MULTI-OBJECTIVE ONLINE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper presents a systematic study of multi-objective online learning. We first formulate the framework of Multi-Objective Online Convex Optimization, which encompasses a novel multi-objective regret. This regret is built upon a sequence-wise extension of the commonly used discrepancy metric Pareto suboptimality gap in zero-order multi-objective bandits. We then derive an equivalent form of the regret, making it amenable to be optimized via first-order iterative methods. To motivate the algorithm design, we give an explicit example in which equipping OMD with the vanilla min-norm solver for gradient composition will incur a linear regret, which shows that merely regularizing the iterates, as in single-objective online learning, is not enough to guarantee sublinear regrets in the multi-objective setting. To resolve this issue, we propose a novel min-regularized-norm solver that regularizes the composite weights. Combining min-regularized-norm with OMD results in the Doubly Regularized Online Mirror Multiple Descent algorithm. We further derive the multi-objective regret bound for the proposed algorithm, which matches the optimal bound in the single-objective setting. Extensive experiments on real-world datasets verify the effectiveness of the proposed algorithm.

1 INTRODUCTION

Traditional optimization methods for machine learning are usually designed to optimize a single objective. However, in many real-world applications, we are often required to optimize multiple correlated objectives concurrently. For example, in autonomous driving (Huang et al., 2019; Lu et al., 2019b), self-driving vehicles need to solve multiple tasks such as self-localization and object identification at the same time. In online advertising (Ma et al., 2018a;b), advertisers need to decide on the exposure of items to different users to maximize both the Click-Through Rate (CTR) and the Post-Click Conversion Rate (CVR). In many multi-objective scenarios, the objectives may conflict with each other (Kendall et al., 2018). Hence, there may not exist any single solution that can optimize all the objectives simultaneously. For example, merely optimizing CTR or CVR will degrade the performance of the other (Ma et al., 2018a;b).

Multi-objective optimization (MOO) (Marler & Arora, 2004; Deb, 2014) is concerned with optimizing multiple conflicting objectives simultaneously. It seeks Pareto optimality, where no single objective can be improved without hurting the performance of others. Many different methods for MOO have been proposed, including evolutionary methods (Murata et al., 1995; Zitzler & Thiele, 1999), scalarization methods (Fliege & Svaiter, 2000), and gradient-based iterative methods (Désidéri, 2012). Recently, the Multiple Gradient Descent Algorithm (MGDA) and its variants have been introduced to the training of multi-task deep neural networks and achieved great empirical success (Sener & Koltun, 2018), making them regain a significant amount of research interest (Lin et al., 2019; Yu et al., 2020; Liu et al., 2021). These methods compute a composite gradient based on the gradient information of all the individual objectives and then apply the composite gradient to update the model parameters. The composite weights are determined by a min-norm solver (Désidéri, 2012) which yields a common descent direction of all the objectives.

However, compared to the increasingly wide application prospect, the gradient-based iterative algorithms are relatively understudied, especially in the online learning setting. Multi-objective online learning is of essential importance due to the reasons in two folds. First, due to the data explosion in many real-world scenarios, such as web applications, making in-time predictions requires performing online learning. Second, the theoretical investigation of multi-objective online learning will lay a solid foundation for the design of new optimizers for multi-task deep neural networks. This

is analogous to the single-objective setting, where nearly all the optimizers for training DNNs are initially analyzed in the online setting, such as AdaGrad (Duchi et al., 2011), Adam (Kingma & Ba, 2014), and AMSGrad (Reddi et al., 2018).

In this paper, we give a systematic study of multi-objective online learning. To begin with, we formulate the framework of Multi-Objective Online Convex Optimization (MO-OCO). One major challenge in deriving MO-OCO is the lack of an appropriate regret definition. In the multi-objective setting, in general, no single decision can optimize all the objectives simultaneously. Thus, to devise the multi-objective regret, we need to first extend the single fixed comparator used in the single-objective regret, i.e., the fixed optimal decision, to the entire Pareto optimal set. Then we need an appropriate discrepancy metric to evaluate the gap between vectorial losses. Intuitively, the Pareto suboptimality gap (PSG) metric, which is frequently used in zero-order multi-objective bandits (Turgay et al., 2018; Lu et al., 2019a), is a very promising candidate. PSG can yield scalarized measurements from any vector loss to a given comparator set. However, we find that vanilla PSG is unsuitable for our setting since it always yields non-negative values. In a concrete example, we show that the naive PSG-based regret $R_{\mathrm{I}}(T)$ can even be linear w.r.t. T when the decisions are already optimal, which disqualifies it as a regret metric. To overcome the failure of vanilla PSG, we propose its sequence-wise variant termed S-PSG, which measures the suboptimality of the whole decision sequence to the Pareto optimal set. Optimizing the resulting regret $R_{\mathrm{II}}(T)$ will drive the cumulative loss to approach the Pareto front. However, as a zero-order metric motivated geometrically, it is very difficult to design first-order algorithms to optimize the regret. To resolve the issue, we derive a more intuitive equivalent form of $R_{\mathrm{II}}(T)$ via a highly non-trivial transformation.

Based on the MO-OCO framework, we develop a novel multi-objective online algorithm termed Doubly Regularized Online Mirror Multiple Descent. The key module of the algorithm is the gradient composition scheme, which calculates a composite gradient in the form of a convex combination of the gradients of all objectives. Intuitively, the most direct way to determine the composite weights is to apply the min-norm solver (Désidéri, 2012) commonly used in offline multi-objective optimization. However, directly applying min-norm is not workable in the online setting. Specifically, the composite weights in the min-norm are only determined by the gradients at the current round. In the online setting, since the gradients can be adversarial, they may result in undesired composite weights, further producing a composite gradient that reversely optimizes the loss. To rigorously verify this point, we give an example where equipping OMD with vanilla min-norm incurs a linear regret, showing that only regularizing the iterate, as in OMD, is not enough to guarantee sublinear regrets in our setting. To fix the issue, we devise a novel min-regularized-norm solver with an explicit regularization on composite weights. Equipping it with OMD results in our proposed algorithm.

In theory, we derive a regret bound of $O(\sqrt{T})$ for DR-OMMD, which matches the optimal bound in the single-objective setting (Hazan et al., 2016) and is tight w.r.t. the number of objectives. Our analysis also shows that DR-OMMD attains a smaller regret bound than that of linearization with fixed composite weights. We show that, in the two-objective setting with linear losses, the margin between the regret bounds depends on the difference between the composite weights yielded by the two algorithms and the difference between the two gradients of the underlying objectives.

To evaluate the effectiveness of DR-OMMD, we conduct extensive experiments on large-scale real-world datasets. We first realize adaptive regularization via multi-objective optimization on real-world datasets, and find that adaptive regularization with DR-OMMD significantly outperforms fixed regularization with linearization, which verifies the effectiveness of DR-OMMD over linearization in the convex setting. Then we apply DR-OMMD to deep online multi-task learning. The results show that DR-OMMD is also effective in the non-convex setting.

In summary, this paper gives the first systematic study of multi-objective online learning, encompassing a novel framework, a new algorithm, and corresponding non-trivial theoretical analysis. We believe that this work paves the way for future research on more advanced multiple-objective optimization algorithms, which may inspire the design of new optimizers for multi-task deep learning.

2 PRELIMINARIES

In this section, we briefly review the necessary background knowledge of multi-objective optimization and online convex optimization.

2.1 MULTI-OBJECTIVE OPTIMIZATION

Multiple-objective optimization (MOO) is concerned with solving the problems of optimizing multiple objectives simultaneously (Fliege & Svaiter, 2000; Deb, 2014). In general, since different objectives may conflict with each other, there is no single solution that can optimize all the objectives at the same time. Thus, the conventional concept of optimality used in the single-objective setting is no longer suitable for the multi-objective setting. Instead, MOO seeks to achieve Pareto optimality. In the following, we exposit Pareto optimality and related definitions more formally. We use a vector-valued loss $H = (h^1, \ldots, h^m)^\intercal$ to denote the objectives, where $m \geq 2$ and $h^i : \mathcal{K} \to \mathbb{R}$, $i \in \{1, \ldots, m\}$, $\mathcal{K} \subset \mathbb{R}$, is the i-th loss function.

Definition 1 (**Pareto optimality**). **(a)** For any two solutions $x, x' \in \mathcal{K}$, we say that x dominates x', denoted as $x \prec x'$ or $x' \succ x$, if $h^i(x) \leq h^i(x')$ for all i, and there exists one i such that $h^i(x) < h^i(x')$; otherwise, we say that x does not dominate x', denoted as $x \not\prec x'$ or $x' \not\succ x$.
(b) A solution $x^* \in \mathcal{K}$ is called Pareto optimal if it is not dominated by any other solution in $\mathcal{K}$.

Note that there may exist multiple Pareto optimal solutions. For example, it is easy to show that the optimizer of any single objective, i.e., $x_i^* \in \arg\min_{x \in \mathcal{K}} h^i(x), i \in \{1, \ldots, m\}$, is Pareto optimal. Different Pareto optimal solutions reflect different trade-offs among the objectives (Lin et al., 2019).

Definition 2 (**Pareto front**). **(a)** All Pareto optimal solutions form the Pareto set $\mathcal{P}_{\mathcal{K}}(H)$.
(b) The image of $\mathcal{P}_{\mathcal{K}}(H)$ constitutes the Pareto front, denoted as $\mathcal{P}(H) = \{H(x) \mid x \in \mathcal{P}_{\mathcal{K}}(H)\}$.

Now that we have established the notion of optimality in MOO, we proceed to introduce the metrics that measure the discrepancy of an arbitrary solution $x \in \mathcal{K}$ from being optimal. Recall that, in the single-objective setting with merely one loss function $h : \mathcal{Q} \to \mathbb{R}$, where $\mathcal{Q} \subset \mathbb{R}$, for any $z \in \mathcal{Q}$, the loss difference $h(z) - \min_{z'' \in \mathcal{Q}} h(z'')$ is directly the discrepancy measure. However, in MOO with more than one loss, for any $x \in \mathcal{K}$, the loss difference $H(x) - H(x'')$, where $x'' \in \mathcal{P}_{\mathcal{K}}(H)$, is a vector. Intuitionally, the desired discrepancy metric shall scalarize the vector-valued loss difference and yield 0 for any Pareto optimal solution. In general, in MOO, there are two commonly used discrepancy metrics, i.e., Pareto suboptimality gap (PSG) (Turgay et al., 2018) and Hypervolume (HV) (Bradstreet, 2011). As HV is a complex volume-based metric, it is more difficult to optimize via gradient-based algorithms (Zhang & Golovin, 2020). Hence in this paper, we adopt PSG, which has already been extensively used in multi-objective bandits (Turgay et al., 2018; Lu et al., 2019a).

Definition 3 (**Pareto suboptimality gap**[1]). For any $x \in \mathcal{K}$, the Pareto suboptimality gap to a given comparator set $\mathcal{K}^* \subset \mathcal{K}$, denoted as $\Delta(x; \mathcal{K}^*, H)$, is defined as the minimal scalar $\epsilon \geq 0$ that needs to be subtracted from all entries of $H(x)$, such that $H(x) - \epsilon \mathbf{1}$ is not dominated by any point in $\mathcal{K}^*$, where $\mathbf{1}$ denotes the all-one vector in $\mathbb{R}^m$, i.e.,

$$\Delta(x; \mathcal{K}^*, H) = \inf_{\epsilon \geq 0} \epsilon, \text{s.t. } \forall\, x'' \in \mathcal{K}^*,\ \exists\, i \in \{1, \ldots, m\},\ h^i(x) - \epsilon < h^i(x'').$$

Clearly, PSG is a distance-based discrepancy metric motivated from a purely geometric viewpoint. In practice, the comparator set $\mathcal{K}^*$ is often set to be the Pareto set $\mathcal{P}_{\mathcal{K}}(H)$ (Turgay et al., 2018). Then for any $x \in \mathcal{K}$, its PSG is always non-negative and equals zero if and only if $x \in \mathcal{P}_{\mathcal{K}}(H)$.

Multiple Gradient Descent Algorithm (MGDA) is an offline first-order algorithm for MOO (Fliege & Svaiter, 2000; Désidéri, 2012). At each iteration $l \in \{1, \ldots, L\}$, where L is the number of iterations, it first computes the gradient $\nabla h^i(x_l)$ for each objective $i \in \{1, \ldots, m\}$, then derives the composite gradient $g_l^{comp} = \sum_{i=1}^m \lambda_l^i \nabla h^i(x_l)$ as the convex combination of these multiple gradients, and finally applies g_l^{comp} to execute the gradient descent step to update the decision, i.e., $x_{l+1} = x_l - \eta g_l^{comp}$, where η is the step size. The core part of MGDA is the module that determines the composite weights $\lambda_l = (\lambda_l^1, \ldots, \lambda_l^m)$, which is given by

$$\lambda_l = \arg\min_{\lambda_l \in \mathcal{S}_m} \| \sum\nolimits_{i=1}^m \lambda_l^i \nabla h^i(x_l) \|_2^2,$$

where $\mathcal{S}_m = \{\lambda \in \mathbb{R}^m \mid \sum_{i=1}^m \lambda^i = 1, \lambda^i \geq 0, i \in \{1, \ldots, m\}\}$ denotes the probabilistic simplex in $\mathbb{R}^m$. This is a min-norm solver, which finds the weights in the simplex that yield the minimum L_2 norm of the composite gradient. Thus MGDA is also called the *min-norm* method. Existing works (Désidéri, 2012; Sener & Koltun, 2018) have shown that under the convex setting where all h^i are convex functions, MGDA is guaranteed to decrease all the objectives simultaneously until it reaches a Pareto optimal decision.

[1] Our definition looks a bit different from (Turgay et al., 2018). In Appendix B, we show they are equivalent.

2.2 ONLINE CONVEX OPTIMIZATION

Online Convex Optimization (OCO) (Zinkevich, 2003; Hazan et al., 2016) is the most commonly adopted framework for designing online learning algorithms. It can be viewed as a structured repeated game between a learner and an adversary. At each round $t \in \{1, \ldots, T\}$, the learner is required to generate a decision x_t from a convex compact set $\mathcal{X} \subset \mathbb{R}^n$. Then the adversary replies the learner with a convex function $f_t : \mathcal{X} \to \mathbb{R}$ and the learner suffers the loss $f_t(x_t)$. The goal of the learner is to minimize the regret with respect to the best fixed decision in hindsight, i.e.,

$$R(T) = \sum\nolimits_{t=1}^{T} f_t(x_t) - \min_{x^* \in \mathcal{X}} \sum\nolimits_{t=1}^{T} f_t(x^*).$$

A meaningful regret is required to be sublinear in T, i.e., $\lim_{T\to\infty} R(T)/T = 0$, which implies that when T is large enough, the learner can perform as well as the best fixed decision in hindsight.

Online Mirror Descent (OMD) (Hazan et al., 2016) is a classic first-order online learning algorithm. At each round $t \in \{1, \ldots, T\}$, OMD yields its decision via

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \eta \langle \nabla f_t(x_t), x \rangle + B_R(x, x_t),$$

where η is the step size, $R : \mathcal{X} \to \mathbb{R}$ is the regularization function, and $B_R(x, x') = R(x) - R(x') - \langle \nabla R(x'), x - x' \rangle$ is the Bregman divergence induced by R. As a meta-algorithm, by instantiating different regularization functions, OMD can induce two important algorithms, i.e., Online Gradient Descent (Zinkevich, 2003) and Online Exponentiated Gradient (Hazan et al., 2016).

3 MULTI-OBJECTIVE ONLINE CONVEX OPTIMIZATION

In this section, we formally formulate the framework of Multi-Objective Online Convex Optimization (MO-OCO).

Framework overview. Analogously to single-objective OCO, MO-OCO can also be viewed as a repeated game between an online learner and the adversarial environment. The major difference is that in MO-OCO, the feedback to the learner given by the environment is now multi-objective. The general framework of MO-OCO can be stated as follows. At each round $t \in \{1, \ldots, T\}$, the learner generates a decision x_t from a given convex compact decision set $\mathcal{X} \subset \mathbb{R}^n$. Then the adversary replies the decision with a vector-valued loss function $F_t : \mathcal{X} \to \mathbb{R}^m$, where its i-th component $f_t^i(x) : \mathcal{X} \to \mathbb{R}$ is a convex function that corresponds to the i-th objective, and the learner suffers the vector loss $F_t(x_t) \in \mathbb{R}^m$. The goal of the learner is to generate a sequence of decisions $\{x_t\}_{t=1}^T$ to minimize a certain kind of multi-objective regret.

The remaining work to accomplish the formulation of MO-OCO is to give an appropriate definition of the multi-objective regret, which is also the most challenging part. Recall that the single-objective regret $R(T) = \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(x^*)$ is defined as the difference between the cumulative loss of the actual decisions $\{x_t\}_{t=1}^T$ and that of the fixed optimal decision in hindsight $x^* \in \min_{x\in\mathcal{X}} \sum_{t=1}^T f_t(x)$. When defining the multi-objective analogy to $R(T)$, we encounter two issues. First, in the multi-objective setting, no single decision can optimize all the objectives at the same time in general. Thus, in MO-OCO, we cannot compare the cumulative loss with that of a single fixed decision. Instead, we use the the Pareto optimal set $\mathcal{X}^*$ of the cumulative loss function $\sum_{t=1}^T F_t$, i.e., $\mathcal{X}^* = \mathcal{P}_X(\sum_{t=1}^T F_t)$, which naturally aligns with the optimality concept in MOO. Second, to compare $\{x_t\}_{t=1}^T$ and $\mathcal{X}^*$ in the loss space, we need a discrepancy metric to measure the gap between vectorial losses. Intuitively, we can adopt the commonly used discrepancy metric PSG (Turgay et al., 2018). But we find that vanilla PSG is not appropriate for the OCO setting that is largely different from the bandits setting. In the following, we explicate the reason in detail.

3.1 THE NAIVE REGRET BASED ON VANILLA PSG FAILS IN MO-OCO

By definition, at each round t, the difference between the decision x_t and the Pareto optimal set can be evaluated by PSG $\Delta(x_t; \mathcal{X}^*, F_t)$. Naturally, we can formulate the multi-objective regret by accumulating $\Delta(x_t; \mathcal{X}^*, F_t)$ over all rounds, i.e.,

$$R_{\mathrm{I}}(T) := \sum\nolimits_{t=1}^{T} \Delta(x_t; \mathcal{X}^*, F_t).$$

Recall that the single-objective regret can also expressed as $R(T) = \sum_{t=1}^{T}(f_t(x_t) - f_t(x^*))$. Hence, $R_{\mathrm{I}}(T)$ essentially extends the scalar discrepancy $f_t(x_t) - f_t(x^*)$ to the PSG metric $\Delta(x_t; \mathcal{X}^*, F_t)$. However, these two discrepancy metrics have a major difference, i.e., $f_t(x_t) - f_t(x^*)$ can be negative, whereas $\Delta(x_t; \mathcal{X}^*, F_t)$ is always non-negative. In previous bandits settings (Turgay et al., 2018), the discrepancy is intrinsically non-negative, since the comparator set is exactly the Pareto optimal set of the evaluated loss function. However, the non-negative property of PSG can be problematic in our setting, where the comparator set $\mathcal{X}^*$ is the Pareto set of the cumulative loss function, rather than the instantaneous loss F_t that is used for evaluation. Specifically, at some round t, the decision x_t may Pareto dominate all points in $\mathcal{X}^*$ w.r.t. F_t, which corresponds to the single-objective setting where it is possible that $f_t(x_t) < f_t(x^*)$ at some specific round. In this case, we would expect the discrepancy metric at this round to be negative. However, PSG can only yield 0 in this case, making the regret much looser than we expect. In the following, we provide an example in which the naive regret $R_{\mathrm{I}}(T)$ is **linear** w.r.t. T even when the decisions x_t are already optimal.

Problem instance. Set $\mathcal{X} = [-2, 2]$. Let the loss function be identical among all objectives, i.e., $f_t^1(x) = ... = f_t^m(x)$, and alternate between x and $-x$. Suppose the time horizon T is an even number, then the Pareto optimal set $\mathcal{X}^* = \mathcal{X}$. Now consider the decisions $x_t = 1, t \in \{1, ..., T\}$. In this case, it can easily be checked that the single-objective regret of each objective is zero, indicating that these decisions are optimal for each objective. To calculate $R_{\mathrm{I}}(T)$, notice that when all the objectives are identical, PSG reduces to $\Delta(x_t; \mathcal{X}^*, f_t^1) = \sup_{x^* \in \mathcal{X}} \max\{f_t^1(x_t) - f_t^1(x^*), 0\}$ at each round t. Hence, in this case we have $R_{\mathrm{I}}(T) = \sum_{1 \le k \le T/2}(\sup_{x^* \in [-2,2]} \max\{1 - x^*, 0\} + \sup_{x^* \in [-2,2]} \max\{x^* - 1, 0\}) = 3T$, which is linear w.r.t. T. Therefore, $R_{\mathrm{I}}(T)$ is too loose to measure the suboptimality of decisions, which is **unqualified as a regret metric**.

3.2 The Alternative Regret based on Sequence-wise PSG

In light of the failure of the naive regret, we need to modify the discrepancy metric in our setting. Recall that the single-objective regret can be interpreted as the gap between the actual cumulative loss $\sum_{t=1}^{T} f_t(x_t)$ and its optimal value $\min_{x \in \mathcal{X}} \sum_{t=1}^{T} f_t(x)$. In analogy, we can measure the gap between $\sum_{t=1}^{T} F_t(x_t)$ and the Pareto front $\mathcal{P}^* = \mathcal{P}_{\mathcal{X}}(\sum_{t=1}^{T} F_t)$. However, vanilla PSG is a point-wise metric, i.e., it can only measure the suboptimality of a decision point. To evaluate the decision sequence $\{x_t\}_{t=1}^{T}$, we modify its definition and propose a sequence-wise variant of PSG.

Definition 4 (**Sequence-wise PSG**). For any decision sequence $\{x_t\}_{t=1}^{T}$, the sequence-wise PSG (S-PSG) to a given comparator set $\mathcal{X}^*$ w.r.t. the loss sequence $\{F_t\}_{t=1}^{T}$ is defined as

$$\Delta(\{x_t\}_{t=1}^{T}; \mathcal{X}^*, \{F_t\}_{t=1}^{T}) = \inf_{\epsilon \ge 0} \epsilon, \text{ s.t. } \forall\, x'' \in \mathcal{X}^*, \exists\, i \in \{1, \dots, m\}, \sum_{t=1}^{T} f_t^i(x_t) - \epsilon < \sum_{t=1}^{T} f_t^i(x'').$$

Since $\mathcal{X}^*$ is the Pareto set of $\sum_{t=1}^{T} F_t$, S-PSG measures the discrepancy from the cumulative loss of the decision sequence to the Pareto front $\mathcal{P}^*$. Now the regret can be directly given as

$$R_{\mathrm{II}}(T) := \Delta(\{x_t\}_{t=1}^{T}; \mathcal{X}^*, \{F_t\}_{t=1}^{T}).$$

$R_{\mathrm{II}}(T)$ has a clear physical meaning that optimizing it will impose the cumulative loss to be close to the Pareto front $\mathcal{P}^*$. However, since PSG (or S-PSG) is a zero-order metric motivated in a purely geometric sense, i.e., its calculation needs to solve a constrained optimization problem with an unknown boundary $\{F_t(x'') \mid x'' \in \mathcal{X}^*\}$, it is difficult to design a first-order algorithm to optimize PSG-based regrets, not to mention the analysis. To resolve this issue, we derive an equivalent form via highly non-trivial transformations, which is more intuitive than its original form.

Proposition 1. *The multi-objective regret $R_{\mathrm{II}}(T)$ based on S-PSG has an equivalent form, i.e.,*

$$R_{\mathrm{II}}(T) = \max\Big\{ \sup_{x^* \in \mathcal{X}^*} \inf_{\lambda^* \in \mathcal{S}_m} \sum\nolimits_{t=1}^{T} {\lambda^*}^{\top}(F_t(x_t) - F_t(x^*)),\ \ 0 \Big\}.$$

Remark. (i) The above form is closely related to the single-objective regret $R(T)$. Specifically, when $m = 1$, we can prove that $R_{\mathrm{II}}(T) = \max\{\sum_{t=1}^{T} F_t(x_t) - \min_{x^* \in \mathcal{X}^*} \sum_{t=1}^{T} F_t(x^*), 0\} = \max\{R(T), 0\}$. Note that in the regret analysis, we are more interested in the case of $R(T) \ge 0$ (where $R_{\mathrm{II}}(T) = R(T)$), since when $R(T) < 0$, it is naturally bounded by any sublinear regret

Algorithm 1 Doubly Regularized Online Mirror Multiple Descent **(DR-OMMD)**

Input: Convex set $\mathcal{X}$, time horizon T, regularization parameter α_t, learning rate η_t, regularization function R, user preference λ_0.
Initialize: $x_1 \in \mathcal{X}$.
for $t = 1, \ldots, T$ **do**
  Predict x_t and receive a loss function $F_t : \mathcal{X} \to \mathbb{R}^m$.
  Compute the multiple gradients $\nabla F_t(x_t) = [\nabla f_t^1(x_t), \ldots, \nabla f_t^m(x_t)] \in \mathbb{R}^{n\times m}$.
  Determine the weights for the gradient composition via **min-regularized-norm**
$$\lambda_t = \arg\min_{\lambda \in \mathcal{S}_m} \|\nabla F_t(x_t)\lambda\|_2^2 + \alpha_t \|\lambda - \lambda_0\|_1.$$
  Compute the composite gradient $g_t = \nabla F_t(x_t)\lambda_t$.
  Perform online mirror descent using g_t
$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \eta_t \langle g_t, x\rangle + B_R(x, x_t).$$
end for

bound. Hence, $R_{\mathrm{II}}(T)$ is essentially aligned with $R(T)$ in the single-objective setting.
(ii) At its first glance, $R_{\mathrm{II}}(T)$ can be optimized via linearization with fixed weights $\lambda_0 \in \mathcal{S}_m$, or alternatively, optimizing a single objective $i \in \{1, ..., m\}$. We remark that this is not a problem of our regret definition, but **an intrinsic requirement** of Pareto optimality. Specifically, Pareto optimality characterizes the status where no objective can be improved without hurting others. Hence merely optimizing a single objective naturally achieves Pareto optimality. Please refer to Proposition 8 in (Emmerich & Deutz, 2018) for the rigorous proof. As a general performance metric, our regret should incorporate this special case. Later, we will design a novel algorithm based on the concept of common descent, which outperforms linearization in both theory and experiment.

4 Doubly Regularized Online Mirror Multiple Descent

In this section, we present the Doubly Robust Online Mirror Multiple Descent (DR-OMMD) algorithm, the protocol of which is given in Algorithm 1. At each round t, the learner first computes the gradient of the loss regarding each objective, then determines the composite weights of all these gradients, and finally applies the composite gradient to the online mirror descent step.

4.1 Vanilla Min-Norm May Incur Linear Regrets

The core module of DR-OMMD is the composition of gradients. For simplicity, denote the gradients at round t in a matrix form $\nabla F_t(x_t) = [\nabla f_t^1(x_t), \ldots, \nabla f_t^m(x_t)] \in \mathbb{R}^{n\times m}$. Then the composite gradient is $g_t = \nabla F_t(x_t)\lambda_t$, where λ_t is the composite weights. As illustrated in the preliminary, in the offline setting, the min-norm method (Désidéri, 2012; Sener & Koltun, 2018) is a classic method to determine the composite weights, which produces a common descent direction that can descend all the losses simultaneously. Thus, it is tempting to consider applying it to the online setting.

However, directly applying min-norm to the online setting is not workable, which may even incur linear regrets. In vanilla min-norm, the composite weights λ_t are determined solely by the gradients $\nabla F_t(x_t)$ at the current round t, which are very sensitive to the instantaneous loss F_t. In the online setting, the losses at each round can be adversarially chosen, and thus the corresponding gradients can be adversarial. These adversarial gradients may result in undesired composite weights, which may further produce a composite gradient that even deteriorates the next prediction. In the following, we provide an example in which min-norm incurs a linear regret. We extend OMD (Hazan et al., 2016) to the multi-objective setting, where the composite weights are directly yielded by min-norm.

Problem instance. We consider a two-objective problem. The decision domain is $\mathcal{X} = \{(u, v) \mid u + v \le \frac{1}{2}, v - u \le \frac{1}{2}, v \ge 0\}$ and the loss function at each round is

$$F_t(x) = \begin{cases} (\|x-a\|^2, \|x-b\|^2), & t = 2k-1, \quad k = 1, 2, ...; \\ (\|x-b\|^2, \|x-c\|^2), & t = 2k, \qquad k = 1, 2, ..., \end{cases}$$

where $a = (-2, -1), b = (0, 1), c = (2, -1)$. For simplicity, we first analyze the case where the total time horizon T is an even number. Then we can compute the Pareto set of the cumulative loss

$\sum_{t=1}^{T} F_t$, i.e., $\mathcal{X}^* = \{(u, 0) \mid -\frac{1}{2} \leq u \leq \frac{1}{2}\}$, which locates at the x-axis. For conciseness of analysis, we instantiate OMD with $L2$-regularization, which results in the simple OGD algorithm (McMahan, 2011). We start at an arbitrary point $x_1 = (u_1, v_1) \in \mathcal{X}$ satisfying $v_1 > 0$. At each round t, suppose the decision $x_t = (u_t, v_t)$, then the gradient of each objective w.r.t. x_t takes

$$g_t^1 = \begin{cases} (2u_t + 4, & 2v_t + 2), \quad t = 2k - 1; \\ (2u_t, & 2v_t - 2), \quad t = 2k. \end{cases} \qquad g_t^2 = \begin{cases} (2u_t, & 2v_t - 2), \quad t = 2k - 1; \\ (2u_t - 4, & 2v_t + 2), \quad t = 2k. \end{cases}$$

Since $0 \leq v_t \leq \frac{1}{2}$, we observe that the second entry of either gradient alternates between positive and negative. By using min-norm, the composite weights λ_t can be computed as

$$\lambda_t = \begin{cases} ((1 - u_t - v_t)/4, & (3 + u_t + v_t)/4), \quad t = 2k - 1; \\ ((3 - u_t + v_t)/4, & (1 + u_t - v_t)/4), \quad t = 2k. \end{cases}$$

We observe that both entries of composite weights alternative between above $\frac{1}{2}$ and below $\frac{1}{2}$, and $\|\lambda_{t+1} - \lambda_t\|_1 \geq 1$. Recall that $\|\lambda_t\|_1 = 1$, hence the composite weights at two consecutive rounds change radically. The resulting composite gradient takes

$$g_t^{comp} = \begin{cases} (u_t - v_t + 1, & -u_t + v_t - 1), \quad t = 2k - 1; \\ (-u_t - v_t - 1, & -u_t - v_t - 1), \quad t = 2k. \end{cases}$$

The fluctuating composite weights mix with the positive and negative second entries of gradients, making the second entry of g_t^{comp} always negative, i.e., $-u_t + v_t - 1 < 0$ and $-u_t - v_t - 1 < 0$. Hence g_t^{comp} always drives x_t away from the Pareto set $\mathcal{X}^*$ that coincides with the x-axis. This essentially reversely optimizes the loss, hence increasing the regret. In fact, we can prove that it even incurs a linear regret. Due to the lack of space, we leave the proof of linear regret when T is an odd number in Appendix G. The above results of the problem instance are summarized as follows.

Proposition 2. *For OMD equipped with vanilla min-norm, there exists a multi-objective online convex optimization problem, in which the resulting algorithm incurs a linear regret.*

Remark. Stability is a basic requirement to ensure meaningful regrets in online learning (McMahan, 2017). In the single-objective setting, directly regularizing the iterate x_t (e.g., OMD) is enough. However, as shown in the above analysis, merely regularizing x_t is not enough to attain sublinear regrets in the multi-objective setting, since there is another source of instability, i.e., the composite weights, that affects the direction of composite gradients. Therefore, in multi-objective online learning, besides regularizing the iterates, we also need to explicitly regularize the composite weights.

4.2 The Algorithm

Enlightened by the design of regularization in FTRL (McMahan, 2017), we consider the regularizer $r(\lambda, \lambda_0)$, where λ_0 is the pre-defined composite weights that may reflect the user preference. This results in a new solver called *min-regularized-norm*, i.e.,

$$\lambda_t = \arg\min_{\lambda \in \mathcal{S}_m} \|\nabla F_t(x_t)\lambda\|_2^2 + \alpha_t \, r(\lambda, \lambda_0),$$

where α_t is the regularization strength. Equipping OMD with the new solver, we derive the proposed algorithm. Note that beyond the regularization on the iterate x_t that is intrinsic in online learning, there is another regularization on the composite weights λ_t in min-regularized-norm. Both regularizations are fundamental, and they together ensure stability in the multi-objective online setting. Hence we call the algorithm Doubly Regularized Online Mirror Multiple Descent (DR-OMMD).

In principle, r can take various forms such as L_1-norm, L_2-norm, etc. Here we adopt L_1-norm since it aligns well with the simplex constraint of λ. Min-regularized-norm can be computed very efficiently. When $m = 2$, it has a closed-form solution. Specifically, suppose the gradients at round t are g_t^1 and g_t^2. Set $\gamma_L = (g_2^\top (g_2 - g_1) - \alpha_t)/\|g_2 - g_1\|^2$ and $\gamma_R = (g_2^\top (g_2 - g_1) + \alpha_t)/\|g_2 - g_1\|^2$. Given any $\lambda_0 = (\gamma_0, 1 - \gamma_0) \in \mathcal{S}_2$, we can compute the composite weights λ_t as $(\gamma_t, 1 - \gamma_t)$ where

$$\gamma_t = \max\{\min\{\gamma_t'', 1\}, 0\}, \quad \text{where } \gamma_t'' = \max\{\min\{\gamma_0, \gamma_R\}, \gamma_L\}.$$

When $m > 2$, since the constraint $\mathcal{S}_m$ is a simplex, we can introduce a Frank-Wolfe solver (Jaggi, 2013) (see detailed protocol in Appendix D.1). We also discuss the L_2-norm case in Appendix D.2.

Compared to vanilla min-norm, the composite weights in min-regularized-norm are not fully determined by the adversarial gradients. The resulting relative stability of composite weights makes the composite gradients more robust to the adversarial environment. In the following, we give a general analysis and prove that DR-OMMD indeed guarantees sublinear regrets.

4.3 THEORETICAL ANALYSIS

We now analyze the regret of DR-OMMD. Our analysis is based on two commonly used assumptions (Jadbabaie et al., 2015; Hazan et al., 2016).

Assumption 1 (**Bregman divergence**). The regularization function R is 1-strongly convex. In addition, the Bregman divergence is γ-Lipschitz continuous, i.e., $B_R(x, z) - B_R(y, z) \leq \gamma\|x - y\|, \forall x, y, z \in \mathrm{dom}R$, where $\mathrm{dom}R$ is the domain of R and satisfies $\mathcal{X} \subset \mathrm{dom}R \subset \mathbb{R}^n$.

Assumption 2 (**Lipschitz continuity**). For each $i \in \{1, \ldots, m\}$, there exists some positive and finite G such that, the i-th loss f_t^i at each round $t \in \{1, \ldots, T\}$ is differentiable and G-Lipschitz continuous w.r.t. $\|\cdot\|$, i.e., $|f_t^i(x) - f_t^i(x')| \leq G\|x - x'\|$. Note that in the convex setting, this assumption leads to bounded gradients, i.e., $\|\nabla f_t^i(x)\|_* \leq G$ for any $t \in \{1, \ldots, T\}, i \in \{1, \ldots, m\}, x \in \mathcal{X}$.

Theorem 1. *Suppose the diameter of $\mathcal{X}$ is D. Assume F_t is bounded, i.e., $|f_t^i(x)| \leq F, \forall x \in \mathcal{X}, t \in \{1, \ldots, T\}, i \in \{1, \ldots, m\}$. For any $\lambda_0 \in \mathcal{S}_m$, DR-OMMD attains*

$$R(T) \leq \frac{\gamma D}{\eta_T} + \sum_{t=1}^{T} \frac{\eta_t}{2}(\|\nabla F_t(x_t)\lambda_t\|_2^2 + \frac{4F}{\eta_t}\|\lambda_t - \lambda_0\|_1).$$

Remark. When $\eta_t = \frac{\sqrt{2\gamma D}}{G\sqrt{T}}$ or $\frac{\sqrt{2\gamma D}}{G\sqrt{t}}$, $\alpha_t = \frac{4F}{\eta_t}$, the bound attains $O(\sqrt{T})$. It matches the optimal single-objective bound w.r.t. T (Hazan et al., 2016) and is tight w.r.t. m (justified in Appendix E.1).

Regret comparison with linearization. Linearization with fixed weights $\lambda_0 \in \mathcal{S}_m$ can be viewed as single-objective optimization on scalar loss $\lambda_0^\top F_t$, whose gradient is $g_t = \nabla F_t(x_t)\lambda_0$. Hence we can directly borrow the tight bound of OMD (Theorem 6.8 in (Orabona, 2019)) and derive a bound $\frac{\gamma D}{\eta_T} + \sum_{t=1}^{T} \frac{\eta_t}{2}\|\nabla F_t(x_t)\lambda_0\|_2^2$ for linearization. In comparison, when $\alpha_t = \frac{4F}{\eta_t}$, DR-OMMD attains a regret bound $\frac{\gamma D}{\eta_T} + \sum_{t=1}^{T} \frac{\eta_t}{2} \min_{\lambda \in \mathcal{S}_m}\{\|\nabla F_t(x_t)\lambda\|^2 + \alpha_t\|\lambda - \lambda_0\|_1\}$, which is smaller than that of linearization. Note that although the bound of linearization refers to single-objective regret $R(T)$, the comparison is reasonable since in Proposition 1 we have proved that the two regret metrics are consistent, such that $R_{\mathrm{II}}(T) = \max\{R(T), 0\}$ when $m = 1$. We further investigate the margin in the two-objective setting with linear losses. Suppose the loss functions are $f_t^1(x) = x^\top g_t^1$ and $f_t^2(x) = x^\top g_t^2$ for some $g_t^1, g_t^2 \in \mathbb{R}^n$ at each round. Via a non-trivial computation, we can show that the margin is at least (see Appendix E.2 for the detailed proof)

$$M \geq \sum_{t=1}^{T} \frac{\eta_t}{4}\|\lambda_t - \lambda_0\|_2^2 \cdot \|g_t^1 - g_t^2\|_2^2.$$

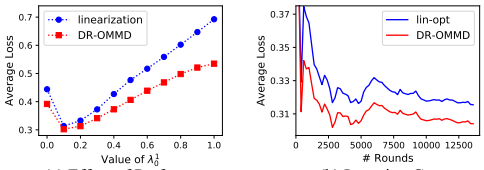
This shows the benefit of DR-OMMD over linearization. First, the performance of linearization relies on the proper choice of λ_0, while DR-OMMD can adaptively select more reasonable λ_t at each round. Second, when the gradients of different objectives vary wildly, DR-OMMD will be much better than linearization. This matches our intuition that linearization suffers from conflict gradients (Yu et al., 2020), while DR-OMMD can alleviate the conflict by pursuing common descent.

5 EXPERIMENTS

In this section, we conduct extensive experiments to evaluate the effectiveness of DR-OMMD. We consider two baselines: (i) *linearization* performs single-objective online learning on linearized losses $\lambda_0^\top F_t$ with pre-defined fixed weights $\lambda_0 \in \mathcal{S}_m$; (ii) *min-norm* equips OMD with vanilla min-norm (Désidéri, 2012) for gradient composition.

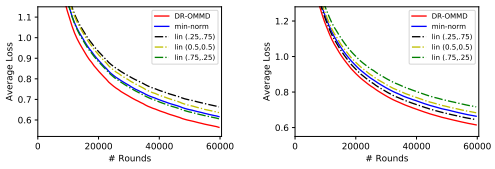
5.1 CONVEX EXPERIMENTS: ADAPTIVE REGULARIZATION

In many real-world online scenarios, regularization is adopted to avoid overfitting. A standard realization is to add a term $r(x)$ to the loss $f_t(x)$ at each round and optimize the regularized loss $f_t(x) + \sigma r(x)$ (McMahan, 2011), where σ is a hyperparameter that needs to be fixed beforehand. The formalism of multi-objective online learning provides a novel way to realize regularization. Since $r(x)$ measures the complexity of x, it can be regarded as the second objective alongside the primary goal $f_t(x)$. We can augment the loss to $F_t(x) = (f_t(x), r(x))$ at each round and thereby cast regularized online learning into a bi-objective online optimization problem. Compared to using fixed σ, the new approach effectively chooses the hyperparameter $\sigma_t = \lambda_t^2/\lambda_t^1$ in an adaptive way.



(a) Effect of Preference (b) Learning Curve

Figure 1: Results to verify the effectiveness of adaptive regularization on *protein*. (a) Performance of DR-OMMD and linearization under varying $\lambda_0 = (\lambda_0^1, 1 - \lambda_0^1)$. (b) Performance using the optimal weights $\lambda_0 = (0.1, 0.9)$.



(a) Task L (b) Task R

Figure 2: Results to verify the effectiveness of DR-OMMD in the non-convex setting. The two plots show the performance of DR-OMMD and various baselines on both tasks (Task L and Task R) of MultiMNIST.

We use two large-scale online benchmark datasets. (i) *protein* is a bioinformatics dataset for protein type classification (Wang, 2002), which has 17 thousand instances with 357 features. (ii) *covtype* is a biological dataset collected from a non-stationary environment for forest cover type prediction (Blackard & Dean, 1999), which has 50 thousand instances with 54 features. We set the logistic classification loss as the first objective, and the squared $L2$-norm of model parameters as the second objective. Since the ultimate goal of regularization is to lift predictive performance, we measure the average loss, i.e., $\sum_{t \leq T} l_t(x_t)/T$, where $l_t(x_t)$ is the classification loss at round t.

We adopt a $L2$-norm ball centered at the origin with diameter $K = 100$ as the decision set. The learning rates are decided by a grid search over $\{0.1, 0.2, \ldots, 3.0\}$. For DR-OMMD, the parameter α_t is simply set as 0.1. For fixed regularization, the strength $\sigma = (1-\lambda_0^1)/\lambda_0^1$ is determined by some $\lambda_0^1 \in [0, 1]$, which is exactly *linearization* with weights $\lambda_0 = (\lambda_0^1, 1 - \lambda_0^1)$. We run both algorithms with varying $\lambda_0^1 \in \{0, 0.1, ..., 1\}$. In Figure 1, we plot (a) their final performance w.r.t. the choice of λ_0 and (b) their learning curves with desirable λ_0 (e.g., $(0.1, 0.9)$ on *protein*). Other results are deferred to the appendix due to the lack of space. The results show that DR-OMMD consistently outperforms fixed regularization; the gap becomes more significant when λ_0 is not properly set.

5.2 Non-Convex Experiments: Deep Multi-Task Learning

We conduct non-convex experiments on MultiMNIST (Sabour et al., 2017), which is a multi-task version of the MNIST dataset for image classification and commonly used in deep multi-task learning (Sener & Koltun, 2018; Lin et al., 2019). In MultiMNIST, each sample is constructed by putting a random image of some digit at the top-left and another image at the bottom-right. The goal is to classify the digit at the top-left (*task L*) and that at the bottom-right (*task R*) at the same time.

We follow (Sener & Koltun, 2018)'s setup and adopt LeNet. For linearization, we examine different weights $(0.25, 0.75)$, $(0.5, 0.5)$, and $(0.75, 0.25)$ (Lin et al., 2019). In all methods, learning rates are selected via a grid search over $\{0.0001, 0.001, 0.01, 0.1\}$. For DR-OMMD, the parameter α_t is set according to Theorem 1, and the initial weights are simply set as $\lambda_0 = (0.5, 0.5)$. Note that in the online setting, samples arrive in a sequential manner, which is a core difference from offline experiments where sample batches are randomly sampled from the training set. Figure 2 compares the average cumulative loss of all the examined methods. We also measure two commonly used metrics in offline experiments (Reddi et al., 2018), i.e., the training loss and test loss, whose results are similar to the average loss and deferred to the appendix due to the lack of space. The results show that DR-OMMD outperforms counterpart algorithms using min-norm or linearization in all metrics on both tasks, which verifies the effectiveness of DR-OMMD in the online non-convex setting.

6 Conclusions

In this paper, we give a systematic study of multi-objective optimization in the online setting. We first formulate the framework of Multi-Objective Online Convex Optimization. Then we devise the Doubly Regularized Online Mirror Multiple Descent algorithm, which has a special design for gradient composition in online learning, namely min-regularized-norm. We provide a non-trivial regret analysis for DR-OMMD and conduct extensive experiments to demonstrate its effectiveness.

Reproducibility statement. The proofs of all theoretical statements in this paper are supplemented in the appendix. The core code of the min-regularized-norm solver is provided in the supplementary material.

REFERENCES

Jacob Abernethy, Peter L Bartlett, and Elad Hazan. Blackwell approachability and no-regret learning are equivalent. In *Annual Conference on Learning Theory*, pp. 27–46, 2011.

Amir Beck and Marc Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

Jock A Blackard and Denis J Dean. Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables. *Computers and Electronics in Agriculture*, 24(3):131–151, 1999.

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

Lucas Bradstreet. *The hypervolume indicator for multi-objective optimisation: calculation and use*. University of Western Australia Perth, 2011.

Róbert Busa-Fekete, Balázs Szörényi, Paul Weng, and Shie Mannor. Multi-objective bandits: Optimizing the generalized gini index. In *International Conference on Machine Learning*, pp. 625–634. PMLR, 2017.

Nicolò Cesa-bianchi, Pierre Gaillard, Gabor Lugosi, and Gilles Stoltz. Mirror descent meets fixed share (and feels no regret). In *Advances in Neural Information Processing Systems*, pp. 980–988, 2012.

Zhao Chen, Jiquan Ngiam, Yanping Huang, Thang Luong, Henrik Kretzschmar, Yuning Chai, and Dragomir Anguelov. Just pick a sign: Optimizing deep multitask models with gradient sign dropout. *Advances in Neural Information Processing Systems*, 2020.

Kalyanmoy Deb. Multi-objective optimization. In *Search methodologies*, pp. 403–449. 2014.

Rémy Degenne, Thomas Nedelec, Clément Calauzènes, and Vianney Perchet. Bridging the gap between regret minimization and best arm identification, with application to a/b tests. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1988–1996, 2019.

Jean-Antoine Désidéri. Multiple-gradient descent algorithm for multiobjective optimization. *Comptes Rendus Mathematique*, 350(5-6):313–318, 2012.

Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. URL `http://archive.ics.uci.edu/ml`.

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

Michael TM Emmerich and André H Deutz. A tutorial on multiobjective optimization: Fundamentals and evolutionary methods. *Natural Computing*, 17(3):585–609, 2018.

Jörg Fliege and Benar Fux Svaiter. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, 51(3):479–494, 2000.

Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

Xinyu Huang, Peng Wang, Xinjing Cheng, Dingfu Zhou, Qichuan Geng, and Ruigang Yang. The apolloscape open dataset for autonomous driving and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10):2702–2719, 2019.

Ali Jadbabaie, Alexander Rakhlin, Shahin Shahrampour, and Karthik Sridharan. Online optimization: Competing with dynamic comparators. In *Artificial Intelligence and Statistics*, pp. 398–406, 2015.

Martin Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *International Conference on Machine Learning*, pp. 427–435, 2013.

Adrián Javaloy and Isabel Valera. Rotograd: Gradient homogenization in multitask learning. *arXiv preprint arXiv:2103.02631*, 2021.

Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7482–7491, 2018.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Wouter Koolen. The pareto regret frontier. *Advances in Neural Information Processing Systems*, pp. 1–9, 2013.

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

Xi Lin, Hui-Ling Zhen, Zhenhua Li, Qing-Fu Zhang, and Sam Kwong. Pareto multi-task learning. In *Advances in Neural Information Processing Systems*, pp. 12060–12070, 2019.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. *arXiv preprint arXiv:2110.14048*, 2021.

Shiyin Lu, Guanghui Wang, Yao Hu, and Lijun Zhang. Multi-objective generalized linear bandits. In *International Joint Conference on Artificial Intelligence*, pp. 3080–3086, 2019a.

Weixin Lu, Yao Zhou, Guowei Wan, Shenhua Hou, and Shiyu Song. L3-net: Towards learning based lidar localization for autonomous driving. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6389–6398, 2019b.

Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H Chi. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *International Conference on Knowledge Discovery & Data Mining*, pp. 1930–1939, 2018a.

Xiao Ma, Liqin Zhao, Guan Huang, Zhi Wang, Zelin Hu, Xiaoqiang Zhu, and Kun Gai. Entire space multi-task model: An effective approach for estimating post-click conversion rate. In *International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 1137–1140, 2018b.

Shie Mannor, Vianney Perchet, and Gilles Stoltz. Approachability in unknown games: Online learning meets multi-objective optimization. In *Annual Conference on Learning Theory*, pp. 339–355, 2014.

R Timothy Marler and Jasbir S Arora. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6):369–395, 2004.

Brendan McMahan. Follow-the-regularized-leader and mirror descent: Equivalence theorems and $L1$ regularization. In *International Conference on Artificial Intelligence and Statistics*, pp. 525–533, 2011.

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

Tadahiko Murata, Hisao Ishibuchi, et al. MOGA: multi-objective genetic algorithms. In *IEEE International Conference on Evolutionary Computation*, pp. 289–294, 1995.

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

Sashank J Reddi, Satyen Kale, and Sanjiv Kumar. On the convergence of Adam and beyond. In *International Conference on Learning Representations*, 2018.

Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. In *Advances in Neural Information Processing Systems*, pp. 3859–3869, 2017.

Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. In *Advances in Neural Information Processing Systems*, pp. 525–536, 2018.

Nati Srebro, Karthik Sridharan, and Ambuj Tewari. On the universality of online mirror descent. In *Advances in Neural Information Processing Systems*, pp. 2645–2653, 2011.

Cem Tekin and Eralp Turğay. Multi-objective contextual multi-armed bandit with a dominant objective. *IEEE Transactions on Signal Processing*, 66(14):3799–3813, 2018.

Eralp Turgay, Doruk Oner, and Cem Tekin. Multi-objective contextual bandit problem with similarity information. In *International Conference on Artificial Intelligence and Statistics*, pp. 1673–1681, 2018.

Jung-Ying Wang. *Application of support vector machines in bioinformatics*. PhD thesis, National Taiwan University, 2002.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In *Advances in Neural Information Processing Systems*, pp. 5824–5836, 2020.

Richard Zhang and Daniel Golovin. Random hypervolume scalarizations for provable multi-objective black box optimization. In *International Conference on Machine Learning*, pp. 11096–11105, 2020.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning*, pp. 928–936, 2003.

Eckart Zitzler and Lothar Thiele. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. *IEEE transactions on Evolutionary Computation*, 3(4):257–271, 1999.

APPENDIX

The appendix is organized as follows. Appendix A reviews related work. Appendix B validates the correctness of our definition of PSG. Appendix C provides the detailed derivation of the equivalent form of $R_{\mathrm{II}}(T)$. Appendix D discusses how to efficiently compute the composition weights for the min-regularized-norm solver. Appendix E discusses the tightness of DR-OMMD's regret bound, then provides more details on the regret comparison between DR-OMMD and linearization. Appendix F supplements more details in the experimental setup and empirical results. Appendix G and H provide detailed proofs of the remaining theoretical claims in this work.

A RELATED WORK

In this section, we review previous work in three related fields, i.e., online learning, multi-objective optimization, and multi-objective multi-armed bandits.

A.1 ONLINE LEARNING

Online learning arms to make sequential predictions for streaming data. Please refer to the introduction books (Hazan et al., 2016; Orabona, 2019) for more background knowledges.

Most of the previous works on online learning are conducted in the single-objective setting. As far as we are concerned, there are only two lines of work concerning multi-objective learning. The first line of works provides a multi-objective perspective of the prediction-with-expert-advice (PEA)

problem (Koolen, 2013; Koolen & Van Erven, 2015). Specifically, they view each individual expert as a multi-objective criterion, and characterize the Pareto optimal trade-offs among different experts. These works have two main distinctions from our proposed MO-OCO. First, they are still built upon the original PEA problem where the payoff of each expert (or decision) is a scalar, while we focus on vectoral payoffs. Second, their framework is restricted to an absolute loss game, whereas our framework is general and can be applied to any coordinate-wise convex loss functions.

The second line of work studies online learning with vectoral payoffs via Blackwell approachability (Blackwell, 1956; Mannor et al., 2014; Abernethy et al., 2011). In their framework, the learner is given a target set $\mathcal{T} \subset \mathbb{R}^m$ and its goal is to generate decisions $\{x_t\}_{t=1}^T$ to minimize the distance between the average loss $\frac{1}{T}\sum_{t=1}^T l_t(x_t)$ and the target set $\mathcal{T}$. There are two major differences between Blackwell approachability and our proposed MO-OCO: previous works on Blackwell approachability are zero-order methods and the target set $\mathcal{T}$ is often known beforehand (also see the discussion in (Busa-Fekete et al., 2017)), while in MO-OCO we intend to develop a first-order method to reach the unknown Pareto front.

A.2 Multi-Objective Optimization

Multi-objective optimization aims to optimize multiple objectives concurrently. Most of the previous works on multi-objective optimization are conducted in the offline setting, including the batch optimization setting (Désidéri, 2012; Liu et al., 2021) and the stochastic optimization setting (Sener & Koltun, 2018; Lin et al., 2019; Yu et al., 2020; Chen et al., 2020; Javaloy & Valera, 2021). These methods are based on gradient composition, and have shown very promising results in multi-task learning applications.

Despite the existence of previous works on multi-objective optimization, as the first work of multi-objective optimization in the OCO setting, our work is largely different from them in three aspects. First, we contribute the first formal framework of multi-objective online convex optimization. In particular, our framework is based on a novel equivalent transformation of the PSG metric, which is intrinsically different from previous offline optimization frameworks. Second, we provide a showcase in which a commonly used method in the offline setting, namely min-norm (Désidéri, 2012; Sener & Koltun, 2018), fail to attain sublinear regret in online setting. Our proposed min-regularized-norm is a novel design when tailoring offline methods to the online setting. Third, the regret analysis of multi-objective online learning is intrinsically different from the convergence analysis in the offline setting (Yu et al., 2020).

A.3 Multi-Objective Multi-Armed Bandits

Another branch of related works study multi-objective optimization in the multi-armed bandits setting (Busa-Fekete et al., 2017; Tekin & Turğay, 2018; Turgay et al., 2018; Lu et al., 2019a; Degenne et al., 2019). Among these works, the most relevant one to ours is (Turgay et al., 2018), which introduces the Pareto suboptimality gap (PSG) metric to characterize the multi-objective regret in the bandits setting, and proposes a zero-order zooming algorithm to minimize the regret.

In this work, our regret definition also utilizes the PSG metric (Turgay et al., 2018). However, as the first study of multi-objective optimization in the OCO setting, our work is intrinsically different from these previous works in the following aspects. First, as PSG is a zero-order metric, we perform a novel equivalent transformation, making it amenable to the OCO setting. Second, our proposed algorithm is a first-order multiple gradient algorithm, whose design principles are completely distinct from zero-order algorithms. For example, the concept of the stability of composite weights does not even exist in the design of previous zero-order methods for multi-objective bandits (Turgay et al., 2018; Lu et al., 2019a). Third, the regret analysis of MO-OCO is intrinsically different from that in the bandits setting.

B An Equivalent Definition of PSG

Recall that in Definition 3, we formulate the PSG metric as a constrained optimization problem. We note that, since the PSG metric is based on the notion of "non-dominance" (Turgay et al., 2018), its

most direct form is actually

$$\Delta'(x;\mathcal{K}^*,H) = \inf_{\epsilon\geq 0} \epsilon,$$
$$\text{s.t.} \quad \forall x'' \in \mathcal{K}^*, \exists i \in \{1,\dots,m\}, h^i(x) - \epsilon < h^i(x'')$$
$$\text{or} \quad \forall i \in \{1,\dots,m\}, h^i(x) - \epsilon = h^i(x'').$$

At the first glance, the above definition seems to be quite different from Definition 3, since it has an extra condition "$\forall i \in \{1,\dots,m\}, h^i(x) - \epsilon = h^i(x'')$". In the following, we prove that both definitions actually yield the same value due to the infimum operation on ϵ.

Specifically, for any possible $(x,\mathcal{K}^*,H)$, we denote $\Delta'(x;\mathcal{K}^*,H) = \epsilon_0'$ and $\Delta(x;\mathcal{K}^*,H) = \epsilon_0$. By comparing the constraints of both definitions, it is obvious that ϵ_0 must satisfy the constraint of $\Delta'(x;\mathcal{K}^*,H)$, hence the infimum operation guarantees that $\epsilon_0' \leq \epsilon_0$. It remains to prove that $\epsilon_0' \geq \epsilon_0$. To this end, we only need to show that $\epsilon_0' + \xi$ satisfies the constraint of $\Delta(x;\mathcal{K}^*,H)$ for any $\xi > 0$. Consider an arbitrary $x'' \in \mathcal{K}^*$. From the definition of $\Delta'(x;\mathcal{K}^*,H)$, we know that either $\exists i \in \{1,\dots,m\}, h^i(x) - \epsilon_0' < h^i(x'')$ or $\forall i \in \{1,\dots,m\}, h^i(x) - \epsilon_0' = h^i(x'')$. Whichever condition holds, we must have $\exists i \in \{1,\dots,m\}, h^i(x) - \epsilon_0' - \xi < h^i(x'')$ for any $\xi > 0$. Since it holds for any $x'' \in \mathcal{K}^*$, $\epsilon_0' + \xi$ lies in the feasible region of $\Delta(x;\mathcal{K}^*,H)$, hence we have $\epsilon_0 \leq \epsilon_0' + \xi, \forall \xi > 0$ and thus $\epsilon_0 \leq \epsilon_0'$. In summary, we have $\Delta'(x;\mathcal{K}^*,H) = \Delta(x;\mathcal{K}^*,H)$ for any $(x,\mathcal{K}^*,H)$.

C Derivation of the Equivalent Multi-Objective Regret Form

In this section, We provide the derivation of the equivalent max-min form of $R_{\mathrm{II}}(T)$ in Proposition 1, which is highly non-trivial and forms the basis of the subsequent algorithm design and theoretical analysis.

Proof of Proposition 1. Recall that the PSG metric used in $R_{\mathrm{II}}(T)$ is an extension of vanilla PSG to leverage any decision sequence. To motivate the analysis, we first investigate vanilla PSG $\Delta(x;\mathcal{X}^*,F)$ that deals with a single decision x, and derive a useful lemma as follows.

Lemma 1. *Vanilla PSG has an equivalent form, i.e.,*

$$\Delta(x;\mathcal{X}^*,F) = \sup_{x^*\in\mathcal{X}^*} \inf_{\lambda\in\mathcal{S}_m} \lambda^\top (F(x) - F(x))_+,$$

where for any vector $l \in \mathbb{R}^m$, the truncation $(l)_+$ represent a vector whose i-th entry equals to $\max\{l^i, 0\}$ for all $i \in \{1,...,m\}$.

Proof. In the definition of PSG, the evaluated decision x is compared to all Pareto optimal points $x' \in \mathcal{X}^*$. For any fixed comparator $x' \in \mathcal{X}^*$, we define the **pair-wise suboptimality gap** w.r.t. F between decisions x and x', i.e.,

$$\delta(x;x',F) = \inf_{\epsilon\geq 0}\{\epsilon \mid F(x) - \epsilon\mathbf{1} \not\succ F(x')\}.$$

Hence, vanilla PSG can be expressed as

$$\Delta(x;\mathcal{X}^*,F) = \sup_{x'\in\mathcal{X}^*} \delta(x;x',F).$$

To proceed, we analyze the pair-wise gap $\delta(x;x',F)$. From its formulation, we know that $\delta(x;x',F)$ measures the minimal non-negative value that needs to be subtracted from each entry of $F(x)$ until it is not dominated by x' anymore. We consider the following two cases.

(i) If $F(x) \not\succ F(x')$, i.e., $f^{k_0}(x) \leq f^{k_0}(x')$ for some $k_0 \in \{1,...,m\}$, nothing needs to be subtracted from $F(x)$ and we directly have $\delta(x;x',F) = 0$.

(ii) If $F(x) \succ F(x')$, i.e., $f^k(x) \geq f^k(x')$ for any $k \in \{1,...,m\}$, then the condition $F(x) - \epsilon\mathbf{1} \not\succ F(x')$ is obviously violated when $\epsilon = 0$. Now let us gradually increase ϵ. This condition holds only when there there exists some k_0 such that $f^{k_0}(x) - \epsilon \leq f^{k_0}(x')$, or equivalently $\epsilon \geq f^{k_0}(x) - f^{k_0}(x')$. Hence, in this case, we have $\delta(x;x',F) = \min_{k\in\{1,\dots,m\}}\{f^k(x) - f^k(x')\}$.

Combining the above two cases, we derive an equivalent form of the pair-wise gap. Specifically, we can easily check that the following form holds for both cases, i.e.,

$$\delta(x; x', F) = \min_{k \in \{1,\dots,m\}} \max\{f^k(x) - f^k(x'), 0\}.$$

To relate the above form with F, denote $\mathcal{U}_m = \{e_k \mid 1 \le k \le m\}$ as the set of all unit vector in $\mathbb{R}^m$, then we equivalently have

$$\delta(x; x', F) = \min_{\lambda \in \mathcal{U}_m} \lambda^\top (F(x) - F(x'))_+.$$

Now the calculation of $\delta(x; x', F)$ becomes a minimization problem over $\lambda \in \mathcal{U}_m$. Since $\mathcal{U}_m$ is a discrete set, we can apply a linear relaxation trick. Specifically, we now turn to minimize the scalar $p(\lambda) = \lambda^\top \max\{F(x) - F(x'), 0\}$ over the convex curvature of $\mathcal{U}_m$, which is exactly the probability simplex $\mathcal{S}_m = \{\lambda \in \mathbb{R}^m \mid \lambda \succ \mathbf{0}, \|\lambda\|_1 = 1\}$. Note that $\mathcal{U}_m$ contains all the vertexes of $\mathcal{S}_m$. Since $\inf_{\lambda \in \mathcal{S}_m} p(\lambda)$ is a linear optimization problem, the minimal point ${\lambda^*}^\top$ must be a vertex of the simplex, i.e., ${\lambda^*}^\top \in \mathcal{U}_m$. Thus the relaxed problem is equivalent to the original problem, namely,

$$\delta(x; x', F) = \min_{\lambda \in \mathcal{U}_m} \lambda^\top (F(x) - F(x'))_+ = \inf_{\lambda \in \mathcal{S}_m} \lambda^\top (F(x) - F(x'))_+.$$

Taking the supremum over $x' \in \mathcal{X}^*$, we prove the lemma. ■

The above lemma can be extended to S-PSG. Specifically, we can extend the pair-wise suboptimality gap $\delta(x; x', F)$ to a sequence of decisions, i.e.,

$$\delta(\{x_t\}_{t=1}^T; x', \{F_t\}_{t=1}^T) = \inf_{\epsilon \ge 0} \{\epsilon \mid \sum_{t=1}^T F_t(x_t) - \epsilon \mathbf{1} \not\succ \sum_{t=1}^T F_t(x')\}.$$

Then S-PSG can be expressed as

$$\Delta(\{x_t\}_{t=1}^T; \mathcal{X}^*, \{F_t\}_{t=1}^T) = \sup_{x^* \in \mathcal{X}^*} \delta(\{x_t\}_{t=1}^T; x^*, \{F_t\}_{t=1}^T).$$

Now we investigate the extended pair-wise gap. Similar to the derivation of the above lemma, by discussing the relations between $\sum_{t=1}^T F_t(x_t)$ and $\sum_{t=1}^T F_t(x')$, we can similarly derive an equivalent form of the extended pair-wise gap as

$$\delta(\{x_t\}_{t=1}^T; x', \{F_t\}_{t=1}^T) = \min_{k \in \{1,\dots,m\}} \max\{\sum_{t=1}^T f_t^k(x) - \sum_{t=1}^T f_t^k(x'), 0\},$$

and further

$$\delta(\{x_t\}_{t=1}^T; x', \{F_t\}_{t=1}^T) = \inf_{\lambda \in \mathcal{S}_m} \lambda^\top (\sum_{t=1}^T F_t(x_t) - \sum_{t=1}^T F_t(x'))_+.$$

Hence the S-PSG-based regret form can be expressed as

$$R_{\mathrm{II}}(T) = \sup_{x^* \in \mathcal{X}^*} \inf_{\lambda \in \mathcal{S}_m} \lambda^\top (\sum_{t=1}^T F_t(x_t) - \sum_{t=1}^T F_t(x^*))_+.$$

The max-min form of $R_{\mathrm{II}}(T)$ has a truncation operation $(\cdot)_+$, which brings irregularity to the regret form. To handle the truncation operation, we utilize the following lemma:

Lemma 2. *(a) For any $l \in \mathbb{R}^m$, we have $\inf_{\lambda \in \mathcal{S}_m} \lambda^\top (l)_+ = \max\{\inf_{\lambda \in \mathcal{S}_m} \lambda^\top l, 0\}$.*
(b) For any $h : \mathcal{X} \to \mathbb{R}$, we have $\sup_{x \in \mathcal{X}} \max\{h(x), 0\} = \max\{\sup_{x \in \mathcal{X}} h(x), 0\}$.

Proof. To prove the first statement, we consider the following two cases.
(i) If $l \succ \mathbf{0}$, then $(l)_+ = l$. For any $\lambda \in \mathcal{S}_m$, we have $\lambda^\top (l)_+ = \lambda^\top l > 0$. Taking the infimum over $\lambda \in \mathcal{S}_m$ on both sides, we have $\inf_{\lambda^\top \mathcal{S}_m} \lambda^\top (l)_+ = \inf_{\lambda \in \mathcal{S}_m} \lambda^\top l \ge 0$. Moreover, from the last equation we have $\max\{\inf_{\lambda \in \mathcal{S}_m} \lambda^\top l, 0\} = \inf_{\lambda \in \mathcal{S}_m} \lambda^\top l$. The statement holds in this case.
(ii) If $l \succeq \mathbf{0}$, then $l^i \le 0$ for some $i \in \{1, ..., m\}$. Set e_i as the i-th unit vector in $\mathbb{R}^m$, then we

Algorithm 2 Frank-Wolfe Solver for Min-Regularized-Norm with L1-Norm

Initialize: $\lambda_t = (\gamma_t^1, \ldots, \gamma_t^m) = (\frac{1}{m}, \ldots, \frac{1}{m})$.
Compute the matrix $\mathbf{U} = \nabla F_t(x_t)^\top \nabla F_t(x_t)$, i.e., $\mathbf{U}_{ij} = \nabla f_t^i(x_t)^\top \nabla f_t^j(x_t), \forall i, j \in \{1, \ldots, m\}$.
repeat
  Select an index $k \in \arg\max_{i \in \{1,\ldots,m\}} \{\sum_{j=1}^m \gamma_t^j \mathbf{U}_{ij} + \alpha \operatorname{sgn}(\gamma_t^i - \gamma_0^i)\}$.
  Compute $\delta \in \arg\min_{0 \le \delta \le 1} \left\|\delta \nabla f_t^k(x_t) + (1-\delta)\nabla F_t(x_t)\lambda_t\right\|_2^2 + \alpha\|\delta(e_k - \lambda_t) + \lambda_t - \lambda_0\|_1$.
  Update $\lambda_t = (1-\delta)\lambda_t + \delta e_k$.
until $\delta \sim 0$ **or** Number of Iteration Limits
return λ_t.

have $e_i^\top l \le 0$. On the one hand, since $e_i \in \mathcal{S}_m$, we have $\inf_{\lambda \in \mathcal{S}_m} \lambda^\top l \le e_i^\top l \le 0$, and further $\max\{\inf_{\lambda \in \mathcal{S}_m} \lambda^\top l, 0\} = 0$. On the other hand, notice that $e_i^\top (l)_+ = 0$ and $\lambda^\top (l)_+ \ge 0$ for any $\lambda \in \mathcal{S}_m$, we know that $\inf_{\lambda \in \mathcal{S}_m} \lambda^\top (l)_+ = e_i^\top (l)_+ = 0$. The statement also holds in this case.

To prove the second statement, we also consider two cases.
(i) If $h(x_0) > 0$ for some $x_0 \in \mathcal{X}$, then $\sup_{x \in \mathcal{X}} h(x) \ge h(x_0) > 0$, and $\max\{\sup_{x \in \mathcal{X}} h(x), 0\} = \sup_{x \in \mathcal{X}} h(x)$. We also have $\sup_{x \in \mathcal{X}} \max\{h(x), 0\} = \sup_{x \in \mathcal{X}} h(x)$, which proves the statement.
(ii) If $h(x) \le 0$ for all $x \in \mathcal{X}$, then $\sup_{x \in \mathcal{X}} h(x) \le 0$, and thus $\max\{\sup_{x \in \mathcal{X}} h(x), 0\} = 0$. On the other hand, for any $x \in \mathcal{X}$, we have $\max\{h(x)\} = 0$, which proves the statement. ■

From the above lemma, we directly have

$$\begin{aligned} R_{\mathrm{II}}(T) &= \sup_{x^* \in \mathcal{X}^*} \max\{\inf_{\lambda \in \mathcal{S}_m} \lambda^\top (\sum_{t=1}^T F_t(x_t) - \sum_{t=1}^T F_t(x^*)), 0\} \\ &= \max\{\sup_{x^* \in \mathcal{X}^*} \inf_{\lambda \in \mathcal{S}_m} \lambda^\top (\sum_{t=1}^T F_t(x_t) - \sum_{t=1}^T F_t(x^*)), 0\}, \end{aligned}$$

which derives the desired equivalent form. ■

D Calculation of Min-Regularized-Norm

In this section, we discuss how to efficiently calculate the solutions to min-regularized-norm with $L1$-norm and $L2$-norm.

D.1 L1-Norm

Similar to (Sener & Koltun, 2018), we first consider the setting of two objectives, namely $m = 2$. In this case, for any $\lambda = (\gamma, 1-\gamma), \lambda_0 = (\gamma_0, 1-\gamma_0) \in \mathcal{S}_2$, the L1-regularization $\|\lambda - \lambda_0\|_1$ equals to $2|\gamma - \gamma_0|$. Hence min-regularized-norm with L1-norm at round t reduces to $\lambda_t = (\gamma_t, 1-\gamma_t)$ where

$$\gamma_t \in \arg\min_{0 \le \gamma \le 1} \|\gamma g_1 + (1-\gamma) g_2\|^2 + 2\alpha|\gamma - \gamma_0|.$$

Interestingly, we find that the above problem has a closed-form solution.

Proposition 3. *Set $\gamma_L = (g_2^\top (g_2 - g_1) - \alpha)/\|g_2 - g_1\|^2$, and $\gamma_R = (g_2^\top (g_2 - g_1) + \alpha)/\|g_2 - g_1\|^2$. Then min-regularized-norm with L1-norm produces weights $\lambda_t = (\gamma_t, 1-\gamma_t)$ where*

$$\gamma_t = \max\{\min\{\gamma_t'', 1\}, 0\}, \quad \textit{where } \gamma_t'' = \max\{\min\{\gamma_0, \gamma_R\}, \gamma_L\}.$$

Proof. We solve the following two quadratic sub-problems, i.e.,

$$\min_{0 \le \gamma \le \gamma_0} h_1(\gamma) = \|\gamma g_1 + (1-\gamma) g_2\|^2 + 2\alpha(\gamma_0 - \gamma),$$

as well as

$$\min_{\gamma_0 \le \gamma \le 1} h_2(\gamma) = \|\gamma g_1 + (1-\gamma) g_2\|^2 + 2\alpha(\gamma - \gamma_0).$$

It can be checked that in the former sub-problem, h_1 monotonously decreases on $(-\infty, \gamma_R]$ and increases on $[\gamma_R, +\infty)$; in the latter sub-problem, h_2 monotonously decreases on $(-\infty, \gamma_L]$ and increases on $[\gamma_L, +\infty)$. Since each sub-problem has its constraint ($[0, \gamma_0]$ or $[\gamma_0, 1]$), the solution to the original optimization problem can then be derived by comparing the optimal values of the two sub-problems with their constraints. Specifically, notice that $\gamma_L \leq \gamma_R$ and $0 \leq \gamma_0 \leq 1$, and we can consider the following three cases.

(i) When $0 \leq \gamma_0 \leq \gamma_L \leq \gamma_R$, then h_1 monotonously decreases on $[0, \gamma_0]$ and its minimum on $[0, \gamma_0]$ is $h_1(\gamma_0)$. Notice that $h_1(\gamma_0) = h_2(\gamma_0)$. For the sub-problem of h_2, we further consider two situations:
(i-a) If $\gamma_L \leq 1$, then $\gamma_L \in [\gamma_0, 1]$, hence the minimum of h_2 on $[\gamma_0, 1]$ is $h_2(\gamma_L)$. Since $h_2(\gamma_L) \leq h_2(\gamma_0) = h_1(\gamma_0)$, the minimal point of the original problem is γ_L, and hence $\gamma_t = \gamma_L$.
(i-b) If $\gamma_L > 1$, then h_2 monotonously decreases on $[\gamma_0, 1]$, and we surely have $h_2(1) \leq h_2(\gamma_0) = h_1(\gamma_0)$. Hence $\gamma_t = 1$ in this situation.
Combining the above two situations, we have $\gamma_t = \min\{\gamma_L, 1\}$ in this case.

(ii) When $\gamma_L \leq \gamma_R \leq \gamma_0 \leq 1$, then h_2 monotonously increases on $[\gamma_0, 1]$ and its minimum on $[\gamma_0, 1]$ is $h_2(\gamma_0)$. Notice that $h_1(\gamma_0) = h_2(\gamma_0)$. For the sub-problem of h_1, similar to the first case, we also consider two situations:
(ii-a) If $\gamma_R \geq 0$, then $\gamma_R \in [0, \gamma_0]$, hence the minimum of h_1 on $[0, \gamma_0]$ is $h_1(\gamma_R)$. Since $h_1(\gamma_R) \leq h_1(\gamma_0) = h_2(\gamma_0)$, the minimal point of the original problem is γ_R, and hence $\gamma_t = \gamma_R$.
(ii-b) If $\gamma_R < 0$, then h_1 monotonously increases on $[0, \gamma_0]$. Hence we have $h_1(0) \leq h_1(\gamma_0) = h_2(\gamma_0)$. Hence the solution to the original problem $\gamma_t = 0$.
Combining the above two situations, we have $\gamma_t = \max\{\gamma_R, 0\}$ in this case.

(iii) When $\gamma_L < \gamma_0 < \gamma_R$, then h_1 monotonously decreases on $[0, \gamma_0]$ and h_2 monotonously increases on $[\gamma_0, 1]$. Hence each sub-problem attains its minimum at γ_0, and thus $\gamma_t = \gamma_0$.

Summarizing the above three cases gives

$$\gamma_t = \begin{cases} \min\{\gamma_L, 1\}, & \gamma_0 \leq \gamma_L; \\ \max\{\gamma_R, 0\}, & \gamma_0 \geq \gamma_R; \\ \gamma_0, & \text{otherwise.} \end{cases}$$

We can further rewrite the above formula into a compact form as follows, which can be checked case-by-case.

$$\gamma_t = \max\{\min\{\gamma_t'', 1\}, 0\}, \quad \text{where } \gamma_t'' = \max\{\min\{\gamma_0, \gamma_R\}, \gamma_L\},$$

This gives the closed-form solution of min-regularized-norm when $m = 2$. ■

Now that we have derived the closed-form solution to the min-regularized-norm solver with any two gradients, in principle, we can apply (Sener & Koltun, 2018)'s technique to efficiently compute the solution to the solver with more than two gradients. We provide the full procedure in Algorithm 2, which is an extension of (Sener & Koltun, 2018). By following the exact line search technique (Jaggi, 2013) in MGDA, we get our line search oracle as line 5 in Algorithm 2. The first term is the same as that in MGDA, and the second term is an extra l_1-regularization term related to the design in Algorithm 1. Unlike the oracle of MGDA that has a closed-form solution by a reduction to the case of two gradients, the extra l_1-norm term makes our oracle difficult to get a closed-form solution. The reason is that, such an extra term is the l_1-norm of a m-dimension vector, hence it can not simply reduce to the case of two gradients. To proceed, we can directly apply numerical methods to get the solution (e.g. similar to the implementation in (Liu et al., 2021)).

D.2 L2-Norm

Recall that in min-regularized-norm, the regularization on λ can take various forms. In the following, we discuss an alternative regularization, i.e., L2-regularization $r(\lambda, \lambda_0) = \frac{1}{2}\|\lambda - \lambda_0\|_2^2$. In the discussion, we will show that similar to (Sener & Koltun, 2018), min-regularized-norm with L2-norm can be computed very efficiently via the Frank-Wolfe method.

Similar to the previous discussion on L1-regularization, we first consider the setting of $m = 2$. In this case, for any $\lambda = (\gamma, 1-\gamma), \lambda_0 = (\gamma_0, 1-\gamma_0) \in \mathcal{S}_2$, the L2-regularization $\frac{1}{2}\|\lambda - \lambda_0\|_2^2$ equals

Algorithm 3 Frank-Wolfe Solver for Min-Regularized-Norm with L2-Norm

1: **Initialize:** $\lambda_t = (\gamma_t^1, \ldots, \gamma_t^m) = (\frac{1}{m}, \ldots, \frac{1}{m})$.
2: Compute the matrix $\mathbf{U} = \nabla F_t(x_t)^\top \nabla F_t(x_t)$, i.e., $\mathbf{U}_{ij} = \nabla f_t^i(x_t)^\top \nabla f_t^j(x_t), \forall i, j \in \{1, \ldots, m\}$.
3: **repeat**
4: Select an index $k \in \arg\max_{i \in \{1,\ldots,m\}} \{\sum_{j=1}^m \gamma_t^j \mathrm{U}_{ij} + \alpha(\gamma_t^i - \gamma_0^i)\}$.
5: Compute $\delta \in \arg\min_{0 \le \delta \le 1} \|\delta \nabla f_t^k(x_t)) + (1-\delta)\nabla F_t(x_t)\lambda_t\|_2^2 + \alpha\|\delta(e_k - \lambda_t) + \lambda_t - \lambda_0\|_2^2$, which has an analytical form

$$\delta = \max\{\min\{\frac{(\nabla F_t(x_t)\lambda_t - \nabla f_t^k(x_t))^\top \nabla F_t(x_t)\lambda_t + \alpha\|e_k - \lambda_t\|^2}{\|\nabla F_t(x_t)\lambda_t - \nabla f_t^k(x_t)\|^2 + \alpha(e_k - \lambda_t)^\top(\lambda_t - \lambda_0)}, 1\}, 0\}.$$

6: Update $\lambda_t = (1-\delta)\lambda_t + \delta e_k$.
7: **until** $\delta \sim 0$ **or** Number of Iteration Limits
8: **return** λ_t.

to $(\gamma - \gamma_0)^2$. Hence min-regularized-norm with L2-norm at round t reduces to $\lambda_t = (\gamma_t, 1 - \gamma_t)$ where

$$\gamma_t \in \arg\min_{0 \le \gamma \le 1} \|\gamma g_1 + (1-\gamma) g_2\|^2 + \alpha(\gamma - \gamma_0)^2.$$

Since the above problem is in the quadratic form, it also has a closed-form solution. The proof is elementary and hence omitted.

Proposition 4. *Min-regularized-norm with L2-norm produces weights $\lambda_t = (\gamma_t, 1 - \gamma_t)$ where*

$$\gamma_t = \max\{\min\{\frac{(g_2 - g_1)^\top g_2 + \alpha\gamma_0}{\|g_2 - g_1\|^2 + \alpha}, 1\}, 0\}.$$

When $m > 2$, since λ_t is constrained to the probability simplex Δ_m, similar to the case of L1-regularization, we can use a Frank-Wolfe method to efficiently calculate the composition weights, which is presented in Algorithm 3. Note that since the line search (step 2) has a closed-form solution, its calculation cost is not high, i.e., just the same as the calculation cost of the original min-norm solver (Sener & Koltun, 2018).

E More Details of The Theoretical Results

In this section, we prove the tightness of the regret bound of DR-OMMD, and give a more detailed comparison with linearization from the theoretical aspect.

E.1 The Tightness of DR-OMMD's Bound

Recall that in the remarks below Theorem 1 in our main paper, we have discussed the order of the regret bound $O(\sqrt{T})$ for DR-OMMD with proper choices of hyperparameters. It can be actually derived from the main theorem. Here we restate the result as the corollary below.

Corollary 1. *(i) With $\eta_t = \frac{\sqrt{2\gamma D}}{G\sqrt{T}}$ and $\alpha_t = \frac{4F}{\eta_t}$, for any $\lambda_0 \in \mathcal{S}_m$, DR-OMMD achieves the following multi-objective regret*

$$R(T) \le G\sqrt{2\gamma DT}.$$

(ii) When $\eta_t = \frac{\sqrt{2\gamma D}}{G\sqrt{t}}$, $\alpha_t = \frac{4F}{\eta_t}$, for any $\lambda_0 \in \mathcal{S}_m$, DR-OMMD attains the following multi-objective regret

$$R(T) \le \frac{3}{2} G\sqrt{2\gamma DT}.$$

Next we show that the derived bound is tight w.r.t. m regarding any gradient-based algorithm. Specifically, we follow the standard worst-case analysis of deriving lower bounds and construct a special case in which any gradient-based algorithm will incur a regret in the order of $\Omega(\sqrt{T})$.

Assume $f_t^1 = f_t^2 = \cdots = f_t^m$ at each round t. In this case, the instantaneous gradients of all the objectives are identical, i.e., $g_t^i = \nabla f_t^i(x_t) \equiv \nabla f_t^1(x_t) = g_t^1, \forall i \in \{1, ..., m\}$. For any gradient-based algorithm, since it can only utilize the gradient information of the objectives, it cannot distinguish the objective to which a certain gradient belongs. Alternatively speaking, in this case, any multiple gradient algorithm will **treat all gradients in the same way** and thus behave like a single-objective algorithm using the single gradient g_t^1. Hence, in intuition, for any gradient-based algorithm, the worst-case bounds are **at least independent** of m. In particular, the worst-case bounds of gradient-based algorithms cannot decrease as m increases; otherwise, the above case will be violated.

In the following, we provide a detailed proof of the tightness of the $O(\sqrt{T})$ bound. In the above case, since $f_t^1 = f_t^2 = \cdots = f_t^m$ for any t, the cumulative losses of all the objectives are also identical, i.e., $\sum_{t=1}^T f_t^1 = \sum_{t=1}^T f_t^2 = \cdots = \sum_{t=1}^T f_t^m$. Therefore, the Pareto set $\mathcal{X}^*$ of the cumulative vector loss $\sum_{t=1}^T F_t$ coincides with the optimal decision set of the cumulative loss $\sum_{t=1}^T f_t^1$ of the first objective, i.e., $\mathcal{X}^* = \operatorname{argmin}_{x \in \mathcal{X}} \sum_{t=1}^T f_t^1(x)$. Recall our definition of the multi-objective regret. Since $\lambda^\top F_t(x) = f_t^1(x)$ for any $\lambda \in \mathcal{S}_m$, we have

$$R(T) = \sup_{x^* \in \mathcal{X}^*} (\sum_{t=1}^T f_t^1(x_t) - \sum_{t=1}^T f_t^1(x^*)) = \sum_{t=1}^T f_t^1(x_t) - \min_{x^* \in \mathcal{X}} \sum_{t=1}^T f_t^1(x^*),$$

which exactly reduces to the single-objective regret $R_S(T)$ defined by the losses $\{f_t^1\}_{t=1}^T$ of the first objective. Hence we have $R(T) = R_S(T)$ in this case. Since the losses $\{f_t^1\}_{t=1}^T$ of the first objective can be chosen adversarially, we can follow Section 3.2 in (Hazan et al., 2016) to construct a certain sequence $\{f_t^1\}_{t=1}^T$ that admits a lower single-objective regret bound of $\Omega(\sqrt{T})$. Hence in this certain case, any multiple gradient algorithm will admit a multi-objective regret $R(T) = \Omega(\sqrt{T})$ w.r.t. T and m, matching our derived regret bound for DR-OMMD in terms of both T and m.

Some readers may suspect it unreasonable that in the multi-objective setting, the derived regret bounds do not increase as m increases. Now we explicate the rationality of such independence in the following.

In fact, the independence of m lies in the adoption of PSG in the formulation of the regret. Recall that, in the definition of PSG, "$\exists i \in \{1, \ldots, m\}$" means that it just needs to pick one coordinate i to satisfy $f_t^i(x_t) - \epsilon < f_t^i(x'')$, which omits the dependency of m. We can see this point from another perspective. Recall that in the derivation of Proposition 1, we know that the regret $R(T)$ has an equivalent form, namely $\sup_{x^* \in \mathcal{X}^*} \inf_{\lambda^* \in \mathcal{S}_m} (\lambda^*)^\top \sum_{t=1}^T (F_t(x_t) - F_t(x^*))$, or equivalently $\sup_{x^* \in \mathcal{X}^*} \min_{i \in \{1,\ldots,m\}} \sum_{t=1}^T (f_t^i(x_t) - f_t^i(x^*))$. In particular, PSG takes a minimum operation over all objectives, and thus it does not necessarily increase as m increases.

There is another intuitive way that can help understand the rationality of the independence of m. As is well recognized in existing research in multi-objective optimization (Emmerich & Deutz, 2018), the proportion of the Pareto optimal solutions (or more precisely, non-dominated solutions) in the decision domain tends to increase rapidly as the number of objectives increases. As a consequence, it might not be harder to reach the Pareto optimal set when m turns larger, hence intuitively, the regret bound does not necessarily increase as m increases.

E.2 More Details in the Comparison with Linearization

Recall that in the remark below Theorem 1, we show that our derived bound for DR-OMMD is smaller than that of linearization, and discuss the margin between the two regret bounds in the two-objective setting with linear losses. We now summarize the result in Theorem 2 in the following.

Theorem 2. *Consider a two-objective optimization setting with linear losses. Suppose the loss functions are $f_t^1(x) = x^\top g_t^1$ and $f_t^2(x) = x^\top g_t^2$ at each round t. For any $\lambda_0 = (\gamma_0, 1 - \gamma_0) \in \mathcal{S}_m$, let $\lambda_t = (\gamma_t, 1 - \gamma_t)$ denote the composite weights produced by min-regularized-norm with $L1$-norm. When the regularization strength is set as $\alpha_t = 4F/\eta_t$, the margin between the regret bound of linearization with fixed weights λ_0 and that of DR-OMMD with composite weights λ_t is at least*

$$M \geq \sum_{t=1}^T \frac{\eta_t}{2} (\gamma_t - \gamma_0)^2 \|g_t^1 - g_t^2\|_2^2.$$

Before proving the theorem, we remark that the two bounds are basically in the same order. Note that this theoretical result is also very commonly seen in the offline setting, where multiple gradient algorithms often have the same (convergence) rate as linearization (Yu et al., 2020; Liu et al., 2021). The benefit of multiple gradient algorithms is mainly due to the implementation of gradient composition. For example, the concept of common descent (Sener & Koltun, 2018; Yu et al., 2020) eliminates the gradient conflicting issue; the resulting algorithm achieves substantial performance improvements compared to linearization in their experiments. In this paper, we move one step forward and discuss the margin between DR-OMMD and linearization. We show that such a margin is due to the gradient difference $g_t^1 - g_t^2$ and the gap between the pre-defined weights λ_0 and the adaptive weights λ_t. This is the best we can do for now. The regret bound comparison for $m \geq 3$ is left as future work.

Proof of Theorem 2. We first write out the regret bounds of both methods. For DR-OMMD with $\alpha_t = 4F/\eta_t$, Theorem 1 provides the following regret bound

$$R_{\text{DR-OMMD}}(T) \leq \frac{\gamma D}{\eta_T} + \sum_{t=1}^{T} \frac{\eta_t}{2} \min_{\lambda \in \mathcal{S}_m} \{\|\nabla F_t(x_t)\lambda\|^2 + \frac{4F}{\eta_t}\|\lambda - \lambda_0\|_1\}.$$

For linearization with fixed weights $\lambda_0 \in \mathcal{S}_m$, it can be viewed as single-objective optimization with linearized loss $\lambda_0^\top F_t$. Hence, we can directly borrow the tight bound of OMD (e.g., Theorem 6.8 in (Orabona, 2019)) and derive a bound

$$R_{\text{linear}}(T) \leq \frac{\gamma D}{\eta_T} + \sum_{t=1}^{T} \frac{\eta_t}{2} \|\nabla F_t(x_t)\lambda_0\|_2^2.$$

The margin between the above two bounds takes

$$M = \sum_{t=1}^{T} \frac{\eta_t}{2} (\|\nabla F_t(x_t)\lambda_0\|^2 - \min_{\lambda \in \mathcal{S}_m} \{\|\nabla F_t(x_t)\lambda\|^2 + \frac{4F}{\eta_t}\|\lambda - \lambda_0\|_1\}),$$

which stems from different choices of composite weights. We investigate the margin at each round.

Lemma 3. *In a two-objective setting, suppose the gradients are g_1 and g_2 at some specific round t, and the corresponding gradient matrix $G = [g_1, g_2]$. For any $\lambda_0 = (\gamma_0, 1 - \gamma_0) \in \mathcal{S}_m$, let $\lambda_t = (\gamma_t, 1 - \gamma_t)$ denote the composite weights produced by min-regularized-norm with $L1$-norm, then the following inequality holds, i.e.,*

$$\|G\lambda_0\|_2^2 - (\|G\lambda_t\|_2^2 + \alpha\|\lambda_t - \lambda_0\|_1) \geq (\gamma_0 - \gamma_t)^2 \|g_2 - g_1\|_2^2.$$

Proof. Denote the left side of the target inequality as $M(\lambda_t, \lambda_0)$, then it can be simplified as

$$\begin{aligned} M(\lambda_t, \lambda_0) &= (G(x_t)\lambda_0 - G\lambda_t)^\top (G\lambda_0 + G\lambda_t) - \alpha\|\lambda_t - \lambda_0\|_1 \\ &= (\lambda_0 - \lambda_t)^\top G^\top G(\lambda_0 + \lambda_t) - \alpha\|\lambda_t - \lambda_0\|_1 \end{aligned}$$

To leverage this term, we need to plug the derived composite weights λ_t into $M(\lambda_t, \lambda_0)$. Recall that in the two-objective setting, the weight γ_t is given as

$$\gamma_t = \max\{\min\{\gamma_t'', 1\}, 0\}, \quad \text{where } \gamma_t'' = \max\{\min\{\gamma_0, \gamma_R\}, \gamma_L\},$$

where $\gamma_L = (g_2^\top (g_2 - g_1) - \alpha)/\|g_2 - g_1\|^2$ and $\gamma_R = (g_2^\top (g_2 - g_1) + \alpha)/\|g_2 - g_1\|^2$. Since the maximum and minimum operations will truncate the value of the produced weight, we now calculate $M(\lambda_t, \lambda_0)$ by case. Specifically, notice that $\gamma_L < \gamma_R$ and $0 \leq \gamma_0 \leq 1$, we consider the following cases.

Case 1: When $\gamma_R < 0$, we must have $\gamma_L < \gamma_R < 0 \leq \gamma_0$, which leads to $\gamma_t = 0$. In this case, we have $\lambda_0 - \lambda_t = (\gamma_0, -\gamma_0)$ and $\lambda_0 + \lambda_t = (\gamma_0, 2 - \gamma_0)$. Therefore, $M(\lambda_t, \lambda_0)$ can be computed as

$$M(\lambda_t, \lambda_0) = (\gamma_0 g_1 - \gamma_0 g_2)^\top (\gamma_0 g_1 + (2 - \gamma_0) g_2) - 2\alpha\lambda_0.$$

Also, from the condition $\gamma_R < 0$, we have $\alpha < g_2^\top (g_1 - g_2)$. Since $\lambda_0 \geq 0$, plugging it into the above inequality, we have

$$M(\lambda_t, \lambda_0) \geq (\gamma_0 g_1 - \gamma_0 g_2)^\top (\gamma_0 g_1 - \gamma_0 g_2) = \gamma_0^2 \|g_1 - g_2\|^2.$$

In this case, since $\gamma_t = 0$, we derive the desired inequality.

Case 2: When $\gamma_L > 1$, we must have $\gamma_0 \leq 1 < \gamma_L < \gamma_R$, which results in $\gamma_t = 1$. In this case, we have $\lambda_0 - \lambda_t = (\gamma_0 - 1, 1 - \gamma_0)$ and $\lambda_0 + \lambda_t = (\gamma_0 + 1, 1 - \gamma_0)$. Now $M(\lambda_t, \lambda_0)$ can be calculated as

$$M(\lambda_t, \lambda_0) = ((\gamma_0 - 1)g_1 + (1 - \gamma_0)g_2)^\top((\gamma_0 + 1)g_1 + (1 - \gamma_0)g_2) - 2\alpha(1 - \gamma_0).$$

Notice that the condition $\gamma_L > 1$ gives $\alpha < g_1^\top(g_2 - g_1)$. Since $1 - \gamma_0 \geq 0$, plugging it into the above inequality, we have

$$M(\lambda_t, \lambda_0) \geq ((\gamma_0 - 1)g_1 + (1 - \gamma_0)g_2)^\top((\gamma_0 - 1)g_1 + (1 - \gamma_0)g_2) = (1 - \gamma_0)^2\|g_1 - g_2\|^2.$$

In this case, since $\gamma_t = 1$, we derive the desired inequality.

Case 3: When $0 \leq \gamma_L \leq \gamma_R \leq 1$, the margin is a bit more complex since the value of λ_t further depends on the relation between λ_0, λ_L, and λ_R. Specifically, we consider the following cases.

(i) If $0 \leq \gamma_L \leq \gamma_0 \leq \gamma_R \leq 1$, then $\gamma_t = \gamma_0$. In this case, the inequality trivially holds.

(ii) If $0 \leq \gamma_0 \leq \gamma_L \leq 1$, then $\gamma_t = \gamma_L$. In this case, since $\|\lambda_t - \lambda_0\|_1 = 2(\gamma_L - \gamma_0)$, we can calculate $M(\lambda_t, \lambda_0)$ as

$$\begin{aligned} M(\lambda_t, \lambda_0) &= ((\gamma_0 - \gamma_L)g_1 + (\gamma_L - \gamma_0)g_2)^\top((\gamma_0 + \gamma_L)g_1 + (2 - \gamma_0 - \gamma_L)g_2) - 2\alpha(\gamma_L - \gamma_0) \\ &= (\gamma_0 - \gamma_L)((g_1 - g_2)^\top((\gamma_0 + \gamma_L)g_1 + (2 - \gamma_0 - \gamma_L)g_2) + 2\alpha). \end{aligned}$$

Since $\gamma_t = \gamma_L = (g_2^\top(g_2 - g_1) - \alpha)/\|g_2 - g_1\|^2$, we have $\alpha = (g_1 - g_2)^\top(-\gamma_L g_1 + (\gamma_L - 1)g_2)$. Therefore, we further have

$$\begin{aligned} (g_1 - g_2)^\top((\gamma_0 + \gamma_L)g_1 + (2 - \gamma_0 - \gamma_L)g_2) + 2\alpha &= (g_1 - g_2)^\top((\gamma_0 - \gamma_L)g_1 + (\gamma_L - \gamma_0)g_2) \\ &= (g_1 - g_2)^\top(\gamma_0 - \gamma_L)(g_1 - g_2). \end{aligned}$$

Plugging it into the above equation on $M(\lambda_t, \lambda_0)$, we derive

$$M(\lambda_t, \lambda_0) = (\gamma_L - \gamma_0)^2\|g_1 - g_2\|_2^2$$

(iii) If $0 \leq \gamma_R \leq \gamma_0 \leq 1$, then $\gamma_t = \gamma_R$. In this case, $\|\lambda_t - \lambda_0\|_1 = 2(\gamma_0 - \gamma_R)$, and $M(\lambda_t, \lambda_0)$ can be calculated as

$$\begin{aligned} M(\lambda_t, \lambda_0) &= ((\gamma_0 - \gamma_R)g_1 + (\gamma_R - \gamma_0)g_2)^\top((\gamma_0 + \gamma_R)g_1 + (2 - \gamma_0 - \gamma_R)g_2) - 2\alpha(\gamma_0 - \gamma_R) \\ &= (\gamma_0 - \gamma_R)((g_1 - g_2)^\top((\gamma_0 + \gamma_R)g_1 + (2 - \gamma_0 - \gamma_R)g_2) - 2\alpha). \end{aligned}$$

Since $\gamma_t = (g_2^\top(g_2 - g_1) + \alpha)/\|g_2 - g_1\|^2$, we have $\alpha = (g_1 - g_2)^\top(\gamma_R g_1 + (1 - \gamma_R)g_2)$, then

$$(g_1 - g_2)^\top((\gamma_0 + \gamma_R)g_1 + (2 - \gamma_0 - \gamma_R)g_2) - 2\alpha = (g_1 - g_2)^\top(\gamma_0 - \gamma_R)(g_1 - g_2).$$

Plugging it into the above equation on $M(\lambda_t, \lambda_0)$, we derive

$$M(\lambda_t, \lambda_0) = (\gamma_R - \gamma_0)^2\|g_1 - g_2\|_2^2$$

Combining all of the above cases, we prove the lemma. ■

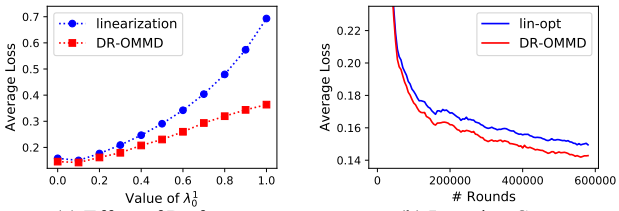
For any $t \in \{1, ..., T\}$, set $g_1 = g_t^1, g_2 = g_t^2$ (i.e., $G = G_t$), and $\alpha = \frac{4F}{\eta_t}$ in Lemma 3, we have

$$\|G_t\lambda_0\|_2^2 - (\|G_t\lambda_t\|_2^2 + \frac{4F}{\eta_t}\|\lambda_t - \lambda_0\|_1) \geq (\gamma_t - \gamma_0)^2\|g_t^2 - g_t^1\|_2^2.$$

Since $\|\lambda_t - \lambda_0\|_2^2 = 2(\gamma_t - \gamma_0)^2$, summing the above inequality over $t \in \{1, ..., T\}$, we can directly calculate the margin as

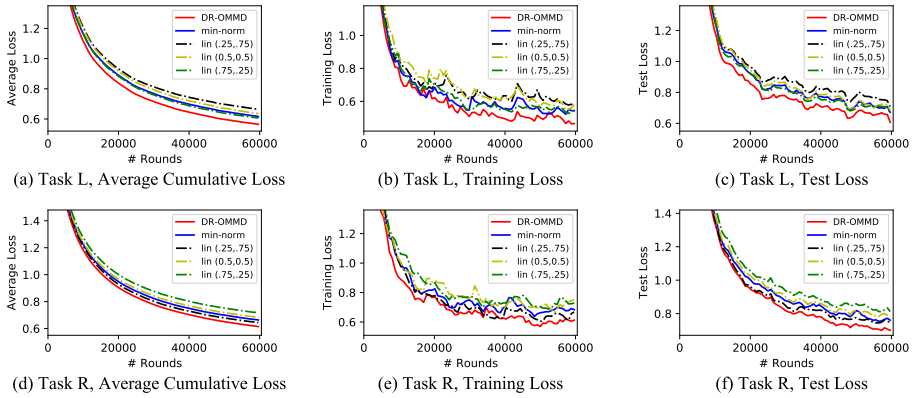
$$M \geq \sum_{t=1}^{T} \frac{\eta_t}{4}\|\lambda_t - \lambda_0\|_2^2 \cdot \|g_t^2 - g_t^1\|_2^2,$$

which proves the theorem. ■



(a) Effect of Preference (b) Learning Curve

Figure 3: Results to verify the effectiveness of adaptive regularization on *covtype*. (a) Performance of DR-OMMD and linearization under varying $\lambda_0 = (\lambda_0^1, 1-\lambda_0^1)$. (b) Performance using the optimal weights $\lambda_0 = (0.1, 0.9)$.



(a) Task L, Average Cumulative Loss (b) Task L, Training Loss (c) Task L, Test Loss

(d) Task R, Average Cumulative Loss (e) Task R, Training Loss (f) Task R, Test Loss

Figure 4: More empirical results of the multi-task deep online experiments on MultiMNIST. The plots show the average cumulative loss, training loss and test loss of DR-OMMD and various baselines on both tasks (*task L* and *task R*).

F More Experimental Results

F.1 More Details of the Experimental Setup

The *protein* and *covtype* datasets used in our experiments are publicly available in (Dua & Graff, 2017). The *MultiMNIST* dataset is acquired by the code provided by (Sener & Koltun, 2018).

All runs are deployed on Xeon(R) E5-2699 @ 2.2GHz.

F.2 More Results for Adaptive Regularization

We supplement the empirical results on *covtype* in Figure 3, which have been omitted from our main paper due to the lack of space. These results are consistent with the results on *protein* as presented in our main paper.

F.3 More Results for Online Deep Multi-Task Learning

In our main paper, due to the page limit, Figure 2 only reports the average cumulative loss of DR-OMMD and various baselines on MultiMNIST. Here we supplement the results on the training loss and the test loss in Figure 4. The results are consistent with the average cumulative loss, showing superiority of DR-OMMD over linearization and MGDA in the non-convex setting.

G Omitted Proofs of Proposition 2 (Min-Norm May Incur Linear Regrets)

Proof. As we have described in our main paper, we consider the following two-objective optimization problem. Decision domain is set as $\mathcal{X} = \{(u, v) \mid u + v \leq \frac{1}{2}, v - u \leq \frac{1}{2}, v \geq 0\}$. At each

round t, the loss function $F_t : \mathcal{X} \to \mathbb{R}^2$ takes

$$F_t(x) = \begin{cases} (\|x-a\|^2, \|x-b\|^2), & t = 2k-1, k = 1, 2, ...; \\ (\|x-b\|^2, \|x-c\|^2), & t = 2k, k = 1, 2, ..., \end{cases}$$

where $a = (-2, -1), b = (0, 1), c = (2, -1)$. For simplicity of analysis, we first consider the case when the total time horizon T is an even number. Then it can be checked that the cumulative loss function takes

$$\begin{aligned} \sum_{t=1}^{T} F_t(x) &= \frac{T}{2} \cdot (\|x-a\|^2 + \|x-b\|^2, \|x-b\|^2 + \|x-c\|^2) \\ &= T \cdot ((u+1)^2 + v^2 + 2, (u-1)^2 + v^2 + 2), \end{aligned}$$

for any $x = (u, v) \in \mathcal{X}$. Obviously the Pareto optimal set $\mathcal{X}^*$ of the cumulative loss coincides with the line segment between $(-1, 0)$ and $(1, 0)$, i.e., $\mathcal{X}^* = \{(u, v) \mid -\frac{1}{2} \leq u \leq \frac{1}{2}, v = 0\}$ (note that $\mathcal{X}^*$ is the intersection of the line segment and $\mathcal{X}$).

Now consider equipping OMD with vanilla min-norm, where the composite gradients are produced by the min-norm method. Suppose the learning process starts at any $x_1 = (u_1, v_1) \in \mathcal{X}$ such that $v_1 > 0$. Note that this is true if and only if $x_1 \notin \mathcal{X}^*$. Then for the iterate $x_t = (u_t, v_t)$ at each round t, we can directly calculate the gradients as

$$g_t^1 = \begin{cases} 2(x_t - a) = (2u_t + 4, & 2v_t + 2), & t = 2k-1; \\ 2(x_t - b) = (2u_t, & 2v_t - 2), & t = 2k. \end{cases}$$

$$g_t^2 = \begin{cases} 2(x_t - b) = (2u_t, & 2v_t - 2), & t = 2k-1; \\ 2(x_t - c) = (2u_t - 4, & 2v_t + 2), & t = 2k. \end{cases}$$

The min-norm weights can be computed as $\lambda_t = (\gamma_t, 1 - \gamma_t)$ where

$$\gamma_t = \begin{cases} \dfrac{(x_t - b)^\top (a - b)}{\|a - b\|^2} = \dfrac{1 - u_t - v_t}{4}, & t = 2k-1; \\ \dfrac{(x_t - c)^\top (b - c)}{\|b - c\|^2} = \dfrac{3 - u_t + v_t}{4}, & t = 2k. \end{cases}$$

The composite gradient

$$g_t^{comp} = \begin{cases} \gamma_t \cdot 2(x-a) + (1-\gamma_t) \cdot 2(x-b) = (u_t - v_t + 1, & -u_t + v_t - 1), & t = 2k-1; \\ \gamma_t \cdot 2(x-b) + (1-\gamma_t) \cdot 2(x-c) = (-u_t - v_t - 1, & -u_t - v_t - 1), & t = 2k. \end{cases}$$

Recall that the update form of OMD takes

$$x_{t+1} = \Pi_{\mathcal{X}}(x_t - \eta_t g_t^{comp}),$$

where $\eta_t > 0$ is the learning rate and $\Pi_{\mathcal{X}}$ is the projection operation onto $\mathcal{X}$. Denote the iterate $x_t = (u_t, v_t)$ at each round. Now we can investigate the relation between x_t and x_{t+1} by considering the following two cases:

(i) If $x_t - \eta_t g_t^{comp} \in \mathcal{X}$, then we do not need projection, and directly have $x_{t+1} = x_t - \eta_t g_t^{comp}$.

(ii) If $x_t - \eta_t g_t^{comp} \notin \mathcal{X}$, then we need to project $x_t - \eta_t g_t^{comp}$ back to $\mathcal{X}$. Denote $x'_{t+1} = x_t - \eta_t g_t^{comp}$. For simplicity we consider the projection based on the Euclidean distance, namely $\Pi_{\mathcal{X}}(x) = \arg\min_{x' \in \mathcal{X}} \|x - x'\|_2^2$. Since the composite gradient is orthogonal to the boundary on which the iterate after projection $x_{t+1} = \Pi_{\mathcal{X}}(x'_{t+1})$ is located, it can be checked that x_{t+1} lies on the line segment linking x_t and x'_{t+1}. Alternatively speaking, x_{t+1} can be expressed as $x_t - \eta'_t g_t^{comp}$ for some $0 \leq \eta'_t < \eta_t$.

Combining the above two cases, we know that at each round t, there exists some $\eta'_t \in [0, \eta_t]$ such that $x_{t+1} = x_t - \eta'_t g_t^{comp}$. Now we can analyze the relation between each entry of x_t and x_{t+1}. Specifically, since the second entry of the composite gradient is always non-positive, namely $-u_t + v_t - 1 \leq 0$ and $-u_t - v_t - 1 \leq 0$, we have $v_{t+1} \geq v_t$ for any t. Moreover, since the first entry of g_t^{comp} is non-negative when $t = 2k-1$, namely $u_{2k-1} - v_{2k-1} + 1 \geq 0$, we have $u_{2k} \leq u_{2k-1}$ for

any k; since the first entry of g_t^{comp} is non-positive when $t = 2k$, namely $-u_{2k} - v_{2k} - 1 \leq 0$, we have $u_{2k+1} \geq u_{2k}$ for any k.

Now we can go back to analyze the gap between the composite weights at any two consecutive rounds. It is easy to verify that $\gamma_{2k-1} < \gamma_{2k}$ and $\gamma_{2k} > \gamma_{2k+1}$, hence we have

$$\|\lambda_{2k} - \lambda_{2k-1}\|_1 = 2(\gamma_{2k} - \gamma_{2k-1}) = \frac{2 - (u_{2k} - u_{2k-1}) + (v_{2k} + v_{2k-1})}{2} \geq 1 + v_1,$$
$$\|\lambda_{2k+1} - \lambda_{2k}\|_1 = 2(\gamma_{2k} - \gamma_{2k+1}) = \frac{2 - (u_{2k} - u_{2k+1}) + (v_{2k} + v_{2k+1})}{2} \geq 1 + v_1.$$

Therefore, the composite weights λ_t indeed change radically at any two consecutive rounds.

The above analysis on v_t also implies the failure of min-norm in this problem. Recall that any Pareto optimal solution $x^* = (u^*, v^*) \in \mathcal{X}^*$ must satisfy $v^* = 0$. Suppose the initial iterate $x_1 = (u_1, v_1)$ does not lie in $\mathcal{X}^*$, i.e., $v_1 > 0$, which is almost sure for random initialization $x_1 \in \mathcal{X}$. Then we iteratively have $0 < v_1 \leq v_2 \leq ... \leq v_T$, which means that x_t moves away from the Pareto set $\mathcal{X}^*$.

In the following, we strictly prove that min-norm indeed incurs a **linear** multi-objective regret. To calculate $R(T)$, we first investigate the quantity $R(x^*, \lambda) = \lambda^\top \sum_{t=1}^T (F_t(x_t) - F_t(x^*))$ for any fixed weights $\lambda = (\gamma, 1-\gamma) \in \mathcal{S}_2$ and best fixed decision $x^* = (u^*, 0) \in \mathcal{X}^*$. Specifically, recall the form of $\sum_{t=1}^T F_t$ derived above, then we have

$$\lambda^\top \sum_{t=1}^T F_t(x^*) = (\gamma(u^* + 1)^2 + (1-\gamma)(u^* - 1)^2 + 2)T.$$

Denote the cumulative loss $\sum_{t=1}^T F_t(x_t) = (L_1, L_2)$, we now consider the loss of each objective L_1 and L_2 separately. Specifically, for the first objective, we have

$$L_1 = \sum_{k=1}^{T/2} ((u_{2k-1} + 2)^2 + u_{2k}^2 + (v_{2k-1} + 1)^2 + (v_{2k} - 1)^2).$$

Since $0 < v_1 \leq v_2 \leq ... \leq v_T \leq 1$, for the term regarding v_t we have

$$\begin{aligned}\sum_{k=1}^{T/2} ((v_{2k-1} + 1)^2 + (v_{2k} - 1)^2) &= (v_1 + 1)^2 + (v_T - 1)^2 + \sum_{k=1}^{T/2-1} ((v_{2k} - 1)^2 + (v_{2k+1} + 1)^2) \\ &\geq \sum_{k=1}^{T/2-1} ((v_{2k} - 1)^2 + (v_{2k} + 1)^2) = \sum_{k=1}^{T/2-1} (2v_{2k}^2 + 2) \\ &\geq \sum_{k=1}^{T/2-1} (2v_1^2 + 2) = (2v_1^2 + 2)(\frac{T}{2} - 1) \geq v_1^2 T + T - 2.\end{aligned}$$

For the k-th term regarding u_t, we have

$$(u_{2k-1} + 2)^2 + u_{2k}^2 = (u_{2k-1} + 1)^2 + (u_{2k} + 1)^2 + 2(u_{2k-1} - u_{2k}) + 2.$$

Recall that we have derived $u_{2k} \leq u_{2k-1}$, thus we have

$$\sum_{k=1}^{T/2} (u_{2k-1} + 2)^2 + u_{2k}^2 \geq \sum_{k=1}^{T/2} ((u_{2k-1} + 1)^2 + (u_{2k} + 1)^2 + 2) \geq \sum_{t=1}^{T} (u_t + 1)^2 + T \geq (\bar{u} + 1)^2 T + T,$$

where $\bar{u} = \frac{1}{T} \sum_{t=1}^T u_t$ and the last inequality is derived from Jensen's inequality. In summary, for the cumulative loss L_1 of the first objective, we have

$$L_1 \geq (\bar{u} + 1)^2 T + v_1^2 T + 2T - 2.$$

Similarly, we can analyze the cumulative loss L_2 of the second objective

$$L_2 = \sum_{k=1}^{T/2} (u_{2k-1}^2 + (u_{2k} - 2)^2 + (v_{2k-1} - 1)^2 + (v_{2k} + 1)^2).$$

Since $0 < v_1 \leq v_2 \leq ... \leq v_T \leq 1$, for the term regarding v_t we have

$$\sum_{k=1}^{T/2}((v_{2k-1}-1)^2+(v_{2k}+1)^2) \geq \sum_{k=1}^{T/2}((v_{2k-1}-1)^2+(v_{2k-1}+1)^2) \geq v_1^2T+T.$$

For the term regarding u_t, we also have

$$\begin{aligned}\sum_{k=1}^{T/2}(u_{2k-1}^2+(u_{2k}-2)^2) &= \sum_{k=1}^{T/2}((u_{2k-1}-1)^2+(u_{2k}-1)^2+2(u_{2k-1}-u_{2k})+2)\\ &\geq \sum_{t=1}^{T}(u_t-1)^2+T \geq (\bar{u}-1)^2T+T,\end{aligned}$$

where the last inequality is derived from Jensen's inequality. Therefore, we have

$$L_2 \geq (\bar{u}-1)^2T+v_1^2T+2T.$$

Combining the above inequalities, we have

$$\begin{aligned}R(x^*,\lambda) &= \gamma L_1+(1-\gamma)L_2-\lambda^\top\sum_{t=1}^{T}F_t(x^*)\\ &\geq \gamma((\bar{u}+1)^2-(u^*+1)^2)+(1-\gamma)((\bar{u}-1)^2-(u^*-1)^2)+v_1^2T-2\gamma.\end{aligned}$$

For any $\lambda \in \mathcal{S}_2$ (i.e., $\gamma \in [0,1]$), set $x' = (\bar{u}, 0) \in \mathcal{X}^*$, then it holds that

$$R(x',\lambda) \geq v_1^2T-2.$$

Equivalently, the multi-objective regret satisfies

$$R(T) = \sup_{x^*\in\mathcal{X}^*}\inf_{\lambda\in\mathcal{S}_2}R(x^*,\lambda) \geq \inf_{\lambda\in\mathcal{S}_2}R(x',\lambda) \geq v_1^2T-2,$$

which is linear w.r.t. T for any $x_1 = (u_1, v_1) \in \mathcal{X}$ such that $v_1 > 0$.

We now investigate the case when T is an odd number. Since the calculation of the composite weights λ_t and the composite gradient g_t^{comp} at each round is independent of the total time horizon T, we still have $\|\lambda_{t+1}-\lambda_t\|_1 \geq v_1+1$ for any t. Hence the first desired property also holds for any odd T.

It remains to prove that OMD with min-norm still incurs a linear regret when T is odd. In this case, the Pareto optimal set $\mathcal{X}^*$ does not lie in the x-axis anymore, hence it is difficult to directly compute $R(T)$. However, we can still use our derived $R(x^*,\lambda)$ for any even T to estimate the regret. Specifically, set $x' = (\frac{1}{T-1}\sum_{t=1}^{T-1}u_t, 0)$; from the above derivation with even T, for any $\lambda \in \mathcal{S}_2$, we still have (note that now $T-1$ is an even number)

$$\lambda^\top\sum_{t=1}^{T-1}F_t(x_t)-\lambda^\top\sum_{t=1}^{T-1}F_t(x') \geq v_1^2T-2.$$

Since for any $x \in \mathcal{X}$, we have $0 \leq \|x-a\|^2, \|x-b\|^2, \|x-c\|^2 \leq 10$, we have

$$R(x',\lambda) = \lambda^\top\sum_{t=1}^{T}F_t(x_t)-\lambda^\top\sum_{t=1}^{T}F_t(x') \geq v_1^2T-12.$$

Furthermore, from the definition of Pareto optimality, there exists some $x'' \in \mathcal{X}^*$ that Pareto dominates x' regarding the cumulative loss $\sum_{t=1}^{T}F_t$, namely $\sum_{t=1}^{T}F_t(x'') \preceq \sum_{t=1}^{T}F_t(x')$. Hence

$$R(x'',\lambda) = \lambda^\top\sum_{t=1}^{T}F_t(x_t)-\lambda^\top\sum_{t=1}^{T}F_t(x'') \geq R(x',\lambda),$$

for any $\lambda \in \mathcal{S}_2$. Therefore, the multi-objective regret

$$R(T) = \sup_{x^*\in\mathcal{X}^*}\inf_{\lambda\in\mathcal{S}_2}R(x^*,\lambda) \geq \inf_{\lambda\in\mathcal{S}_2}R(x'',\lambda) \geq \inf_{\lambda\in\mathcal{S}_2}R(x',\lambda) \geq v_1^2T-12,$$

which is also linear w.r.t. T for any $x_1 = (u_1, v_1) \in \mathcal{X}$ such that $v_1 > 0$. ■

H OMITTED PROOFS OF THE REGRET BOUND FOR DR-OMMD

H.1 PROOF OF THEOREM 1

Proof. We start from the definition of the multi-objective regret $R_{\mathrm{II}}(T)$ (which is abbreviated as $R(T)$). Specifically, for any $\lambda \in \mathcal{S}_m$ and $\lambda_1 \ldots, \lambda_T \in \mathcal{S}_m$, it holds that

$$\begin{aligned}
R(T) &= \sup_{x^* \in \mathcal{X}^*} \inf_{\lambda^* \in \mathcal{S}_m} \sum_{t=1}^{T} {\lambda^*}^\top (F_t(x_t) - F_t(x^*)) \leq \sup_{x^* \in \mathcal{X}^*} \sum_{t=1}^{T} \lambda^\top (F_t(x_t) - F_t(x^*)) \\
&= \sup_{x^* \in \mathcal{X}^*} \sum_{t=1}^{T} \Big((\lambda - \lambda_t)^\top F_t(x_t) + {\lambda_t}^\top (F_t(x_t) - F_t(x^*)) + (\lambda_t - \lambda)^\top F_t(x^*) \Big) \\
&\leq \sum_{t=1}^{T} F \|\lambda - \lambda_t\|_1 + \sup_{x^* \in \mathcal{X}^*} \sum_{t=1}^{T} {\lambda_t}^\top (F_t(x_t) - F_t(x^*)) + \sum_{t=1}^{T} F \|\lambda - \lambda_t\|_1 \\
&= 2F \sum_{t=1}^{T} \|\lambda - \lambda_t\|_1 + \sup_{x^* \in \mathcal{X}^*} \sum_{t=1}^{T} {\lambda_t}^\top (F_t(x_t) - F_t(x^*)).
\end{aligned}$$

To proceed, notice that if the composite weights λ_t are given beforehand instead of being calculated via min-regularized-norm, DR-OMMD acts just like standard OMD using linearized loss $\lambda_t F_t$. Hence, the second term in the above regret bound can be further analyzed in a similar way as single-objective OMD (Srebro et al., 2011; Cesa-bianchi et al., 2012). Specifically, at each round t, since F_t is coordinate-wise convex, the linearized loss $\lambda_t^\top F_t$ is also convex. Also notice that the composite gradient $g_t = \nabla F_t(x_t)\lambda_t$ is exactly the gradient of $\lambda_t^\top F_t$ at x_t. Hence for any $x^* \in \mathcal{X}^*$, we have

$$\lambda_t^\top F_t(x_t) - \lambda_t^\top F_t(x^*) \leq g_t^\top (x_t - x^*) = g_t^\top (x_{t+1} - x^*) + g_t^\top (x_t - x_{t+1}).$$

From the first-order optimal condition of x_{t+1}, for any $x' \in \mathcal{X}$, we have

$$(\eta_t \nabla F_t(x_t)\lambda_t + \nabla R(x_{t+1}) - \nabla R(x_t))^\top (x' - x_{t+1}) \geq 0.$$

Recall that $g_t = \nabla F_t(x_t)\lambda_t$. We set $x' = x^*$ and combine the above two inequalities, which derives

$$\lambda_t^\top F_t(x_t) - \lambda_t^\top F_t(x^*) \leq \frac{1}{\eta_t} (\nabla R(x_{t+1}) - \nabla R(x_t))^\top (x^* - x_{t+1}) + g_t^\top (x_t - x_{t+1}).$$

Recall the definition of Bregman divergence B_R. We can check that (also see (Beck & Teboulle, 2003))

$$B_R(x^*, x_t) - B_R(x^*, x_{t+1}) - B_R(x_{t+1}, x_t) = (\nabla R(x_{t+1}) - \nabla R(x_t))^\top (x^* - x_{t+1}).$$

Since R is 1-strongly convex, we have $B_R(x_{t+1}, x_t) \geq \|x_{t+1} - x_t\|^2/2$. Hence

$$\lambda_t^\top F_t(x_t) - \lambda_t^\top F_t(x^*) \leq \frac{1}{\eta_t} (B_R(x^*, x_t) - B_R(x^*, x_{t+1}) - \frac{1}{2}\|x_{t+1} - x_t\|^2) + g_t^\top (x_t - x_{t+1}).$$

Moreover, from the Cauchy-Schwartz inequality we have

$$g_t^\top (x_t - x_{t+1}) \leq \frac{\eta_t}{2} \|g_t\|_*^2 + \frac{1}{2\eta_t} \|x_t - x_{t+1}\|^2.$$

Combining the above two inequalities, we derive

$${\lambda_t}^\top F_t(x_t) - {\lambda_t}^\top F_t(x^*) \leq \frac{1}{\eta_t} (B_R(x^*; x_t) - B_R(x^*; x_{t+1})) + \frac{\eta_t}{2} \|\nabla F_t(x_t)\lambda_t\|_*^2,$$

for any $x^* \in \mathcal{X}^*$. Summing it over $t \in \{1, ..., T\}$ and utilizing $B_R(x^*; x_{T+1}) \geq 0$, we have

$$\sum_{t=1}^{T} {\lambda_t}^\top (F_t(x_t) - F_t(x^*)) \leq \frac{1}{\eta_1} B_R(x^*; x_1) + \sum_{t=2}^{T} (\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}) B_R(x^*; x_T) + \sum_{t=1}^{T} \frac{\eta_t}{2} \|\nabla F_t(x_t)\lambda_t\|_*^2.$$

Since $B_R(x^*; x_t) \leq \gamma D$ and $\eta_t \leq \eta_{t-1}$ for any t, we have $(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}) B_R(x^*; x_T) \leq (\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}})\gamma D$. Hence we further have

$$\sum_{t=1}^{T} {\lambda_t}^\top (F_t(x_t) - F_t(x^*)) \leq \frac{\gamma D}{\eta_T} + \sum_{t=1}^{T} \frac{\eta_t}{2} \|\nabla F_t(x_t)\lambda_t\|_*^2.$$

Taking the supremum over $x^* \in \mathcal{X}^*$ and plugging it back to the above regret bound, we prove the theorem. ■

H.2 PROOF OF COROLLARY 1

Proof. We set $\alpha_t = \frac{4F}{\eta_t}$ in DR-OMMD, and specify λ_t to be the composition weights generated by the algorithm at each round t. From the formulation of min-regularized-norm with $L1$-norm, we know that $\lambda_t \in \arg\min_{\lambda \in \mathcal{S}_m} \|\nabla F_t(x_t)\lambda\|_*^2 + \frac{4F}{\eta_t}\|\lambda - \lambda_0\|_1$. In particular, we have $\|\nabla F_t(x_t)\lambda_t\|_*^2 + \frac{4F}{\eta_t}\|\lambda_t - \lambda_0\|_1 \leq \|\nabla F_t(x_t)\lambda_0\|_*^2$. Therefore, from Theorem 1 we know that

$$R(T) \leq \frac{\gamma D}{\eta_T} + \frac{\eta_t}{2}\sum_{t=1}^{T}\|\nabla F_t(x_t)\lambda_0\|_*^2.$$

Utilize Assumption 2, and set $\eta_t = \frac{\sqrt{2\gamma D}}{G\sqrt{T}}$, then we have

$$R(T) \leq \frac{\gamma D}{\eta_T} + \frac{\eta_t}{2}G^2 T = G\sqrt{2\gamma DT},$$

which proves the bound with the fixed optimal learning rate.

Alternatively, set $\eta_t = \frac{\sqrt{2\gamma D}}{G\sqrt{t}}$ and utilize $\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \leq 2\sqrt{T}$, we also have

$$R(T) \leq \frac{1}{2}G\sqrt{2\gamma DT}(1 + \sum_{t=1}^{T}\frac{1}{\sqrt{t}}) \leq \frac{3}{2}G\sqrt{2\gamma DT},$$

which proves the bound with the adaptive learning rate. ■