}}\right), \forall m \leq j-1\right\}, \tag{3}$$

which depends on $\mathbf{q}$. With this in mind, and given $\varepsilon \in (0, 1)$, we say that $\mathbf{q}$ is ε-lex-min-max optimal *if, for each $j \in [k]$,*

$$\sigma_j(\mathbf{r}_{\mathbf{q},\mathcal{D}}) \leq \min_{\mathbf{q}'\in\mathcal{Q}_j} \sigma_j\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right) + \varepsilon. \tag{4}$$

To build intuition for this definition, consider a simple case where the number of distributions $k = 2$. Then, for $j = 1$, $\mathbf{r}_{\mathbf{q},\mathcal{D}}$ satisfies:

$$\sigma_1(\mathbf{r}_{\mathbf{q},\mathcal{D}}) \leq \min_{\mathbf{q}'\in\Delta(\mathcal{H})} \sigma_1\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right) + \varepsilon, \tag{5}$$

which is exactly the min-max optimality given in Definition 1. Next, consider $j = 2$. Then Definition 2 further gives a $\mathbf{q} \in \Delta(\mathcal{H})$ which (approximately) minimizes the second largest loss:

$$\sigma_2\left(\mathbf{r}_{\mathbf{q},\mathcal{D}}\right) \leq \min_{\mathbf{q}' \in \mathcal{Q}_2} \sigma_2\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right) + \varepsilon, \tag{6}$$

compared with the policies in:

$$\mathcal{Q}_2 = \left\{ \mathbf{q}'' \in \Delta(\mathcal{H}) \middle| \sigma_1\left(\mathbf{r}_{\mathbf{q}'',\mathcal{D}}\right) \leq \sigma_1\left(\mathbf{r}_{\mathbf{q},\mathcal{D}}\right) \right\},$$

the set of policies which is as good as $\mathbf{q}$ in terms of the maximum risk. In other words, since $\mathbf{q}$ is already approximately optimal for the largest risk as given in $j = 1$, we aim to further make sure $\mathbf{q}$ is approximately optimal in terms of the second largest risk, among all policies that minimize the largest risk. This ensures that optimality of $\mathbf{q}$ at the second level does not compromise its first-level optimality, which is fundamental to the lexicographic optimization principle. This lexicographic refinement naturally extends to larger values of j, progressively tightening control over the entire sorted risk profile across distributions.

Comparison with lexicographic definition of Diana et al. (2021) Note that lexicographic min-max optimality is challenging for MDL, especially because the function $\sigma_i(\cdot)$ is not convex, making both the objective function and the constraint set at each level non-convex. Diana et al. (2021) introduce the following *convexified* lexicographic min-max fairness definition for MDL:

Definition 3 (α-convex-lex-min-max optimality) *Given a target accuracy level $\alpha \in (0, 1)$, let $\epsilon \in (0, \alpha)$ be an error term. For $j \in [k]$, recursively define the following set, where $\mathcal{Q}_1 = \Delta(\mathcal{H})$, and*

$$\mathcal{Q}_j = \left\{ \mathbf{q}' \in \mathcal{Q}_{j-1} \middle| \sum_{n=1}^{j-1} \sigma_n\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right) \leq \min_{\mathbf{q}'' \in \mathcal{Q}_{j-1}} \sum_{n=1}^{j-1} \sigma_n\left(\mathbf{r}_{\mathbf{q}'',\mathcal{D}}\right) + \epsilon \right\} \tag{7}$$

Then, $\mathbf{q} \in \Delta(\mathcal{H})$ is α-convex-lex-min-max optimal if $\forall j \in [k]$,

$$\sum_{n=1}^{j} \sigma_n\left(\mathbf{r}_{\mathbf{q},\mathcal{D}}\right) \leq \min_{\mathbf{q}'' \in \mathcal{Q}_j} \sum_{n=1}^{j} \sigma_n\left(\mathbf{r}_{\mathbf{q}'',\mathcal{D}}\right) + \epsilon + \alpha. \tag{8}$$

Compared with Definition 2, there are several key differences: 1) the j-th largest function in (4) and (3) is replaced by the *sum of the top j largest* risks in (8) and (7) respectively, to ensure the *convexity* of both the objective and the constraints; 2) the leveled constraint set $\mathcal{Q}_j$ is defined in a different nested structure, which is independent of $\mathbf{q}$ itself; 3) an additional error term ϵ is introduced in (7) and (8) to enhance the stability of the results.

These differences lead to several consequences. First, although the convexified objective is easier to optimize, the sum of the top-j risks is a weaker notion than the exact j-th largest risk. This discrepancy becomes particularly less effective when the largest error dominates the others. Moreover, it remains unclear how the convexified formulation relates to the original lexicographic optimality when $\alpha > 0$. Finally, the definition of $\mathcal{Q}_j$ makes the problem almost unsolvable in an online fashion: to optimize this nested objective, one must first gather a sufficiently large dataset and solve the problem level by level, as the constraint set of the lower-level problem depends on the upper-level one. In this paper, we propose an efficient method that can obtain an ε-lex-min-max optimal decision efficiently and in an *online* fashion with standard no-regret dynamics.

3. Lexicographic min-max MDL as a Zero-Sum Game

In this section, we present how to efficiently achieve ε-lex-min-max optimality in MDL. Before introducing our strategy, we first briefly recall how min-max optimality, as defined in Definition 1, is typically achieved. More specifically, to obtain a min-max optimal solution, the goal is to solve the following min-max optimization problem:

$$\min_{\mathbf{q}\in\Delta(\mathcal{H})} \sigma_1(\mathbf{r}_{\mathbf{q},\mathcal{D}}) = \min_{\mathbf{q}\in\Delta(\mathcal{H})} \max_{i\in[k]} R\left(\mathbf{q}, \mathcal{D}^{(i)}\right) = \min_{\mathbf{q}\in\Delta(\mathcal{H})} \max_{i\in[k]} \sum_{h\in\mathcal{H}} \mathbf{q}(h)\mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(i)}}[\ell(h,(\mathbf{x},y))], \tag{9}$$

where $\mathbf{q}(h)$ denotes the probability assigned to hypothesis h of $\mathbf{q}$. The above problem can be equivalently written as

$$\min_{\mathbf{q}\in\Delta(\mathcal{H})} \max_{\mathbf{p}\in\Delta_k} \sum_{i=1}^{k} p_i \sum_{h\in\mathcal{H}} \mathbf{q}(h)\mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(i)}}[\ell(h,(\mathbf{x},y))], \tag{10}$$

where $\mathbf{p}$ is a distribution over the candidate distributions. The equivalence naturally holds as the optimal $\mathbf{p}$ will always assign all weights to the single item with the highest risk. Note that the problem in (10) can be considered as a (stochastic) bi-linear game with two players: a $\mathbf{q}$-player picking a distribution over $\Delta(\mathcal{H})$, while a $\mathbf{p}$-player chooses a distribution over k candidate data distributions. Therefore, one can essentially solve it by applying online learning dynamics for both players. As summarized by Awasthi et al. (2023), *all* existing multi-distribution learning algorithms can be expressed as finding an ε-equilibrium for (10) using no-regret dynamics.

Note that (9) and (10) focus on minimizing the maximum risk, and thus are too "greedy" for achieving more stringent performance metrics like lexicographic min-max optimality. In this paper, we consider a simple yet natural generalized version of (10), given by:

$$\min_{\mathbf{q}\in\Delta(\mathcal{H})} \max_{\mathbf{p}\in\mathcal{P}} c\sum_{i=1}^{k} p_i \sum_{h\in\mathcal{H}} \mathbf{q}(h)\mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(i)}}[\ell(h,(\mathbf{x},y))] - \mathrm{KL}\left(\mathbf{p}\middle\|\frac{\mathbf{1}}{k}\right) := \psi(\mathbf{q},\mathbf{p}), \tag{11}$$

where $\mathbf{1}$ is the k-dimensional all-1 vector, and

$$\mathcal{P} = \left\{\mathbf{p}\in\mathbb{R}_+^k \,\middle|\, \frac{1}{k} \leq p_i \leq \frac{\exp(c)}{k}, \forall i\in[k]\right\}. \tag{12}$$

We note two key differences between our formulation in (11) and the classical min-max objective in (10). First, we introduce a KL-divergence regularizer on the $\mathbf{p}$-player, which encourages a smoother weighting over distributions and avoids the overly greedy behavior of focusing solely on the worst-case distribution. Second, we relax the feasible set of $\mathbf{p}$ from the probability simplex Δ_k to a box constraint, where $c > 0$ is a parameter.

We denote the resulting objective by $\psi(\mathbf{q},\mathbf{p})$. This function is linear in $\mathbf{q}$ and strictly concave in $\mathbf{p}$, making it well-suited for applying online learning dynamics. Importantly, this simple yet principled modification leads to a powerful result: as we show next, simply optimizing this regularized objective suffices to obtain an approximate lex-min-max optimal solution.

Theorem 1 *Suppose there exists a policy $\overline{\mathbf{q}}$ such that*

$$\max_{\mathbf{p}\in\mathcal{P}} \psi(\overline{\mathbf{q}},\mathbf{p}) - \min_{\mathbf{q}\in\Delta(\mathcal{H})} \max_{\mathbf{p}\in\mathcal{P}} \psi(\mathbf{q},\mathbf{p}) \leq g(c),$$

for some function $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$. Then $\overline{\mathbf{q}}$ is $\frac{1}{c}\log(k + kg(c))$-lex-min-max optimal.

Algorithm 1

1: **Input**: step size $\eta > 0$, time period T
2: **Initialization:** $\mathbf{q}_1 \in \Delta(\mathcal{H}), \mathbf{p}_1 = \frac{\mathbf{1}}{k}$
3: **for** $t = 1, \dots, T$ **do**
4: Sample $(\mathbf{x}_t^{(i)}, y_t^{(i)}) \sim \mathcal{D}^{(i)}$ for $i \in [k]$
5: Compute estimated gradients $\widetilde{\nabla}_{\mathbf{q},t}$ and $\widetilde{\nabla}_{\mathbf{p},t}$ based on (13) and (14)
6: Update $\mathbf{q}_{t+1}$ and $\mathbf{p}_{t+1}$ based on (17) and (16)
7: **end for**
8: **Output:** $\overline{\mathbf{q}}_T = \frac{1}{T}\sum_{t=1}^{T} \mathbf{q}_t$

Theorem 1 implies that, $\overline{\mathbf{q}}$ can be $o(1)$-lex-min-max optimal, with appropriate choice of c and $g(c)$. In the next section, we develop an algorithm that ensures an $O(\frac{1}{\log T})$-lex-min-max optimality. The key insight behind Theorem 1 is also remarkably simple: Just as the standard min-max formulation (10) smooths out the max operation in (9), our version in (11) acts as a smooth proxy for the following natural objective[2]:

$$\min_{\mathbf{q} \in \Delta(\mathcal{H})} \sum_{i=1}^{k} \exp\left(c\, R\left(\mathbf{q}, \mathcal{D}^{(i)}\right)\right),$$

which captures the full risk profile across distributions in a lexicographically-aware way. Although this exponential objective is challenging to optimize directly—due to its stochastic compositional structure—our formulation allows it to be solved via standard online learning dynamics, with nearly the same ease as solving the original min-max problem. In essence, our approach delivers stronger lexicographic guarantees without sacrificing simplicity, making it both theoretically appealing and practically implementable.

3.1. Proposed Algorithm for Lexicographic Min-Max MDL

Finally, we present an online learning dynamic for solving the min-max optimization problem given in (11). Following previous work (Zhang et al., 2023), our proposed method, detailed in Algorithm 1, is based on the stochastic mirror descent (SMD) method (Nemirovski et al., 2009).

More specifically, in an idealized setting where the distributions $\{\mathcal{D}^{(i)}\}_{i=1}^{k}$ are fully known, we can compute the full gradients of the objective $\psi(\mathbf{q}, \mathbf{p})$ with respect to both players. Specifically, the gradient with respect to the $\mathbf{q}$-player is:

$$\nabla_{\mathbf{q}}\psi(\mathbf{p}, \mathbf{q}) = \left[c\sum_{i=1}^{k} p^{(i)} \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(i)}}[\ell(h_1, (\mathbf{x}, y))], \dots, c\sum_{i=1}^{k} p^{(i)} \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(i)}}[\ell(h_{|\mathcal{H}|}, (\mathbf{x}, y))]\right],$$

and the gradient with respect to the $\mathbf{p}$-player is given by:

$$\nabla_{\mathbf{p}}\psi(\mathbf{p}, \mathbf{q}) = \left[c\sum_{h\in\mathcal{H}} \mathbf{q}(h) \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(1)}}[\ell(h, (\mathbf{x}, y))], \dots, c\sum_{h\in\mathcal{H}} \mathbf{q}(h) \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(k)}}[\ell(h, (\mathbf{x}, y))]\right] - [\log kp_1, \dots, \log kp_k].$$

2. We note that in (12), we relax the feasible set of $\mathbf{p}$ from the probability simplex to a box constraint. Consequently, the objective becomes a sum of exponentials rather than a log-sum-exp. This distinction is important because the log-sum-exp formulation may introduce a multiplicative error in the final approximate optimum, whereas our Definition 3 is based on an additive error criterion.

However, in reality, the distributions are not known in advance. Instead, we assume access only to samples drawn from each $\mathcal{D}^{(i)}$. In each round t, we first obtain the stochastic sample $(\mathbf{x}_t^{(i)}, y_t^{(i)}) \sim \mathcal{D}^{(i)}$ from each distribution respectively (Step 4) (see Appendix B.2 for an extension that requires a sample from only one distribution per round). After that, we can compute the estimated gradients as follows (Step 5):

$$\widetilde{\nabla}_{\mathbf{q},t} = \left[c\sum_{i=1}^{k} p_{t,i}[\ell(h_1, (\mathbf{x}_t^{(i)}, y_t^{(i)}))], \ldots, c\sum_{i=1}^{k} p_{t,i}[\ell(h_{|\mathcal{H}|}, (\mathbf{x}_t^{(i)}, y_t^{(i)}))] \right], \tag{13}$$

and

$$\begin{aligned}\widetilde{\nabla}_{\mathbf{p},t} = & \left[c\sum_{h\in\mathcal{H}} \mathbf{q}_t(h)[\ell(h, (\mathbf{x}_t^{(1)}, y_t^{(1)}))], \ldots, c\sum_{h\in\mathcal{H}} \mathbf{q}_t(h)[\ell(h, (\mathbf{x}_t^{(k)}, y_t^{(k)}))] \right] \\ & - [\log kp_{t,1}, \ldots, \log kp_{t,k}],\end{aligned} \tag{14}$$

where $\mathbf{p}_t$ and $\mathbf{q}_t$ are the decisions from the previous round. Note that these gradients are unbiased:

$$\mathbb{E}[\widetilde{\nabla}_{\mathbf{q},t}|\mathbf{p}_t, \mathbf{q}_t] = \nabla_{\mathbf{q}}\psi(\mathbf{p}_t, \mathbf{q}_t), \text{ and } \mathbb{E}[\widetilde{\nabla}_{\mathbf{p},t}|\mathbf{p}_t, \mathbf{q}_t] = \nabla_{\mathbf{p}}\psi(\mathbf{p}_t, \mathbf{q}_t).$$

With these unbiased stochastic gradients, we update the decisions with mirror descent:

$$\mathbf{q}_{t+1} = \operatorname*{argmin}_{\mathbf{q}\in\Delta(\mathcal{H})} \eta \left\langle \mathbf{q}, \widetilde{\nabla}_{\mathbf{q},t} \right\rangle + B_\rho(\mathbf{q}, \mathbf{q}_t), \tag{15}$$

and

$$\mathbf{p}_{t+1} = \operatorname*{argmin}_{\mathbf{p}\in\mathcal{P}} -\eta \left\langle \mathbf{p}, \widetilde{\nabla}_{\mathbf{p},t} \right\rangle + B_\phi(\mathbf{p}, \mathbf{p}_t). \tag{16}$$

Here, B is the Bregman divergence, which is defined in (1), η is the step size which well be configured later. For the $\mathbf{q}$ player, we set ρ as negative entropy: $\rho(\mathbf{q}) = \sum_{h\in\mathcal{H}} \mathbf{q}(h)\log\mathbf{q}(h)$. In this way, we do not need to solve the optimization problem in (15) directly, instead, we can update $\mathbf{q}_{t+1}$ with the classical Hedge update. For all $h \in \mathcal{H}$,

$$\mathbf{q}_{t+1}(h) = \frac{\mathbf{q}_t(h)\exp\left(-\eta c\sum_{i=1}^{k} p_{t,i}[\ell(h, (\mathbf{x}_t^{(i)}, y_t^{(i)}))]\right)}{\sum_{h'\in\mathcal{H}} \mathbf{q}_t(h')\exp\left(-\eta c\sum_{i=1}^{k} p_{t,i}[\ell(h', (\mathbf{x}_t^{(i)}, y_t^{(i)}))]\right)}. \tag{17}$$

On the other hand, for the $\mathbf{p}$-player, we simply set $\phi(\mathbf{p}) = \frac{1}{2}\|\mathbf{p}\|_2^2$. After T iterations, we output the average of $\mathbf{q}_t$, which is denoted as $\overline{\mathbf{q}}_T$ (Step 9). Following the classical theoretical guarantees of stochastic mirror descent (Nemirovski et al., 2009; Zhang et al., 2023), we have the following conclusion for Algorithm 1. The proof is given in Appendix A.

Lemma 1 *For Algorithm 1, let* $\eta = \frac{\ln|\mathcal{H}|}{2k\sqrt{10T}}$*. Then, with probability at least* $1-\delta$*, we have*

$$\max_{\mathbf{p}\in\mathcal{P}} \psi(\overline{\mathbf{q}}_T, \mathbf{p}) - \min_{\mathbf{q}\in\Delta(\mathcal{H})} \max_{\mathbf{p}\in\mathcal{P}} \psi(\mathbf{q}, \mathbf{p}) = O\left(\frac{(\exp(c))^2 \log|\mathcal{H}| \log\frac{1}{\delta}}{\sqrt{T}} \right).$$

Combining with Theorem 1, we obtain the following conclusion.

Theorem 2 *Let $c = \log T^{1/4}$. Then with a probability $1-\delta$, the output of Algorithm 1, $\overline{\mathbf{q}}_T$, is $O\left(\frac{1}{\log T}\log\left(k + k\log|\mathcal{H}|\log\frac{1}{\delta}\right)\right)$-lex-min-max optimal.*

The theoretical guarantees provided by Theorem 2 show that Algorithm 1 achieves $O\big(\frac{\log(k\log(|\mathcal{H}|)\log(1/\delta))}{\log T}\big)$-min-max optimality, which converges to 0 as $T\to\infty$. While this rate is slower than the $O(1/\sqrt{T})$ rate of the methods for min-max MDL, it does so across all hierarchy levels simultaneously. On the other hand, the dependence on $|\mathcal{H}|$ is reduced to $\log\log(|\mathcal{H}|)$ from $\log(|\mathcal{H}|)$. Finally, we note that in this section we consider the setting with a finite hypothesis class for simplicity. However, our method can be extended to other MDL settings—such as adaptive sampling and infinite hypothesis classes (Haghtalab et al., 2022)—by modifying the corresponding min-max objective to our regularized version and applying corresponding no-regret methods for the $\mathbf{q}$-player. See Appendix B.

4. Theoretical Analysis

In this section, we provide the main proof for Theorem 1. Firstly, we have the following observation:

Lemma 2 *For any $\mathbf{q}\in\Delta(\mathcal{H})$, we have*

$$\max_{\mathbf{p}\in\mathcal{P}}\psi(\mathbf{q},\mathbf{p}) = \frac{1}{k}\sum_{i=1}^{k}\exp\left(c\sum_{h\in\mathcal{H}}\mathbf{q}(h)\mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(i)}}[\ell(h,(\mathbf{x},y))]\right) - 1.$$

Proof Note that ψ is strictly concave with respect to $\mathbf{p}$, as it includes a negative Bregman divergence. Based on the first-order optimality condition, since

$$\psi(\mathbf{q},\mathbf{p}) = c\sum_{i=1}^{k}p_i\sum_{h\in\mathcal{H}}\mathbf{q}(h)\mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(i)}}[\ell(h,(\mathbf{x},y))] - \sum_{i=1}^{k}p_i\log p_i - \sum_{i=1}^{k}p_i\log k + \sum_i p_i - 1, \quad (18)$$

setting

$$\frac{\partial\psi}{\partial p_i} = c\sum_{h\in\mathcal{H}}\mathbf{q}(h)\mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(i)}}[\ell(h,(\mathbf{x},y))] - \log p_i - \log k = 0,$$

we could obtain a closed form for the optimal solution $\mathbf{p}^*$:

$$\begin{aligned} p_i^* &= \exp\left(c\sum_{h\in\mathcal{H}}\mathbf{q}(h)\mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(i)}}[\ell(h,(\mathbf{x},y))] - \log k\right) \\ &= \frac{\exp\left(c\sum_{h\in\mathcal{H}}\mathbf{q}(h)\mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(i)}}[\ell(h,(\mathbf{x},y))]\right)}{k}. \end{aligned}$$

Note that, since the loss are bounded in $[0,1]$, we know $\frac{1}{k}\le p_i^*\le\frac{\exp(c)}{k}$, which is inside the feasible region. Substituting $\mathbf{p}^*$ back to (18), we get

$$\begin{aligned} &\psi(\mathbf{q},\mathbf{p}^*) \\ &= \frac{c}{k}\sum_{i=1}^{k}\exp(c\Gamma_i)\Gamma_i - \frac{1}{k}\sum_{i=1}^{k}\exp(c\Gamma_i)(c\Gamma_i - \log k) - \frac{1}{k}\sum_{i=1}^{k}\exp(c\Gamma_i)\log k + \frac{1}{k}\sum_{i=1}^{k}\exp(c\Gamma_i) - 1 \\ &= \frac{1}{k}\exp(c\Gamma_i) - 1, \end{aligned}$$

where $\Gamma_i = \sum_{h\in\mathcal{H}} \mathbf{q}(h)\mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}^{(i)}}[\ell(h,(\mathbf{x},y))]$. ■

The lemma above indicates that, if we obtain a policy $\overline{\mathbf{q}}$ such that

$$\max_{\mathbf{p}\in\mathcal{P}} \psi(\overline{\mathbf{q}},\mathbf{p}) - \min_{\mathbf{q}\in\Delta(\mathcal{H})}\max_{\mathbf{p}\in\mathcal{P}} \psi(\mathbf{q},\mathbf{p}) \leq g(c),$$

we immediately have:

$$\sum_{i=1}^{k} \exp\left(cR(\overline{\mathbf{q}},\mathcal{D}^{(i)})\right) - \min_{\mathbf{q}\in\Delta(\mathcal{H})}\sum_{i=1}^{k}\exp\left(cR(\mathbf{q},\mathcal{D}^{(i)})\right) \leq kg(c).$$

Note that this also can be written as

$$\sum_{n=1}^{k} \exp\left(c\sigma_n\left(\mathbf{r}_{\overline{\mathbf{q}}_T,\mathcal{D}}\right)\right) - \min_{\mathbf{q}\in\Delta(\mathcal{H})}\sum_{n=1}^{k}\exp\left(c\sigma_n\left(\mathbf{r}_{\mathbf{q},\mathcal{D}}\right)\right) \leq kg(c),$$

Since introducing the ordered sum does not change the results of the sum overall. Next, consider the j-th level, for $j \in \{1,\ldots,k\}$. For any $\mathbf{q}' \in \mathcal{Q}_j$, where $\mathcal{Q}_j$ is defined in (3), we have for $m \in \{1,\ldots,j-1\}$, $\sigma_m\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right) \leq \sigma_m\left(\mathbf{r}_{\overline{\mathbf{q}}_T,\mathcal{D}}\right)$, which implies that

$$\exp\left(c\sigma_m\left(c\mathbf{r}_{\overline{\mathbf{q}}_T,\mathcal{D}}\right)\right) - \exp\left(c\sigma_m\left(\mathbf{r}_{\mathbf{q}'\mathcal{D}}\right)\right) \geq 0 \tag{19}$$

Therefore, we know

$$\begin{aligned}&\sum_{n=j}^{k} \exp\left(c\sigma_n\left(\mathbf{r}_{\overline{\mathbf{q}}_T,\mathcal{D}}\right)\right) - \sum_{n=j}^{k} \exp\left(c\sigma_n\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right)\right) \leq \sum_{n=1}^{k} \exp\left(c\sigma_n\left(\mathbf{r}_{\overline{\mathbf{q}}_T,\mathcal{D}}\right)\right) - \sum_{n=1}^{k} \exp\left(c\sigma_n\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right)\right)\\ &\leq \sum_{n=1}^{k} \exp\left(c\sigma_n\left(\mathbf{r}_{\overline{\mathbf{q}}_T,\mathcal{D}}\right)\right) - \min_{\mathbf{q}'\in\Delta(\mathcal{H})}\sum_{n=1}^{k} \exp\left(c\sigma_n\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right)\right) \leq kg(c),\end{aligned}$$

where the first inequality is based on the definition of $\mathcal{Q}_j$. To proceed, we have

$$\exp\left(c\sigma_j\left(\mathbf{r}_{\overline{\mathbf{q}}_T,\mathcal{D}}\right)\right) - k\exp\left(c\sigma_j\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right)\right) \leq \sum_{n=j}^{k} \exp\left(c\sigma_n\left(\mathbf{r}_{\overline{\mathbf{q}}_T,\mathcal{D}}\right)\right) - \sum_{n=j}^{k} \exp\left(c\sigma_n\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right)\right) \leq kg(c),$$

where the first inequality is because $\exp\left(c\sigma_m\left(\mathbf{r}_{\overline{\mathbf{q}}_T,\mathcal{D}}\right)\right) \geq 0$ for all $m \in [k]$, and $\exp\left(c\sigma_j\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right)\right) \geq \exp\left(c\sigma_m\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right)\right)$, for $m > j$. We have

$$\exp\left(c\sigma_j\left(\mathbf{r}_{\overline{\mathbf{q}}_T,\mathcal{D}}\right) - c\sigma_j\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right)\right) - k \leq \frac{kg(c)}{\exp\left(c\sigma_j\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right)\right)} \leq kg(c), \tag{20}$$

and thus

$$\sigma_j\left(\mathbf{r}_{\overline{\mathbf{q}}_T,\mathcal{D}}\right) - \sigma_j\left(\mathbf{r}_{\mathbf{q}',\mathcal{D}}\right) \leq \frac{1}{c}\log\left(k + kg(c)\right).$$

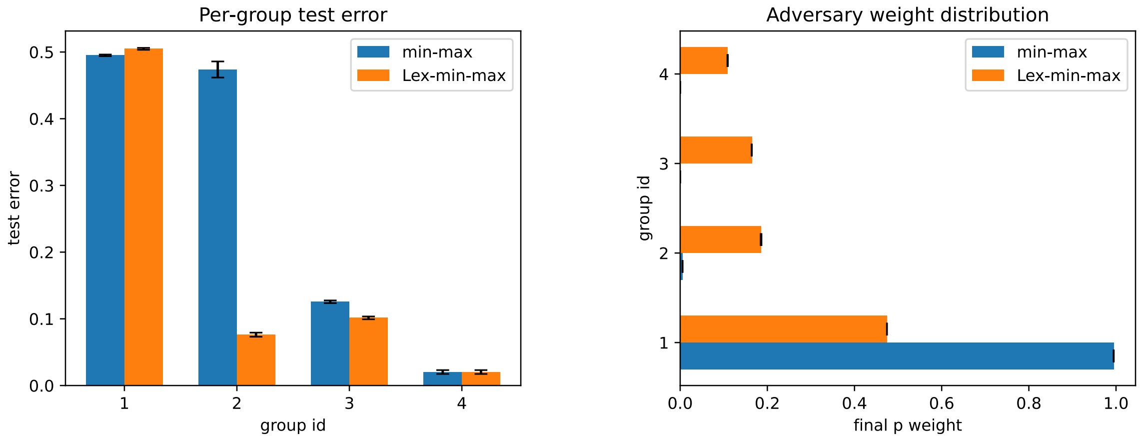

Figure 1: Experimental results on the UCI Adult dataset. Left: Per-group test errors show that lex-min-max achieves more balanced performance across groups. Right: *Group weights* learned after training, which reveals that the lex-min-max method avoids over-focusing on the hardest group.

5. Experiments

In this section, we conduct experiments on both synthetic and real-world data sets to demonstrate the effectiveness of our proposed objective. As a benchmark, we compare with basic min-max algorithm trained with no-regret dynamics (Soma et al., 2022). Note that there exist various advanced MDL methods, but they all focus on the sample-complexity for solving the same min-max objective (Sagawa et al., 2020; Haghtalab et al., 2022; Zhang et al., 2023; Awasthi et al., 2023; Peng, 2024), whereas our focus is a fairer, level-aware lexicographic objective. Thus, we contrast chiefly with the classical min-max baseline to highlight the gains from lex-min-max.

5.1. Experiments on real-world data set

Setup Following previous work (Namkoong and Duchi, 2016; Sagawa et al., 2020), we conduct experiments on the UCI ADULT income data set (Becker and Kohavi, 1996). We retain all records with complete feature information ($\{\approx 30\,\text{k}\}$ examples), binarize the label as $\mathbb{I}\{\text{income} \geq 50\text{K}\}$, and construct four demographic groups by crossing binary sex with educational attainment, yielding 1: Male–HighEdu, 2: Male–LowEdu, 3: Female–HighEdu, 4: Female–LowEdu. We also make one major modification to the dataset: we *corrupt* exactly one of the groups (group 1 in Figure 1) by resetting each label to be an independent Bernoulli. This perturbation of the task is both (a) very natural in practice, as one frequently encounters hard-to-predict groups within multi-distribution settings; and (b) it helps elucidate why the lex-min-max is a superior objective to the classical min-max, which only aims to achieve low risk on the hardest group at the expense of groups that exhibit more favorable prediction tasks. We perform a stratified 70/30 train–test split. Training runs for $T = 8000$ steps, with step sizes $\eta = 5/\sqrt{T}$ for both lex-min-max and min-max methods. We set $c = \log T^{1/4}$ as prescribed by Theorem 2. To accommodate the parameterized model, we use offline

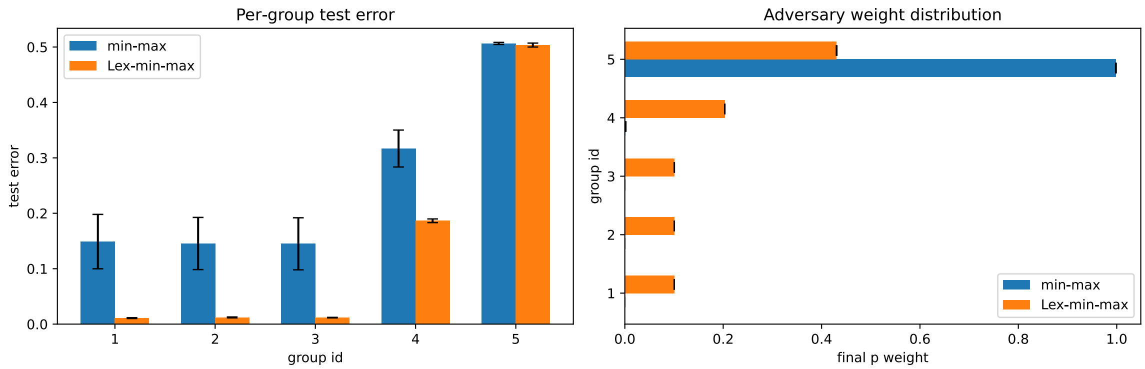

Figure 2: Results on Synthetic data

gradient descent for the p-player instead of Hedge in both algorithms. We run each experiments 10 times.

Results The final test error is shown in Figure 2 (left). As expected, both lex-min-max and min-max methods incur high error on the hard group. However, the lex-min-max method achieves lower test error on the remaining, easier groups, especially group 2. Figure 2 (right) presents the final weights assigned to each group. It can be observed that, due to the large error on the worst-case group, the min-max method concentrates most of the weight on this group—effectively ignoring the others. In contrast, the lex-min-max method distributes weights more evenly in a lexicographically-informed manner, preserving information from all groups.

5.2. Experiments on synthetic data

Following previous work on MDL (Namkoong and Duchi, 2016; Soma et al., 2022), we create five data-generating distributions $\{\mathcal{D}^{(i)}\}_{i=1}^{5}$ in $\mathbb{R}^2$; each distribution is a balanced *two-component Gaussian mixture* with shared isotropic covariance and class-specific means:

$$\mathcal{D}^{(i)}:\ y \sim \mathrm{Unif}\{+1,-1\}, \qquad \mathbf{x} \mid y \sim \mathcal{N}\big(y\,[d_i - o_i,\, 0]^\top,\, I_2\big),$$

where d_i controls the class separation of group i:

$$d_1 = d_2 = d_3 = 3.0\ \ (\text{easy groups}), \qquad d_4 = 1.0\ \ (\text{medium group}), \qquad d_5 = 0.1\ \ (\text{hard group}),$$

and o_i controls the offset from the y-axis:

$$o_1 = o_2 = o_3 = 1.0\ \ (\text{easy groups}), \qquad o_4 = 0.5\ \ (\text{medium group}), \qquad o_5 = -0.5\ \ (\text{hard group}),$$

For the hard group, we additionally flip the class label with probability 0.5, introducing label noise. The remaining four groups are noise-free. Each group contributes 10,000 training and 3,000 test samples. For this task, we train a linear classifier using the logistic loss. Training runs for $T = 8000$ steps, with step sizes $\eta = \frac{5}{\sqrt{T}}$ for both lex-min-max and min-max methods. We set $c = \log T^{1/4}$ as prescribed by the theory. To accommodate the parameterized model, we use gradient descent for the p-player instead of Hedge in both algorithms. The final results is shown in Figure 2.

6. Conclusion and Future Work

In this paper, we studied multi-distribution learning, and introduced approximate lex-min-max optimality as the performance metric. Unlike min-max optimality, which focuses only on the worst-case group, lex-min-max optimality considers all groups in a lexicographic order. We showed that this objective can be achieved with similar ease as min-max optimality, by simply solving a regularized min-max game.

Our paper takes a first step toward understanding full-distributions aware performance guarantees in MDL and raises several open questions. First, we do not know whether the $O\left(\frac{1}{\log T}\right)$ convergence rate of our algorithm can be improved, or whether a matching lower bound can be established. Second, it is important to validate the proposed methods with more empirical experiments.

References

Jacob Abernethy, Robert Schapire, and Umar Syed. Lexicographic optimization: Algorithms and stability. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, 2024.

Pranjal Awasthi, Nika Haghtalab, and Eric Zhao. Open problem: The sample complexity of multi-distribution learning for vc classes. In *Proceedings of 36th Conference on Learning Theory*, 2023.

Barry Becker and Ronny Kohavi. Adult. UCI Machine Learning Repository, 1996.

Fred Alois Behringer. Lexicographic quasiconcave multiobjective programming. *Zeitschrift für Operations Research*, 1977.

Avrim Blum, Nika Haghtalab, Ariel D Procaccia, and Mingda Qiao. Collaborative pac learning. *In Advances in Neural Information Processing Systems 30*, 2017.

Avrim Blum, Nika Haghtalab, Richard Lanas Phillips, and Han Shao. One for one, or all for all: Equilibria and optimality of collaboration in federated learning. In *Proceedings of the 38th International Conference on Machine Learning*, 2021a.

Avrim Blum, Shelby Heinecke, and Lev Reyzin. Communication-aware collaborative learning. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, 2021b.

Jiecao Chen, Qin Zhang, and Yuan Zhou. Tight bounds for collaborative pac learning via multiplicative weights. In *Advances in the 31st Neural Information Processing Systems*, 2018.

Emily Diana, Wesley Gill, Ira Globus-Harris, Michael Kearns, Aaron Roth, and Saeed Sharifi-Malvajerdi. Lexicographically fair learning: Algorithms and generalization. In *Proceedings of the 2nd Symposium on Foundations of Responsible Computing*, 2021.

Nika Haghtalab, Michael Jordan, and Eric Zhao. On-demand sampling: Learning optimally from multiple distributions. *In Advances in Neural Information Processing Systems 35*, 2022.

Eden Hartman, Avinatan Hassidim, Yonatan Aumann, and Erel Segal-Halevi. Leximin approximation: From single-objective to multi-objective. In *ECAI 2023*, pages 996–1003. IOS Press, 2023.

Howard Paul Hayden. *Voice flow control in integrated packet networks*. PhD thesis, Massachusetts Institute of Technology, 1981.

Elon Kohlberg. The nucleolus as a solution of a minimization problem. *SIAM Journal on Applied Mathematics*, 1972.

Jean-Yves Le Boudec. Rate adaptation, congestion control and fairness: A tutorial. *Web page, November*, 2005.

Xi Lin, Xiaoyuan Zhang, Zhiyuan Yang, Fei Liu, Zhenkun Wang, and Qingfu Zhang. Smooth tchebycheff scalarization for multi-objective optimization. In *Proceedings of the 41st International Conference on Machine Learning*, 2024.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 2017.

Jeonghoon Mo and Jean Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on networking*, 2000.

Mehryar Mohri, Gary Sivek, and Ananda Theertha Suresh. Agnostic federated learning. In *Proceedings of the 36th International Conference on Machine Learning*, 2019.

Mor Shpigel Nacson, Suriya Gunasekar, Jason Lee, Nathan Srebro, and Daniel Soudry. Lexicographic and depth-sensitive margins in homogeneous and non-homogeneous deep models. In *International Conference on Machine Learning*, pages 4683–4692. PMLR, 2019.

Hongseok Namkoong and John C Duchi. Stochastic gradient methods for distributionally robust optimization with f-divergences. In *Advances in Neural Information Processing Systems 29*, 2016.

Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on optimization*, 2009.

Włodzimierz Ogryczak. On the lexicographic minimax approach to location problems. *European Journal of Operational Research*, 1997.

Włodzimierz Ogryczak and Tomasz Śliwiński. On direct methods for lexicographic min-max optimization. In *International Conference on Computational Science and Its Applications*, 2006.

Binghui Peng. The sample complexity of multi-distribution learning. In *Proceedings of the 37th Annual Conference on Learning Theory*, 2024.

Shiori Sagawa, Pang Wei Koh*, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust neural networks. In *International Conference on Learning Representations*, 2020.

Tasuku Soma, Khashayar Gatmiry, and Stefanie Jegelka. Optimal algorithms for group distributionally robust optimization and beyond. *arXiv preprint arXiv:2212.13669*, 2022.

Ronald R Yager. On the analytic representation of the leximin ordering and its application to flexible constraint propagation. *European Journal of Operational Research*, 1997.

Lijun Zhang, Peng Zhao, Zhen-Hua Zhuang, Tianbao Yang, and Zhi-Hua Zhou. Stochastic approximation approaches to group distributionally robust optimization. In *Advances in Neural Information Processing Systems 36*, 2023.

Zihan Zhang, Wenhao Zhan, Yuxin Chen, Simon S Du, and Jason D Lee. Optimal multi-distribution learning. In *Proceedings of the 37th Annual Conference on Learning Theory*, 2024.

Appendix A. Proof of Lemma 1

Note that Lemma 1 is a standard conclusion of applying stochastic mirror descent (Nemirovski et al., 2009) for solving a stochastic min-max optimization problem (Zhang et al., 2023). In the following, we include the proof for completeness. To apply the conclusion, we only need to make sure that the boundedness assumption on the feasible domain and gradients are satisfied (Zhang et al., 2023).

Firstly, we verify the domain of $\mathbf{p}$ and $\mathbf{q}$ are bounded with respect to the potential function ϕ and ρ of the Bregman divergence, respectively. Recall that ρ is the negative entropy, and $\phi(\mathbf{p}) = \frac{1}{2}\|\mathbf{p}\|_2^2$. Thus, we have

$$\max_{\mathbf{q}\in\Delta(\mathcal{H})} \rho(\mathbf{q}) - \min_{\mathbf{q}\in\Delta(\mathcal{H})} \rho(\mathbf{q}) \leq 0 + \ln|\mathcal{H}| = \ln|\mathcal{H}|,$$

and

$$\max_{\mathbf{p}\in\mathcal{P}} \phi(\mathbf{p}) - \min_{\mathbf{p}\in\mathcal{P}} \phi(\mathbf{p}) \leq k \max_{\mathbf{p}\in\mathcal{P}} \frac{1}{2}\|\mathbf{p}\|_\infty^2 - \frac{1}{2}\min_{\mathbf{p}\in\mathcal{P}} \|\mathbf{p}\|_2^2 \leq \frac{(\exp(c))^2}{k}.$$

Let $D^2 = \frac{(\exp(c))^2 \ln|\mathcal{H}|}{k}$, and it can be easily seen that both terms above are bounded by D^2 for $|\mathcal{H}| > 2$. Next, we show that the gradients are bounded with respect to the dual norm introduced by ρ and ϕ. More specifically, ρ is strongly convex with respect to ℓ_1 norm, thus the dual norm of the gradient

$$\left\|\widetilde{\nabla}_{\mathbf{q},t}\right\|_\infty \leq \frac{c\exp(c)}{k},$$

on the other hand, ϕ is strongly convex with respect to ℓ_2 norm, thus

$$\left\|\widetilde{\nabla}_{\mathbf{p},t}\right\|_2 \leq c + \sqrt{\sum_{i=1}^{k}(\log kp_i)^2} \leq c + \sqrt{k}c.$$

Similarly, let $G = 4\sqrt{k}\exp(c)$. Then both terms above are bounded by G for $c > 0$. With these bounds, we invoke the standard SMD analysis (Nemirovski et al., 2009; Zhang et al., 2023). We use the following norms over the domain $\mathcal{X} := \Delta_k \times \mathcal{P}$:

$$\|[\mathbf{q};\mathbf{p}]\| = \frac{1}{\sqrt{2}D}\sqrt{\|\mathbf{p}\|_2^2 + \|\mathbf{q}\|_1^2}, \text{ and } \|[\mathbf{u};\mathbf{v}]\|_* = \sqrt{2}D\sqrt{\|\mathbf{u}\|_2^2 + \|\mathbf{v}\|_\infty^2}.$$

Define

$$\nu(\mathbf{x}) = \nu([\mathbf{q};\mathbf{p}]) = \frac{2}{D^2}(\phi(\mathbf{p}) + \rho(\mathbf{q})).$$

Then $\nu(\mathbf{x})$ is 1-strongly convex with respect to $\|\cdot\|$. It can be easily verify that the Bregman divergence defined by ν is given by:

$$B_\nu(\mathbf{x},\mathbf{x}') = \frac{2}{D^2}\left(B_\rho(\mathbf{q},\mathbf{q}') + B_\phi(\mathbf{p},\mathbf{p}')\right).$$

Consider the following MD updates:

$$\mathbf{x}_{t+1} = \operatorname*{argmin}_{\mathbf{x}\in\mathcal{X}} \left\langle \eta_\nu[\widetilde{\nabla}_{\mathbf{q},t}, -\widetilde{\nabla}_{\mathbf{p},t}], \mathbf{x} - \mathbf{x}_t \right\rangle + B_\nu(\mathbf{x},\mathbf{x}_t).$$

Note that it is equivalent to Algorithm 1 with $\eta = 2\eta_\nu D^2$. Moreover, the dual norm of the gradients:

$$\|[\widetilde{\nabla}_{\mathbf{q},t}; -\widetilde{\nabla}_{\mathbf{p},t}]\|_* \leq \sqrt{2}DG.$$

Algorithm 2

1: **Input**: step size $\eta > 0$, time period T
2: **Initialization:** $\mathbf{w}_1 \in \mathcal{W}, \mathbf{p}_1 = \frac{\mathbf{1}}{k}$
3: **for** $t = 1, \ldots, T$ **do**
4: For each $i \in [k]$, obtain $R^{(i)}(\mathbf{w}_t; \mathbf{z}_t^{(i)})$ and $\nabla R^{(i)}(\mathbf{w}_t; \mathbf{z}_t^{(i)})$
5: Compute $\widetilde{\nabla}_{\mathbf{w},t} = \sum_{i=1}^{k} p_{t,i} \nabla R^{(i)}(\mathbf{w}_t; \mathbf{z}_t^{(i)})$, and
$\widetilde{\nabla}_{\mathbf{p},t} = [R^{(1)}(\mathbf{w}_t; \mathbf{z}_t^{(1)}), \ldots, R^{(n)}(\mathbf{w}_t; \mathbf{z}_t^{(n)})] - [\log kp_1, \ldots, \log kp_k]$
6: Update $\mathbf{w}_{t+1}$ and $\mathbf{p}_{t+1}$ by:

$$\mathbf{w}_{t+1} = \operatorname*{argmin}_{\mathbf{w} \in \mathcal{W}} \eta \left\langle \widetilde{\nabla}_{\mathbf{w},t}, \mathbf{w} \right\rangle + B_\phi(\mathbf{w}, \mathbf{w}_t)$$
$$\mathbf{p}_{t+1} = \operatorname*{argmin}_{\mathbf{p} \in P} \eta \left\langle \widetilde{\nabla}_{\mathbf{p},t}, \mathbf{w} \right\rangle + B_\phi(\mathbf{p}, \mathbf{p}_t)$$

7: **end for**
8: **Output:** $\overline{\mathbf{w}}_T = \frac{1}{T} \sum_{t=1}^{T} \mathbf{w}_t$

Thus, the algorithm can considered as applying standard SMD for a convex problem with bounded domain and gradients. The final results follows Lemma 1 of (Zhang et al., 2023) by setting the step size $\eta = \frac{2}{DG\sqrt{10T}}$.

Appendix B. Extensions

In this setting, we show the flexibility of the lexicographic min-max MDL framework by extending it to the infinite hypophysis setting and also incorporate the on-demand sampling technique.

B.1. Infinite hypophysis setting with parameterization

In this section, we study the case where the hypothesis class is infinite. Following prior work (Soma et al., 2022; Zhang et al., 2023), we consider a parametrized setting. Specifically, each hypothesis is associated with a parameter vector $\mathbf{w} \in \mathcal{W} \subseteq \mathbb{R}^d$, where $\mathcal{W}$ is a convex set. For each objective $i \in [k]$, let $R^{(i)}(\mathbf{w})$ denote the risk corresponding to the i-th objective. We introduce the following standard assumption.

Assumption 1 *Without loss of generality, assume $\mathcal{W}$ is bounded: $\max_{\mathbf{w} \in \mathcal{W}} \|\mathbf{w}\|_2 \leq 1$. Moreover, assume for $i \in [k]$, $R^{(i)}(\mathbf{w})$ is convex, and have bounded function value and gradients. More specifically, assume $0 \leq R^{(i)}(\mathbf{w}) \leq 1$ and $\max_{\mathbf{w} \in \mathcal{W}} \|\nabla R^{(i)}(\mathbf{w})\|_2 \leq 1$.*

Similar as before, we collect the risks into the risk vector

$$R_{\mathbf{w}} = \left[R^{(1)}(\mathbf{w}), \ldots, R^{(k)}(\mathbf{w})\right].$$

We then extend the notion of ϵ-lex-min-max optimality to this parametrized setting.

Definition 4 (ε-lex-min-max optimality) *Given any $\mathbf{w} \in \mathcal{W}$, we define a sequence of constraint sets $\mathcal{Q}_1, \ldots, \mathcal{Q}_k$ as follows: let $\mathcal{Q}_1 := \mathcal{W}$ and, for $j \in \{2, \ldots, k\}$, let*

$$\mathcal{Q}_j := \left\{ \mathbf{w}'' \in \mathcal{W} \middle| \sigma_m\left(R_{\mathbf{w}''}\right) \leq \sigma_m\left(R_{\mathbf{w}}\right), \forall m \leq j-1 \right\}.$$

With this in mind, and given $\varepsilon \in (0,1)$*, we say that* $\mathbf{w}$ *is* ε-lex-min-max optimal *if, for each* $j \in [k]$*,*

$$\sigma_j\left(R_{\mathbf{w}}\right) \leq \min_{\mathbf{w}' \in \mathcal{Q}_j} \sigma_j\left(R_{\mathbf{w}'}\right) + \varepsilon.$$

Following (11), we introduce the following objective function:

$$\min_{\mathbf{w}\in\mathcal{W}} \max_{\mathbf{p}\in\mathcal{P}} h(\mathbf{w},\mathbf{p}) = \min_{\mathbf{w}\in\mathcal{W}} \max_{\mathbf{p}\in\mathcal{P}} c\sum_{i=1}^{k} p_i R^{(i)}(\mathbf{w}) - \mathrm{KL}\left(\mathbf{p}\middle\|\frac{\mathbf{1}}{k}\right), \tag{21}$$

where $\mathcal{P}$ is given in (12). We have the following results, which is the counterpart of Theorem 1.

Theorem 3 (Lex-min-max optimality in infinite hypothesis setting) *Suppose Assumption 1 holds, and an algorithm outputs a policy parameter* $\overline{\mathbf{w}}_T(c)$ *after* T *iterations such that*

$$\max_{\mathbf{p}\in\mathcal{P}} h(\overline{\mathbf{w}}_T(c),\mathbf{p}) \;-\; \min_{\mathbf{w}\in\mathcal{W}} \max_{\mathbf{p}\in\mathcal{P}} h(\mathbf{w},\mathbf{p}) \;\leq\; g(c), \tag{22}$$

for some function $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$*. Then* $\overline{\mathbf{w}}_T(c)$ *is* $\frac{1}{c}\log\left(k + kg(c)\right)$*-lex-min-max optimal with respect to Definition 4.*

Proof The proof is basically the same as that of Theorem 1, and we provide the proof for completeness. We first eliminate $\mathbf{p}$ from the regularized game. Consider

$$h(\mathbf{w},\mathbf{p}) = c\sum_{i=1}^{k} p_i R^{(i)}(\mathbf{w}) - \mathrm{KL}\Big(\mathbf{p}\Big\|\frac{\mathbf{1}}{k}\Big), \qquad \mathcal{P} = \Big\{\mathbf{p}\in\mathbb{R}_+^k : \frac{1}{k} \leq p_i \leq \frac{e^c}{k}\Big\}.$$

Writing

$$\mathrm{KL}\Big(\mathbf{p}\Big\|\frac{\mathbf{1}}{k}\Big) = \sum_{i=1}^{k} p_i \log p_i + \sum_{i=1}^{k} p_i \log k - \sum_{i=1}^{k} p_i + 1,$$

we have

$$h(\mathbf{w},\mathbf{p}) = c\sum_{i=1}^{k} p_i R^{(i)}(\mathbf{w}) - \sum_{i=1}^{k} p_i \log p_i - \sum_{i=1}^{k} p_i \log k + \sum_{i=1}^{k} p_i - 1.$$

The first-order optimality condition gives

$$\frac{\partial h}{\partial p_i} = cR^{(i)}(\mathbf{w}) - \log p_i - \log k = 0 \quad\implies\quad p_i^*(\mathbf{w}) = \frac{\exp(cR^{(i)}(\mathbf{w}))}{k}.$$

Since $R^{(i)}(\mathbf{w}) \in [0,1]$, we have $\frac{1}{k} \leq p_i^*(\mathbf{w}) \leq \frac{e^c}{k}$, so $p^*(\mathbf{w}) \in \mathcal{P}$ is feasible. Substituting $p^*(\mathbf{w})$ back yields

$$\max_{\mathbf{p}\in\mathcal{P}} h(\mathbf{w},\mathbf{p}) = \frac{1}{k}\sum_{i=1}^{k} \exp(cR^{(i)}(\mathbf{w})) - 1. \tag{23}$$

By the premise of the theorem, the output $\overline{\mathbf{w}}_T(c)$ satisfies

$$\max_{\mathbf{p}\in\mathcal{P}} h(\overline{\mathbf{w}}_T(c),\mathbf{p}) - \min_{\mathbf{w}\in\mathcal{W}} \max_{\mathbf{p}\in\mathcal{P}} h(\mathbf{w},\mathbf{p}) \leq g(c).$$

Using (23) and multiplying both sides by k, we obtain

$$\sum_{i=1}^{k} \exp(cR^{(i)}(\overline{\mathbf{w}}_T(c))) - \min_{\mathbf{w}\in\mathcal{W}} \sum_{i=1}^{k} \exp(cR^{(i)}(\mathbf{w})) \leq kg(c). \tag{24}$$

Let $R_{\mathbf{w}} = [R^{(1)}(\mathbf{w}), \dots, R^{(k)}(\mathbf{w})]^\top$ and write its entries in descending order as $\sigma_1(R_{\mathbf{w}}) \geq \cdots \geq \sigma_k(R_{\mathbf{w}})$. Since reordering does not change the sum, (24) is equivalent to

$$\sum_{n=1}^{k} \exp(c\sigma_n(R_{\overline{\mathbf{w}}_T(c)})) - \min_{\mathbf{w}\in\mathcal{W}} \sum_{n=1}^{k} \exp(c\sigma_n(R_{\mathbf{w}})) \leq kg(c).$$

Note that we obtain the bound for minizing the sum-of-exp function, and the rest of the proof is exactly the same as that of Theorem 1. ■

Next, we consider solving (21) in the stochastic setting. More specifically, assume that we have access to stochastic estimates $R^{(i)}(\mathbf{w};\mathbf{z})$ and $\nabla R^{(i)}(\mathbf{w};\mathbf{z})$, which are unbiased estimators of $R^{(i)}(\mathbf{w})$ and $\nabla R^{(i)}(\mathbf{w})$, respectively. Our algorithm for this setting is summarized in Algorithm 2, which serves as the counterpart to Algorithm 1. We have the following conclusion, which is a standard conclusion for using SMD for solving stochastic min-max optimization problems.

Lemma 3 *Suppose Assumption 1 holds. Let $\eta = \frac{1}{2k\sqrt{10T}}$. Then, with a probability at least $1-\delta$, Algorithm 2 ensures that*

$$\max_{\mathbf{p}\in\mathcal{P}} \psi(\overline{\mathbf{w}}_T, \mathbf{p}) - \min_{\mathbf{w}\in\mathcal{W}} \max_{\mathbf{p}\in\mathcal{P}} \psi(\mathbf{w}, \mathbf{p}) = O\left(\frac{(\exp(c))^2 \log\frac{1}{\delta}}{\sqrt{T}}\right). \tag{25}$$

Combining with Theorem 3, we obtain the following conclusion.

Theorem 4 *Let $c = \log T^{1/4}$. Then with a probability $1-\delta$, the output of Algorithm 1, $\overline{\mathbf{w}}_T$, is $O\left(\frac{1}{\log T}\log\left(k + k\log\frac{1}{\delta}\right)\right)$-lex-min-max optimal.*

B.2. Adaptive Sampling

In Algorithm 2, each round requires sampling k points, one from each distribution. Prior work in MDL (Soma et al., 2022; Zhang et al., 2023) addresses this challenge by incorporating adversarial multi-armed bandit (MAB) techniques, which reduce the sampling complexity from k to 1. The idea was firstly proposed by Soma et al. (2022) and then refined by Zhang et al. (2023). We extend our framework to adopt this approach. Specifically, following (Soma et al., 2022; Zhang et al., 2023), in each round t, we first sample an index i_t uniformly at random from $\{1, \dots, k\}$. Based on this choice, we construct the following unbiased gradient estimates:

$$\begin{aligned} \overline{\nabla}_{\mathbf{w},t} &= p_{t,i_t}\, k\, \nabla R^{(i_t)}(\mathbf{w}_t; \mathbf{z}_t^{(i_t)}), \\ \overline{\nabla}_{\mathbf{p},t} &= [0, \dots, kR^{(i_t)}(\mathbf{w}_t; \mathbf{z}_t^{(i_t)}), \dots, 0] - k\big[\log(kp_1), \dots, \log(kp_k)\big]. \end{aligned} \tag{26}$$

These updates can be directly incorporated into the SMD framework in Algorithm 2, leading to results similar to those in (25), with the exception of an additional linear dependence on k. Nevertheless, this does not affect the final upper bound in Theorem 2, since the extra linear factor in k only contributes a logarithmic dependence, which is already captured by the existing $\log k$ term in our bound.

B.3. Proof of Lemma 3

The proof is similar to that of Lemma 1, and we provide the proof for completeness. Firstly, we verify the domain of $\mathbf{w}$ and $\mathbf{p}$ are bounded with respect to the potential function $\phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$ of the Bregman divergence. We have

$$\max_{\mathbf{w}\in\mathcal{W}} \phi(\mathbf{w}) - \min_{\mathbf{w}\in\mathcal{W}} \phi(\mathbf{w}) \leq 1,$$

and

$$\max_{\mathbf{p}\in\mathcal{P}} \phi(\mathbf{p}) - \min_{\mathbf{p}\in\mathcal{P}} \phi(\mathbf{p}) \leq k \max_{\mathbf{p}\in\mathcal{P}} \frac{1}{2}\|\mathbf{p}\|_\infty^2 - \frac{1}{2} \min_{\mathbf{p}\in\mathcal{P}} \|\mathbf{p}\|_2^2 \leq \frac{(\exp(c))^2}{k}.$$

Let $D^2 = \frac{(\exp(c))^2}{k}$, and it can be easily seen that both terms above are bounded by D^2. Moreover, the gradients are also bounded: $\|\widetilde{\nabla}_{\mathbf{w},t}\|_2 \leq 1$, and

$$\left\|\widetilde{\nabla}_{\mathbf{p},t}\right\|_2 \leq c + \sqrt{\sum_{i=1}^{k} (\log kp_i)^2} \leq c + \sqrt{k}c.$$

Similarly, let $G = 4\sqrt{k}\exp(c)$. Then both terms above are bounded by G for $c > 0$. Next, we apply the standard SMD analysis, which is exactly the same as what we did in the proof of Lemma 1.