ENHANCE THE DYNAMIC REGRET VIA OPTIMISM

Anonymous authors
Paper under double-blind review

ABSTRACT

In this paper, we study the enhancement method for dynamic regret in online convex optimization. Existing works have shown that adaptive learning for dynamic environment (Ader) enjoys an $O\left(\sqrt{(1+P_T)\,T}\right)$ dynamic regret upper bound, where T is the number of rounds and P_T is the path length of the reference strategy sequence. The basic idea of Ader is to maintain a group of experts, where each expert obtains the best dynamic regret of a specific path length by running Mirror Descent (MD) with specific parameter, and then tracks the best expert by Normalized Exponentiated Subgradient (NES). However, Ader is not environmental adaptive. By introducing the estimated linear loss function $\widehat{x}_t^*$, the dynamic regret for Optimistic Mirror Descent (OMD) is tighter than MD if the environment is not completely adversarial and $\widehat{x}_t^*$ is well-estimated. Based on the fact that optimism can enhance dynamic regret, we develop an algorithm to replace MD and NES in Ader with OMD and Optimistic Normalized Exponentiated Subgradient (ONES) respectively, and utilize the adaptive trick to achieve $O\left(\sqrt{(1+P_T)\,M_T}\right)$ dynamic regret upper bound, where $M_T \leqslant O(T)$ is a measure of estimation accuracy. In particular, if $\widehat{x}_t^* \in \partial\widehat{\varphi}_t$, where $\widehat{\varphi}_t$ represents the estimated convex loss function and $\partial\widehat{\varphi}_t$ is Lipschitz continuous, then the dynamic regret upper bound of OMD has a subgradient variation type. Based on this fact, we develop a variant algorithm whose upper bound has a subgradient variation type. All our algorithms are environmental adaptive.

1 INTRODUCTION

The Online Convex Optimization (OCO), which was introduced by Zinkevich (2003), plays a vital role in online learning as its interesting theory and wide application (Shalev-Shwartz, 2012). The OCO problem can be viewed as repeated games between the learner and the adversary: At round t, the learner chooses a map x_t from a hypothesis class C for prediction, and the adversary feeds back a convex loss function φ_t, then the learner suffers an instantaneous loss $\varphi_t(x_t)$. In general, $\varphi_t(x_t)$ is bounded to exclude the case where the loss can be arbitrarily large.

The appropriate performance metric, namely regret, as described below, comes from game theory since the framework of OCO is game-theoretic and adversarial in nature (Zinkevich, 2003; Hazan, 2019).

$$\underset{(z_1,z_2,\cdots,z_T)}{\text{regret}} \mathscr{A} := \sum_{t=1}^{T} \varphi_t(x_t) - \sum_{t=1}^{T} \varphi_t(z_t), \tag{1}$$

where $z_t \in C$ represents the reference strategy in round t, and $\mathscr{A}$ is the algorithm that generates x_t. Particularly, if $z_t \equiv z$, we have the following static regret,

$$\underset{(z,z,\cdots,z)}{\text{regret}} \mathscr{A} := \sum_{t=1}^{T} \varphi_t(x_t) - \sum_{t=1}^{T} \varphi_t(z).$$

Correspondingly, we call Eq. (1) dynamic regret. The static regret used in most literature is usually written as $\sup_{z\in C} \text{regret}_{(z,z,\cdots,z)} \mathscr{A}$. There are plenty of works devoted to designing online algorithms to minimize static regret (Cesa-Bianchi & Lugosi, 2006; Shalev-Shwartz, 2012; Hazan, 2019; Orabona, 2019). Recently, designing online algorithms to minimize dynamic regret has attracted much attention (Hall & Willett, 2013; Jadbabaie et al., 2015; Mokhtari et al., 2016; Zhang et al., 2018; Zhao et al., 2020; Campolongo & Orabona, 2021; Kalhan et al., 2021).

An online algorithm is environmental adaptive if it maintains the regret upper bound when the environment is adversarial, and tightens the upper bound as much as possible when the environment is not completely adversarial. Optimistic algorithm provides a way to achieve environmental adaptive. The word "optimistic" refers to the idea that if the learner can predict the impending loss when the environment is not completely adversarial, then the regret upper bound may be tightened. How to predict the impending loss is not the focus of the optimistic algorithm. The Optimistic Mirror Descent (OMD) was proposed by Chiang et al. (2012) and extended by Rakhlin & Sridharan (2013).

The regret upper bound usually contains some characteristic terms. The following are three well-known characteristic terms.

- The path length term (Zinkevich, 2003), $P_T = \sum_{t=2}^{T} \|z_t - z_{t-1}\|$.
- The gradient variation term (Chiang et al., 2012), $V_T = \sum_{t=2}^{T} \sup_C \|\nabla\varphi_t - \nabla\varphi_{t-1}\|^2$.
- The function variation term (Besbes et al., 2015), $F_T = \sum_{t=2}^{T} \sup_C \|\varphi_t - \varphi_{t-1}\|$.

Usually the dynamic regret upper bound contains a path length term. Zinkevich (2003) shows that mirror descent achieves an $O\big(\,(1 + P_T)\,\sqrt{T}\big)$ dynamic regret upper bound, where T is the number of games. Zhang et al. (2018) propose a method, namely adaptive learning for dynamic environment (Ader), achieves an $O\big(\sqrt{(1 + P_T)\,T}\,\big)$ dynamic regret upper bound, which is optimal in completely adversarial environment. The main idea of Ader is to run multiple Mirror Descent (MD) in parallel, each with a different step size that is optimal for a specific path length, and track the best one with Normalized Exponentiated Subgradient (NES). Zhao et al. (2020) follow the idea of Ader, and try to utilize smoothness to enhance the dynamic regret.

In this paper, we follow the idea of Ader and develop an algorithm, namely ONES-OMD with adaptive trick, which achieves an $O\big(\sqrt{(1 + P_T)\,M_T}\,\big)$ dynamic regret upper bound, where M_T is a measure of estimation accuracy. The main idea is to replace MD and NES in Ader with OMD and Optimistic Normalized Exponentiated Subgradient (ONES) respectively, and utilizes the adaptive trick. In particular, if the estimated linear loss $\widehat{x}_t^*$ in OMD is the subgradient of an estimated convex loss $\widehat{\varphi}_t$ and $\partial\widehat{\varphi}_t$ is Lipschitz continuous, then its dynamic regret upper bound has a subgradient variation type. For this situation, we develop a variant of ONES-OMD with adaptive trick, the upper bound of which has a subgradient variation type. All our algorithms are environmental adaptive.

The contributions of this article are summarized as follows.

- We develop the ONES-OMD with adaptive trick, which achieves an $O\big(\sqrt{(1 + P_T)\,M_T}\,\big)$ dynamic regret upper bound.
- We develop a variant of ONES-OMD with adaptive trick, whose dynamic regret upper bound has a subgradient variation type.
- We propose the adaptive trick, which is an extension of the doubling trick. The adaptive trick gets rid of the explicit dependence of the dynamic regret upper bound on the number of rounds T.
- ONES-OMD with adaptive trick and its variant version are all environmental adaptive.

2 Problem Formulation

We denote by $\langle\cdot\,,\cdot\rangle$ the bilinear map. Let H be a Hilbert space over $\mathbb{R}$. C is a nonempty subset of H. The bilinear map $\langle\cdot\,,\cdot\rangle$ defined on H represents its inner product. We formalize OCO problem as follows. At round t,

$$\begin{aligned}
&\text{the player chooses the strategy } x_t \in C \text{ according to some algorithm,}\\
&\text{where } C \text{ is closed and convex, and } \rho = \sup_{x,y\in C} \|x - y\| < +\infty,\ 0 \in C,\\
&\text{the adversary (environment) feeds back a convex loss function } \varphi_t\\
&\text{with } \operatorname{dom}\partial\varphi_t \supset C \text{ and } \|\partial\varphi_t\,(C)\| \leqslant \varrho < +\infty,
\end{aligned}$$

where ∂ represents the subdifferential operator. We choose the dynamic regret as the performance metric, and design adaptive algorithm to enhance its upper bound.

3 Optimistic Algorithm

3.1 Optimistic Mirror Descent

Optimistic Mirror Descent (OMD) in the form of projection is formalized as

$$\begin{aligned} \widetilde{x}_{t+1} &= P_C\left(\widetilde{x}_t - \eta x_t^*\right), \qquad x_t^* \in \partial\varphi_t\left(x_t\right), \quad \widetilde{x}_1 = x_1 \in C, \\ x_{t+1} &= P_C\left(\widetilde{x}_{t+1} - \eta \widehat{x}_{t+1}^*\right), \end{aligned} \tag{2}$$

where P_C represents the projection onto the subset C, $\eta > 0$ is the step size, $\widehat{x}_t^* \in H$ is the estimated linear loss function in round t. In Hilbert space, the projection of any point onto a closed convex subset exists and is unique (See Lemma 1 in Appendix A.1), which leads to $\widetilde{x}_t, x_t \in C$, $\forall t \in \mathbb{N}$.

Remark 1 *If φ_t is differentiable and $\widehat{x}_{t+1}^* = \nabla\varphi_t\left(\widetilde{x}_{t+1}\right)$, then OMD (Eq. (2)) becomes*

$$\begin{aligned} \widetilde{x}_{t+1} &= P_C\left(\widetilde{x}_t - \eta\nabla\varphi_t\left(x_t\right)\right), \qquad \widetilde{x}_1 = x_1 \in C, \\ x_{t+1} &= P_C\left(\widetilde{x}_{t+1} - \eta\nabla\varphi_t\left(\widetilde{x}_{t+1}\right)\right), \end{aligned} \tag{3}$$

Chiang et al. (2012) studied the static regret of Eq. (3).

The following theorem states that OMD has dynamic regret upper bound.

Theorem 1 *OMD enjoys the following dynamic regret upper bound,*

$$\underset{(z_1,z_2,\cdots,z_T)}{\operatorname{regret}} \text{OMD} \leqslant \frac{\rho^2}{2\eta} + \frac{\rho}{\eta}\sum_{t=2}^{T}\|z_t - z_{t-1}\| + \frac{\eta}{2}\sum_{t=1}^{T}\left\|x_t^* - \widehat{x}_t^*\right\|^2 - \frac{1}{2\eta}\sum_{t=1}^{T}\|x_t - \widetilde{x}_t\|^2, \tag{4}$$

where $z_t \in C$ represents the reference strategy in round t.

Set $\widehat{x}_t^*$ to be null, then OMD degenerates into Mirror Descent (MD), i.e.,

$$x_{t+1} = P_C\left(x_t - \eta x_t^*\right), \quad x_t^* \in \partial\varphi_t\left(x_t\right), \quad x_1 \in C,$$

and the corresponding dynamic regret upper bound degenerates into the following form,

$$\underset{(z_1,z_2,\cdots,z_T)}{\operatorname{regret}} \text{MD} \leqslant \frac{\rho^2}{2\eta} + \frac{\rho}{\eta}\sum_{t=2}^{T}\|z_t - z_{t-1}\| + \frac{\eta}{2}\sum_{t=1}^{T}\left\|x_t^*\right\|^2. \tag{5}$$

Remark 2 *Eq. (5) is a slight improvement of the following well-known upper bound (Zinkevich, 2003; Zhang et al., 2018).*

$$\underset{(z_1,z_2,\cdots,z_T)}{\operatorname{regret}} \text{MD} \leqslant \frac{7\rho^2}{4\eta} + \frac{\rho}{\eta}\sum_{t=2}^{T}\|z_t - z_{t-1}\| + \frac{\eta}{2}\sum_{t=1}^{T}\left\|x_t^*\right\|^2.$$

Comparing Eq. (4) and Eq. (5), we realize that by introducing the estimated linear loss function $\widehat{x}_t^*$, the dynamic regret upper bound can be tighter in the case the environment is not completely adversarial and $\widehat{x}_t^*$ is well-estimated, and meanwhile guarantees the same upper bound in the worst case.

3.2 Optimistic Normalized Exponentiated Subgradient

Optimistic Normalized Exponentiated Subgradient (ONES) is formalized as

$$\begin{aligned} \widetilde{w}_{t+1} &= \mathcal{N}\left(\widetilde{w}_t \circ \mathrm{e}^{-\theta\ell_t}\right), \qquad \widetilde{w}_1 = w_1 \in \operatorname{ri}\triangle^n, \\ w_{t+1} &= \mathcal{N}\left(\widetilde{w}_{t+1} \circ \mathrm{e}^{-\theta\widehat{\ell}_{t+1}}\right), \end{aligned} \tag{6}$$

where $\mathcal{N}$ is the normalization operator, $\circ$ is the Hadamard product symbol, $\theta > 0$ is the step size, ℓ_t is the loss vector, $\widehat{\ell}_t$ is the corresponding estimated vector, ri is the relative interior operator, and $\triangle^n := \left\{w \,\middle|\, w \in \mathbb{R}_+^{n+1}, \|w\|_1 = 1\right\}$ is the probability simplex. The normalization operator $\mathcal{N}$ guarantees that $\widetilde{w}_t, w_t \in \operatorname{ri}\triangle^n$.

ONES (Eq. (6)) is equivalent to the following iteration,

$$\begin{aligned} \widetilde{v}_{t+1} &= \widetilde{v}_t - \theta \ell_t, & \widetilde{w}_{t+1} &= \mathcal{N} e^{\widetilde{v}_{t+1}}, \quad \widetilde{v}_1 = v_1 \in \mathbb{R}^{n+1}, \\ v_{t+1} &= \widetilde{v}_{t+1} - \theta \widehat{\ell}_{t+1}, & w_{t+1} &= \mathcal{N} e^{v_{t+1}}, \end{aligned} \tag{7}$$

or the following compact version,

$$w_{t+1} = \mathcal{N}\left(w_1 \circ e^{-\theta \sum_{i=1}^{t} \ell_i - \theta \widehat{\ell}_{t+1}}\right), \quad w_1 \in \mathrm{ri}\ \triangle^n .$$

The following theorem states that ONES has static regret upper bound.

Theorem 2 *ONES enjoys the following static regret upper bound,*

$$\underset{(w,w,\cdots,w)}{\mathrm{regret}} \ \mathrm{ONES} \leqslant \frac{1}{\theta} \sum_i w(i) \ln \frac{w(i)}{w_1(i)} + \frac{\theta}{2} \sum_{t=1}^{T} \left\| \ell_t - \widehat{\ell}_t \right\|_\infty^2 - \frac{1}{2\theta} \sum_{t=1}^{T} \| w_t - \widetilde{w}_t \|_1^2 , \tag{8}$$

where $w \in \triangle^n$ represents the reference strategy.

Remark 3 *Theorem 2 is a refined version of Theorem 19 of Syrgkanis et al. (2015). The Kullback-Leibler divergence term allows the regret upper bound to be controlled by the initial value w_1 of ONES.*

Set $\widehat{\ell}_t$ to be null, then ONES degenerates into Normalized Exponentiated Subgradient (NES), i.e.,

$$w_{t+1} = \mathcal{N}\left(w_t \circ e^{-\theta \ell_t}\right), \quad w_1 \in \mathrm{ri}\ \triangle^{n-1},$$

and the corresponding static regret upper bound degenerates into the following form,

$$\underset{(w,w,\cdots,w)}{\mathrm{regret}} \ \mathrm{NES} \leqslant \frac{1}{\theta} \sum_i w(i) \ln \frac{w(i)}{w_1(i)} + \frac{\theta}{2} \sum_{t=1}^{T} \| \ell_t \|_\infty^2 . \tag{9}$$

Remark 4 *If $w_1 = \frac{1}{n+1} \mathbf{1}^{n+1}$ in Eq. (9), where $\mathbf{1}^{n+1}$ is the all-ones vector in $\mathbb{R}^{n+1}$, then*

$$\underset{(w,w,\cdots,w)}{\mathrm{regret}} \ \mathrm{NES} \leqslant \frac{\ln(n+1)}{\theta} + \frac{\theta}{2} \sum_{t=1}^{T} \| \ell_t \|_\infty^2 ,$$

which is a well-known upper bound (Shalev-Shwartz, 2012).

Comparing Eq. (8) and Eq. (9), we realize that by introducing the estimated linear loss vector $\widehat{\ell}_t$, the static regret upper bound can be tighter in the case the environment is not completely adversarial and $\widehat{\ell}_t$ is well-estimated, and meanwhile guarantees the same upper bound in the worst case.

A typical application scenario of ONES is to combine expert advices. Suppose a group of experts $\{e_i\}_{i \in I}$ provide suggestions to a player, where I is an index set. At round t, the expert e_i provides a suggestion strategy $x_t(i) \in C$, the player combines experts' suggestions with weight w_t to generate the final strategy $\overline{x}_t = \langle w_t, \boldsymbol{x}_t \rangle$, where $\boldsymbol{x}_t = \{x_t(i)\}_{i \in I}$ and w_t is generated by ONES. Then

$$\sum_{t=1}^{T} \varphi_t(\overline{x}_t) - \varphi_t(\langle w, \boldsymbol{x}_t \rangle) \leqslant \sum_{t=1}^{T} \left\langle \partial \varphi_t(\overline{x}_t), \langle w_t - w, \boldsymbol{x}_t \rangle \right\rangle = \sum_{t=1}^{T} \left\langle \langle \partial \varphi_t(\overline{x}_t), \boldsymbol{x}_t \rangle, w_t - w \right\rangle.$$

Choose $\ell_t \in \langle \partial \varphi_t(\overline{x}_t), \boldsymbol{x}_t \rangle$ as the surrogate linear loss, we have

$$\sum_{t=1}^{T} \varphi_t(\overline{x}_t) - \varphi_t(\langle w, \boldsymbol{x}_t \rangle) \leqslant \sum_{t=1}^{T} \langle \ell_t, w_t - w \rangle = \underset{(w,w,\cdots,w)}{\mathrm{regret}} \ \mathrm{ONES}.$$

4 Enhancement Method for Dynamic Regret

In this section, we follow the idea of Ader (Zhang et al., 2018) and attempt to enhance the dynamic regret by replacing MD and NES in Ader with OMD and ONES respectively.

We modify the dynamic regret upper bound for OMD (Eq. (4)) by dropping the negative term,

$$\underset{(z_1,z_2,\cdots,z_T)}{\text{regret}} \text{OMD} \leqslant \frac{\rho^2}{2\eta} + \frac{\rho}{\eta}\sum_{t=2}^{T}\|z_t - z_{t-1}\| + \frac{\eta}{2}\sum_{t=1}^{T}\left\|x_t^* - \widehat{x}_t^*\right\|^2 \leqslant \frac{\rho^2}{2\eta} + \frac{\rho}{\eta}P_T + \frac{\eta}{2}\varrho^2 S_T, \tag{10}$$

where

$$P_T = \sum_{t=2}^{T}\|z_t - z_{t-1}\|, \quad S_T = 4 + \varrho^{-2}\sum_{t=1}^{T-1}\left\|x_t^* - \widehat{x}_t^*\right\|^2, \quad \text{and} \quad \left\|\widehat{x}_t^*\right\| \leqslant \varrho.$$

After going for the game for T rounds, the value of S_T is fixed, however, the path length P_T remains unknown. This implies that the optimal parameter $\dot{\eta} = \sqrt{\rho\left(\rho + 2P_T\right)/S_T}/\varrho$ cannot be determined. A feasible method is to maintain a group of experts $\{e_i\}_{i\in E}$ (E is unknown temporarily), where the expert e_i operates OMD with a certain parameter η_i, and then composite the experts' suggestions by weight w_t to obtain the final strategy, i.e., $\overline{x}_t = \langle w_t, \boldsymbol{x}_t\rangle$, where $\boldsymbol{x}_t = \{x_t(i)\}_{i\in E}$ and $x_t(i)$ represent the suggestion of the expert e_i. This ingenious way to solve parameter difficulties comes from Zhang et al. (2018). Note that the dynamic regret can be decomposed as

$$\begin{aligned}\sum_{t=1}^{T}\varphi_t\left(\overline{x}_t\right) - \varphi_t\left(z_t\right) &= \sum_{t=1}^{T}\varphi_t\left(\langle w_t, \boldsymbol{x}_t\rangle\right) - \varphi_t\left(\left\langle 1_j, \boldsymbol{x}_t\right\rangle\right) + \sum_{t=1}^{T}\varphi_t\left(x_t(j)\right) - \varphi_t\left(z_t\right) \\ &\leqslant \sum_{t=1}^{T}\left\langle \ell_t, w_t - 1_j\right\rangle + \sum_{t=1}^{T}\varphi_t\left(x_t(j)\right) - \varphi_t\left(z_t\right),\end{aligned} \tag{11}$$

where 1_j is the one-hot vector corresponding to the expert e_j, and $\ell_t \in \langle\partial\varphi_t\left(\overline{x}_t\right), \boldsymbol{x}_t\rangle$, we use ONES to generate w_t. Since the ONES guarantees

$$\underset{(w,w,\cdots,w)}{\text{regret}} \text{ONES} \leqslant \frac{1}{\theta}\sum_i w(i)\ln\frac{w(i)}{w_1(i)} + \frac{\theta}{2}\sum_{t=1}^{T}\left\|\ell_t - \widehat{\ell}_t\right\|_\infty^2$$

by dropping the negative term in Eq. (8), we have

$$\sum_{t=1}^{T}\left\langle \ell_t, w_t - 1_j\right\rangle \leqslant \frac{1}{\theta}\sum_i 1_j(i)\ln\frac{1_j(i)}{w_1(i)} + \frac{\theta}{2}\sum_{t=1}^{T}\left\|\ell_t - \widehat{\ell}_t\right\|_\infty^2 \leqslant \frac{-\ln w_1(j)}{\theta} + \frac{\theta\rho^2\varrho^2}{2}L_T, \tag{12}$$

where

$$L_T = 4 + \rho^{-2}\varrho^{-2}\sum_{t=1}^{T-1}\left\|\ell_t - \widehat{\ell}_t\right\|_\infty^2, \quad \text{and} \quad \left\|\widehat{\ell}_t\right\| \leqslant \rho\varrho.$$

We rearrange Eq. (10) and Eq. (12) as follows,

$$\begin{aligned}\sum_{t=1}^{T}\varphi_t\left(x_t(j)\right) - \varphi_t\left(z_t\right) &\leqslant \frac{\rho\left(\rho + 2P_T\right)}{2\eta_j} + \frac{\eta_j\varrho^2}{2}S_T(j) \leqslant \frac{\rho\left(\rho + 2P_T\right)}{2\eta_j} + \frac{\eta_j\varrho^2}{2}M_T, \\ \sum_{t=1}^{T}\left\langle \ell_t, w_t - 1_j\right\rangle &\leqslant \frac{-\ln w_1(j)}{\theta} + \frac{\theta\rho^2\varrho^2}{2}L_T \leqslant \frac{-\ln w_1(j)}{\theta} + \frac{\theta\rho^2\varrho^2}{2}M_T,\end{aligned} \tag{13}$$

where

$$M_T = \max\left\{L_T, \max_j S_T(j)\right\}, \quad S_T(j) = 4 + \varrho^{-2}\sum_{t=1}^{T-1}\left\|x_t^*(j) - \widehat{x}_t^*(j)\right\|^2, \tag{14}$$

$x_t^*(j) \in \partial\varphi_t\left(x_t(j)\right)$, $x_t(j)$ is the suggestion strategy of e_j, $\widehat{x}_t^*(j)$ is the corresponding estimated linear loss function for e_j with $\left\|\widehat{x}_t^*(j)\right\| \leqslant \varrho$. We call M_T the measure of estimation accuracy. When the environment is not completely adversarial and all $\widehat{x}_t^*(j)$ and $\widehat{\ell}_t$ predict accurately, then M_T grows slowly. On the contrary, when the environment is completely adversarial, the prediction will fail and M_T grows linearly.

Now we need to allocate the group of experts. Let M_T be fixed. According to $0 \leqslant P_T \leqslant (T-1)\rho$, the optimal parameter

$$\dot{\eta} = \sqrt{\frac{\rho\left(\rho + 2P_T\right)}{\varrho^2 M_T}} \in \frac{\rho}{\varrho\sqrt{M_T}}\left[1, \sqrt{2T-1}\right].$$

Note that

$$\exists\, j \in \left\{0, 1, \cdots, \left\lfloor \log_2 \sqrt{2T-1} \right\rfloor\right\} =: E, \quad \text{such that} \quad \dot{\eta} \in \frac{\rho}{\varrho\sqrt{M_T}}\left[2^j, 2^{j+1}\right),$$

we assign the expert group as $\{e_i\}_{i\in E}$, where the expert e_i operates OMD with $\eta_i = \frac{\rho}{\varrho\sqrt{M_T}} 2^i$. Since $\eta_j \leqslant \dot{\eta} < 2\eta_j$, then for expert e_j, the following bound holds,

$$\underset{(z_1,z_2,\cdots,z_T)}{\text{regret}}\ e_j \leqslant \frac{\rho\,(\rho + 2P_T)}{2\eta_j} + \frac{\eta_j \varrho^2}{2} M_T < \frac{\rho\,(\rho+2P_T)}{\dot{\eta}} + \frac{\dot{\eta}\varrho^2}{2} M_T = \frac{3}{2}\varrho\sqrt{\rho\,(\rho+2P_T)\,M_T}, \quad (15)$$

which implies that the expert e_j reaches an almost optimal upper bound.

Substitute Eq. (13) and Eq. (15) into Eq. (11) yields

$$\sum_{t=1}^{T} \varphi_t\,(\bar{x}_t) - \varphi_t\,(z_t) < \frac{-\ln w_1\,(j)}{\theta} + \frac{\theta\rho^2\varrho^2}{2} M_T + \frac{3}{2}\varrho\sqrt{\rho\,(\rho+2P_T)\,M_T}.$$

To determine this upper bound, it suffices to choose some appropriate w_1 and θ. Let $w_1\,(i) = \beta\,(i+2)^{-\alpha}$, where $\alpha \geqslant \zeta^{-1}\,(2)$, $\beta^{-1} = \sum_{i\in E}\,(i+2)^{-\alpha}$. $\zeta^{-1}\,(2) \approx 1.728647238998183$ is the root of equation $\zeta\,(\alpha) = 2$ on $\mathbb{R}_+$, ζ represents the Riemann ζ function, i.e.,

$$\zeta\,(\alpha) = \sum_{n=1}^{\infty} \frac{1}{n^{\alpha}}, \quad \alpha > 0.$$

Note that $\beta > 1$ and $\eta_j \leqslant \dot{\eta}$, we have

$$-\ln w_1\,(j) < \alpha \ln\,(j+2) \quad \text{and} \quad j \leqslant \log_2 \sqrt{1 + \frac{2P_T}{\rho}}.$$

Thus,

$$\sum_{t=1}^{T} \varphi_t\,(\bar{x}_t) - \varphi_t\,(z_t) < \frac{\alpha}{\theta} \ln\left(2 + \log_2\sqrt{1+\frac{2P_T}{\rho}}\right) + \frac{\theta\rho^2\varrho^2}{2} M_T + \frac{3}{2}\varrho\sqrt{\rho\,(\rho+2P_T)\,M_T}. \quad (16)$$

Let $\theta \propto \frac{1}{\sqrt{M_T}}$, we have

$$\sum_{t=1}^{T} \varphi_t\,(\bar{x}_t) - \varphi_t\,(z_t) < O\left(\sqrt{(1+P_T)\,M_T}\right).$$

We call the above algorithm ONES-OMD. Comparing with $O\big(\sqrt{(1+P_T)\,T}\,\big)$, the upper bound of Ader proposed by Zhang et al. (2018), and noting that $M_T \leqslant O\,(T)$, the dynamic regret upper bound of ONES-OMD is tighter in the case the environment is not completely adversarial and $\widehat{x}_t^*$, $\widehat{\ell}_t$ are well predicted, and meanwhile guarantees the same rate in the worst case. The estimator $\widehat{x}_t^*$ and $\widehat{\ell}_t$ play the pivot role in enhancing the dynamic regret.

Note that the above analysis is based on the premise that M_T is fixed, we can utilize the following adaptive trick to unfreeze M_T, like utilizing the doubling trick to unfreeze T to anytime.

Theorem 3 (Adaptive Trick) *The adaptive trick*

$$\textit{calls ONES-OMD with } \theta \propto 2^{-m} \textit{ and } \eta_i = \frac{\rho}{\varrho} 2^{i-m}, \textit{ for } i = 0, 1, \cdots, n,$$

$$\textit{under the constraints that } M_T \in \left[4^m, 4^{m+1}\right) \textit{ and } T \in \frac{1}{2}\left[4^n, 4^{n+1}\right) + 1,$$

where m indicates the stage index of the game. The above execution process achieves an $O\big(\sqrt{(1+P_T)\,M_T}\,\big)$ *dynamic regret upper bound.*

Remark 5 *The idea of adaptive trick is to divide the range of M_T into stages of exponentially increasing size and runs ONES-OMD on each stage. This inspiration comes from the doubling trick, which divides T into stages of doubling size and runs some appropriate algorithm on each stage. Shifting from monitoring T to monitoring M_T is a crucial step in achieving environmental adaptation.*

To be understood easy, we illustrate the specific execution process for *ONES-OMD with adaptive trick* in Algorithm 1.

Algorithm 1 ONES-OMD with adaptive trick

$m \leftarrow -1,\ n \leftarrow -1$
for round $t = 1, 2, \cdots$ **do**
  $n \leftarrow \left\lfloor \log_2 \sqrt{2t-1} \right\rfloor$, $m \leftarrow \lfloor \log_4 M_t \rfloor$, where M_t is calculated according to Eq. (14)
  if m changed or n changed **then**
    Construct a set of experts $\{e_i\}_{i=0}^{n}$ and invoke Algorithm 3 with $\eta_i = \frac{\rho}{\varrho} 2^{i-m}$ for e_i
    Call Algorithm 2 with parameter n and $\theta \propto 2^{-m}$
  end if
  Receive the estimated loss vector $\widehat{\ell}_t$ from an arbitrary estimating process and send it to Algorithm 2, receive a group of estimated linear losses $\left\{\widehat{x}_t^*(0), \widehat{x}_t^*(1), \cdots, \widehat{x}_t^*(n)\right\}$ from an arbitrary estimating process and send them to each expert
  Get expert advice strategies $\boldsymbol{x}_t = \{x_t(0), x_t(1), \cdots, x_t(n)\}$, call Algorithm 2 to get the weight w_t
  Output strategy $\overline{x}_t = \langle w_t, \boldsymbol{x}_t \rangle$, and then observe loss function φ_t
  Send $\ell_t \in \langle \partial\varphi_t(\overline{x}_t), \boldsymbol{x}_t \rangle$ to Algorithm 2, send $\partial\varphi_t$ to each expert
end for

Algorithm 2 Subprogram: ONES with parameter n and θ

Require: ℓ_τ and $\widehat{\ell}_{\tau+1}$ from Algorithm 1
Output $w_1(i) = \beta (i+2)^{-\alpha}$, $i = 0, 1, \cdots, n$ at the first call, and each subsequent call follows the ONES (Eq. (6))

Algorithm 3 Subprogram: OMD with parameter η

Require: $\partial\varphi_\tau$ and $\widehat{x}_{\tau+1}^*$ from Algorithm 1
Output $x_1 \in C$ at the first call, and each subsequent call follows the ONES (Eq. (2))

5 Enhancement Method for Dynamic Regret in Subgradient Variation Type

In this section, we follow the idea of Section 4 to study the enhancement method for dynamic regret in subgradient variation type.

The following corollary states that, under the assumptions that $\widehat{x}_t^*$ is the subgradient of an estimated convex loss $\widehat{\varphi}_t$, and $\partial\widehat{\varphi}_t$ is Lipschitz continuous, the dynamic regret upper bound of OMD has subgradient variation type.

Corollary 1 *If $\widehat{x}_t^* \in \partial\widehat{\varphi}_t(\widetilde{x}_t)$ and $\partial\widehat{\varphi}_t$ is Lipschitz continuous, i.e.,*

$$\exists L > 0, \quad \text{such that} \quad \|\partial\widehat{\varphi}_t(x) - \partial\widehat{\varphi}_t(y)\| \leqslant L\|x-y\|, \quad \forall x, y \in C,$$

where $\widehat{\varphi}_t$ represents the estimated convex loss function, then

$$\underset{(z_1, z_2, \cdots, z_T)}{\text{regret}} \text{OMD} \leqslant \frac{\rho(\rho + 2P_T)}{2\eta} + \eta \sum_{t=1}^{T} \left(\sup_{x \in C} \left\| x^{\varphi_t} - x^{\widehat{\varphi}_t} \right\|^2 + 1_{\eta > \frac{1}{\sqrt{2}L}} L^2 \|x_t - \widetilde{x}_t\|^2 \right),$$

where $x^{\varphi_t} \in \partial\varphi_t(x)$, $x^{\widehat{\varphi}_t} \in \partial\widehat{\varphi}_t(x)$, and $1_{\eta > \frac{1}{\sqrt{2}L}}$ is the zero-one indicator function w.r.t. $1_{\eta > \frac{1}{\sqrt{2}L}} = 1$ iff $\eta > \frac{1}{\sqrt{2}L}$.

Remark 6 *Corollary 1 is inspired by Zhao et al. (2020). However, we have not restricted $\partial\varphi_t$ to be Lipschitz continuous.*

Similar to Section 4, we also maintain a group of experts, and each expert operates OMD with a specific parameter. Denote by $\boldsymbol{x}_t$ the vector of expert advice strategies and $\widetilde{\boldsymbol{x}}_t$ the vector of all $\widetilde{x}_t$s.

Let

$$V_T = 4 + \varrho^{-2} \sum_{t=1}^{T-1} \sup_{x \in C} \left\| x^{\varphi_t} - x^{\widehat{\varphi}_t} \right\|^2, \quad D_T = L^2 \varrho^{-2} \left(\rho^2 + \sum_{t=1}^{T-1} \max \| \boldsymbol{x}_t - \widetilde{\boldsymbol{x}}_t \|^2 \right). \tag{17}$$

The expert who operates OMD with the parameter η yields the global dynamic regret upper bound as follows,

$$\frac{\rho \left(\rho + 2P_T \right)}{2\eta} + \eta \varrho^2 \left(V_T + D_T \right),$$

and correspondingly, yields the local dynamic regret upper bound as follows,

$$\frac{\rho \left(\rho + 2P_T \right)}{2\eta} + \eta \varrho^2 V_T, \quad \eta \leqslant \frac{1}{\sqrt{2}L}.$$

In order to be compatible to V_T, we choose $\widehat{\ell}_t \in \left\langle \partial \widehat{\varphi}_t \left(\widetilde{\overline{x}}_t \right), \boldsymbol{x}_t \right\rangle$ in ONES, where $\widetilde{\overline{x}}_t = \langle \widetilde{w}_t, \boldsymbol{x}_t \rangle$. Note that $\ell_t \in \langle \partial \varphi_t \left(\overline{x}_t \right), \boldsymbol{x}_t \rangle$, we have

$$\left\| \ell_t - \widehat{\ell}_t \right\|_\infty^2 = \left\| \left\langle \overline{x}_t^{\varphi_t} - \widetilde{\overline{x}}_t^{\widehat{\varphi}_t}, \boldsymbol{x}_t \right\rangle \right\|_\infty^2 \leqslant \rho^2 \left\| \overline{x}_t^{\varphi_t} - \widetilde{\overline{x}}_t^{\widehat{\varphi}_t} \right\|^2 \leqslant \rho^2 \left(\left\| \overline{x}_t^{\varphi_t} - \overline{x}_t^{\widehat{\varphi}_t} \right\| + \left\| \overline{x}_t^{\widehat{\varphi}_t} - \widetilde{\overline{x}}_t^{\widehat{\varphi}_t} \right\| \right)^2,$$

where

$$\left\| \overline{x}_t^{\widehat{\varphi}_t} - \widetilde{\overline{x}}_t^{\widehat{\varphi}_t} \right\| \leqslant L \left\| \overline{x}_t - \widetilde{\overline{x}}_t \right\| = L \left\| \langle w_t - \widetilde{w}_t, \boldsymbol{x}_t \rangle \right\| \leqslant \rho L \left\| w_t - \widetilde{w}_t \right\|_1.$$

According to Theorem 2, the static regret upper bound for ONES is

$$\begin{aligned} & \frac{1}{\theta} \sum_i w(i) \ln \frac{w(i)}{w_1(i)} + \theta \rho^2 \sum_{t=1}^{T} \sup_{x \in C} \left\| x^{\varphi_t} - x^{\widehat{\varphi}_t} \right\|^2 + \left(\theta \rho^4 L^2 - \frac{1}{2\theta} \right) \sum_{t=1}^{T} \| w_t - \widetilde{w}_t \|_1^2 \\ \leqslant & \frac{1}{\theta} \sum_i w(i) \ln \frac{w(i)}{w_1(i)} + \theta \rho^2 \varrho^2 V_T, \quad \text{if } \theta \leqslant \frac{1}{\sqrt{2}\rho^2 L}. \end{aligned}$$

If we choose $V_T + D_T$ as the measure of estimation accuracy, then the global dynamic regret upper bound is $O\left(\sqrt{(1 + P_T)(V_T + D_T)}\right)$, and the group of experts is $\{e_\lambda\}_{\lambda \in \mathcal{E}}$, where

$$\mathcal{E} = \left\{ 0, 1, \cdots, \left\lfloor \log_2 \sqrt{2T - 1} \right\rfloor \right\},$$

the expert e_λ operates OMD with $\eta_{e_\lambda} = \frac{\rho}{\varrho \sqrt{V_T + D_T}} 2^\lambda$. If we choose V_T as the measure of estimation accuracy, then the local dynamic regret upper bound is $O\left(\sqrt{(1 + P_T) V_T}\right)$, and the group of experts is $\left\{ \epsilon_\mu \right\}_{\mu \in \mathscr{E}}$, where

$$\mathscr{E} = \left\{ \mu \in \left\{ 0, 1, \cdots, \left\lfloor \log_2 \sqrt{2T - 1} \right\rfloor \right\} \middle| \frac{\rho}{\varrho \sqrt{V_T}} 2^\mu \leqslant \frac{1}{\sqrt{2}L} \right\},$$

the expert ϵ_μ operates OMD with $\eta_{\epsilon_\mu} = \frac{\rho}{\varrho \sqrt{V_T}} 2^\mu$.

We merge two expert groups and utilize ONES to track the best expert. In this case, the initial value of ONES is

$$\begin{aligned} w_1(e_\lambda) &= \beta (\lambda + 2)^{-\alpha}, \quad \lambda \in \mathcal{E}, \\ w_1(\epsilon_\mu) &= \beta (\mu + 2)^{-\alpha}, \quad \mu \in \mathscr{E}, \end{aligned}$$

where $\alpha \geqslant \zeta^{-1}(1.5)$, $\beta^{-1} = \sum_{\lambda \in \mathcal{E}} (\lambda + 2)^{-\alpha} + \sum_{\mu \in \mathscr{E}} (\mu + 2)^{-\alpha}$. $\zeta^{-1}(1.5) \approx 2.185285451787483$ is the root of equation $\zeta(\alpha) = 1.5$ on $\mathbb{R}_+$, ζ represents the Riemann ζ function. Let $\theta \propto \frac{1}{\sqrt{V_T}}$, $\theta \leqslant \frac{1}{\sqrt{2}\rho^2 L}$, we have

$$\sum_{t=1}^{T} \varphi_t(\overline{x}_t) - \varphi_t(z_t) \leqslant \begin{cases} O\left(\sqrt{(1 + P_T)(V_T + D_T)}\right), & \\ O\left(\sqrt{(1 + P_T) V_T}\right), & \dot{\eta} \leqslant \frac{1}{\sqrt{2}L}, \end{cases}$$

where $\dot{\eta} = \sqrt{\rho(\rho + 2P_T)/(2V_T)}/\varrho$. We recombine this dynamic regret upper bound as

$$O\left(\sqrt{(1 + P_T)\left(V_T + 1_{L^2\rho(\rho + 2P_T) \leqslant \varrho^2 V_T} D_T\right)}\right).$$

To make it easier to follow, we depict the above specific execution process in Algorithm 4.

Algorithm 4 Subgradient variation version of ONES-OMD with adaptive trick

$m \leftarrow -1,\ m' \leftarrow -1,\ n \leftarrow -1,\ n' \leftarrow -1$
for round $t = 1, 2, \cdots$ **do**
  $n \leftarrow |\mathcal{E}| - 1,\ n' \leftarrow |\mathscr{E}| - 1,\ m \leftarrow \lfloor \log_4 (V_t + D_t) \rfloor,\ m' \leftarrow \lfloor \log_4 V_t \rfloor$, where V_t and D_t are calculated by Eq. (17)
  if (m or m' or n or n') changed **then**
    Construct a set of experts $\{e_\lambda\}_{\lambda \in \mathcal{E}} \cup \{\epsilon_\mu\}_{\mu \in \mathscr{E}}$ and invoke Algorithm 6 with $\eta_{e_\lambda} = \frac{\rho}{\varrho} 2^{\lambda - m}$ for e_λ, invoke Algorithm 6 with $\eta_{\epsilon_\mu} = \frac{\rho}{\varrho} 2^{\mu - m'}$ for ϵ_μ if $\mathscr{E} \neq \varnothing$
    Call Algorithm 5 with parameter n, n' and $\theta \propto 2^{-m'}$, where $\theta \leqslant \frac{1}{\sqrt{2}\rho^2 L}$
  end if
  Receive the estimated convex loss $\widehat{\varphi}_t$ from an arbitrary estimating process with $\partial \widehat{\varphi}_t$ to be Lipschitz continuous, send $\partial \widehat{\varphi}_t$ to Algorithm 5 and each expert
  Call Algorithm 5 to get expert advice strategies $\boldsymbol{x}_t, \widetilde{\boldsymbol{x}}_t$, and the weight w_t
  Output strategy $\overline{x}_t = \langle w_t, \boldsymbol{x}_t \rangle$, and then observe loss function φ_t
  Send $\ell_t \in \langle \partial \varphi_t (\overline{x}_t), \boldsymbol{x}_t \rangle$ to Algorithm 5, send $\partial \varphi_t$ to each expert
end for

Algorithm 5 Subprogram: ONES with parameter n, n' and θ

Require: ℓ_τ and $\partial \widehat{\varphi}_{\tau+1}$ from Algorithm 4
Get expert advice strategies $\boldsymbol{x}_\tau$ and $\widetilde{\boldsymbol{x}}_\tau$, send them to Algorithm 4
Output $w_1 (e_\lambda) = \beta (\lambda + 2)^{-\alpha}$, $\lambda = 0, 1, \cdots, n$, $w_1 (\epsilon_\mu) = \beta (\mu + 2)^{-\alpha}$, $\mu = 0, 1, \cdots, n'$ at the first call, and each subsequent call follows the following rule

$$\begin{aligned} \widetilde{w}_{\tau+1} &= \mathcal{N}\left(\widetilde{w}_\tau \circ \mathrm{e}^{-\theta \ell_\tau}\right), \qquad \widetilde{w}_1 = w_1, \\ \widehat{\ell}_{\tau+1} &\in \langle \partial \widehat{\varphi}_{\tau+1} (\langle \widetilde{w}_{\tau+1}, \boldsymbol{x}_\tau \rangle), \boldsymbol{x}_\tau \rangle, \\ w_{\tau+1} &= \mathcal{N}\left(\widetilde{w}_{\tau+1} \circ \mathrm{e}^{-\theta \widehat{\ell}_{\tau+1}}\right) \end{aligned}$$

Algorithm 6 Subprogram: OMD with parameter η

Require: $\partial \varphi_\tau$ and $\partial \widehat{\varphi}_{\tau+1}$ from Algorithm 4
Output $\widetilde{x}_1 = x_1 \in C$ at the first call, and each subsequent call follows the ONES (Eq. (2))

6 Conclusions and Future Work

In this paper, we study the enhancement method for dynamic regret in a non-adversarial environment under the premise of guaranteeing the worst-case dynamic regret in OCO problem. We develop an algorithm, named as ONES-OMD with adaptive trick. Theoretical analysis shows that our algorithm achieves an $O\left(\sqrt{(1 + P_T) M_T}\right)$ dynamic regret upper bound. We also develop a variant of ONES-OMD with adaptive trick that makes the dynamic regret upper bound have a subgradient variation type.

Tracking the best expert may be the general approach for online learning with dynamic regret. Optimism combined with adaptive trick provides a way to achieve environmental adaptation. We hope this work encourages further research on smoothed online learning and online learning with delayed feedback.

Acknowledgments

Paper under double-blind review.

REFERENCES

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, 2015. doi: 10.1287/opre.2015.1408.

Haïm Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer-Verlag New York, 2011. ISBN 9780387709130. doi: 10.1007/978-0-387-70914-7.

Nicolò Campolongo and Francesco Orabona. A closer look at temporal variability in dynamic online learning. *arXiv e-prints*, art. arXiv:2102.07666, February 2021.

Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006. doi: 10.1017/CBO9780511546921.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In Shie Mannor, Nathan Srebro, and Robert C. Williamson (eds.), *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pp. 6.1–6.20, Edinburgh, Scotland, 25–27 Jun 2012. JMLR Workshop and Conference Proceedings. URL `https://proceedings.mlr.press/v23/chiang12.html`.

Eric Hall and Rebecca Willett. Dynamical models and tracking regret in online convex programming. In Sanjoy Dasgupta and David McAllester (eds.), *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 579–587, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR. URL `https://proceedings.mlr.press/v28/hall13.html`.

Elad Hazan. Introduction to Online Convex Optimization. *arXiv e-prints*, art. arXiv:1909.05207, September 2019.

Ali Jadbabaie, Alexander Rakhlin, Shahin Shahrampour, and Karthik Sridharan. Online Optimization : Competing with Dynamic Comparators. In Guy Lebanon and S. V. N. Vishwanathan (eds.), *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*, pp. 398–406, San Diego, California, USA, 09–12 May 2015. PMLR. URL `https://proceedings.mlr.press/v38/jadbabaie15.html`.

Deepak S. Kalhan, Amrit Singh Bedi, Alec Koppel, Ketan Rajawat, Hamed Hassani, Abhishek K. Gupta, and Adrish Banerjee. Dynamic online learning via frank-wolfe algorithm. *IEEE Transactions on Signal Processing*, 69:932–947, 2021. doi: 10.1109/TSP.2021.3051871.

Aryan Mokhtari, Shahin Shahrampour, Ali Jadbabaie, and Alejandro Ribeiro. Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 7195–7201, 2016. ISBN 978-1-5090-1838-3. doi: 10.1109/CDC.2016.7799379.

Francesco Orabona. A modern introduction to online learning. *arXiv e-prints*, art. arXiv:1912.13213, December 2019.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In Shai Shalev-Shwartz and Ingo Steinwart (eds.), *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pp. 993–1019, Princeton, NJ, USA, 12–14 Jun 2013. PMLR. URL `https://proceedings.mlr.press/v30/Rakhlin13.html`.

Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012. ISSN 1935-8237. doi: 10.1561/2200000018.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL `https://proceedings.neurips.cc/paper/2015/file/7fea637fd6d02b8f0adf6f7dc36aed93-Paper.pdf`.

Lijun Zhang, Shiyin Lu, and Zhi-Hua Zhou. Adaptive online learning in dynamic environments. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31, pp. 1323–1333. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/file/10a5ab2db37feedfdeaab192ead4ac0e-Paper.pdf.

Peng Zhao, Yu-Jie Zhang, Lijun Zhang, and Zhi-Hua Zhou. Dynamic regret of convex and smooth functions. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 12510–12520. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/939314105ce8701e67489642ef4d49e8-Paper.pdf.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the Twentieth International Conference on Machine Learning*, ICML'03, pp. 928–935. AAAI Press, 2003. ISBN 1577351894.

A Appendix

A.1 Proof of Theorem 1

The proof of Theorem 1 relies on the following lemma. Part of the proof is inspired by Zhao et al. (2020).

Lemma 1 (Theorem 5.2 of Brezis (2011)) *Let H be a Hilbert space, and let $C \subset H$ be a nonempty closed convex set. Then $\forall x \in H$, $\exists! x_0 = P_C(x)$, such that $\langle C - x_0, x - x_0 \rangle \leqslant 0$.*

We rearrange OMD as follows,

$$\widetilde{y}_{t+1} = \widetilde{x}_t - \eta x_t^*, \qquad \widetilde{x}_{t+1} = P_C(\widetilde{y}_{t+1}),$$
$$y_{t+1} = \widetilde{x}_{t+1} - \eta \widehat{x}_{t+1}^*, \quad x_{t+1} = P_C(y_{t+1}).$$

Note that

$$\varphi_t(x_t) - \varphi_t(z_t) \leqslant \frac{1}{\eta}\left\langle \eta x_t^*, x_t - z_t \right\rangle, \quad x_t^* \in \partial \varphi_t(x_t),$$

and

$$\begin{aligned}\left\langle \eta x_t^*, x_t - z_t \right\rangle &= \eta \left\langle x_t^* - \widehat{x}_t^*, x_t - \widetilde{x}_{t+1} \right\rangle + \left\langle \eta x_t^*, \widetilde{x}_{t+1} - z_t \right\rangle + \left\langle \eta \widehat{x}_t^*, x_t - \widetilde{x}_{t+1} \right\rangle \\ &= \eta \left\langle x_t^* - \widehat{x}_t^*, x_t - \widetilde{x}_{t+1} \right\rangle - \langle \widetilde{x}_t - \widetilde{y}_{t+1}, z_t - \widetilde{x}_{t+1} \rangle - \langle \widetilde{x}_t - y_t, \widetilde{x}_{t+1} - x_t \rangle,\end{aligned}$$

where

$$\eta \left\langle x_t^* - \widehat{x}_t^*, x_t - \widetilde{x}_{t+1} \right\rangle \leqslant \eta \left\| x_t^* - \widehat{x}_t^* \right\| \| x_t - \widetilde{x}_{t+1} \| \leqslant \frac{\eta^2}{2} \left\| x_t^* - \widehat{x}_t^* \right\|^2 + \frac{1}{2} \| x_t - \widetilde{x}_{t+1} \|^2,$$

and

$$\frac{1}{2}\| x_t - \widetilde{x}_{t+1} \|^2 \leqslant \frac{1}{2}\|\widetilde{x}_{t+1}\|^2 - \frac{1}{2}\|x_t\|^2 + \langle y_t, x_t - \widetilde{x}_{t+1} \rangle,$$

since $\langle \widetilde{x}_{t+1} - x_t, y_t - x_t \rangle \leqslant 0$ according to Lemma 1. Thus

$$\begin{aligned}\left\langle \eta x_t^*, x_t - z_t \right\rangle &\leqslant \frac{\eta^2}{2}\left\| x_t^* - \widehat{x}_t^* \right\|^2 + \frac{1}{2}\|\widetilde{x}_{t+1}\|^2 - \frac{1}{2}\|x_t\|^2 - \langle \widetilde{x}_t - \widetilde{y}_{t+1}, z_t - \widetilde{x}_{t+1} \rangle - \langle \widetilde{x}_t, \widetilde{x}_{t+1} - x_t \rangle \\ &\leqslant \frac{\eta^2}{2}\left\| x_t^* - \widehat{x}_t^* \right\|^2 + \frac{1}{2}\|z_t - \widetilde{x}_t\|^2 - \frac{1}{2}\|z_t - \widetilde{x}_{t+1}\|^2 - \frac{1}{2}\|x_t - \widetilde{x}_t\|^2,\end{aligned}$$

since $\langle z_t - \widetilde{x}_{t+1}, \widetilde{y}_{t+1} - \widetilde{x}_{t+1} \rangle \leqslant 0$ according to Lemma 1. So we have

$$\begin{aligned}&\sum_{t=1}^T \varphi_t(x_t) - \varphi_t(z_t) \leqslant \frac{1}{2\eta}\sum_{t=1}^T \left(\|z_t - \widetilde{x}_t\|^2 - \|z_t - \widetilde{x}_{t+1}\|^2 \right) + \frac{\eta}{2}\sum_{t=1}^T \left\| x_t^* - \widehat{x}_t^* \right\|^2 - \frac{1}{2\eta}\sum_{t=1}^T \|x_t - \widetilde{x}_t\|^2 \\ &\leqslant \frac{1}{2\eta}\|z_1 - \widetilde{x}_1\|^2 + \frac{1}{\eta}\sum_{t=2}^T \left\| \frac{z_t + z_{t-1}}{2} - \widetilde{x}_t \right\| \|z_t - z_{t-1}\| + \frac{\eta}{2}\sum_{t=1}^T \left\| x_t^* - \widehat{x}_t^* \right\|^2 - \frac{1}{2\eta}\sum_{t=1}^T \|x_t - \widetilde{x}_t\|^2 \\ &\leqslant \frac{\rho^2}{2\eta} + \frac{\rho}{\eta}\sum_{t=2}^T \|z_t - z_{t-1}\| + \frac{\eta}{2}\sum_{t=1}^T \left\| x_t^* - \widehat{x}_t^* \right\|^2 - \frac{1}{2\eta}\sum_{t=1}^T \|x_t - \widetilde{x}_t\|^2.\end{aligned}$$

A.2 Proof of Theorem 2

The proof of Theorem 2 relies on the following lemma.

Lemma 2 (Example 2.5 of Shalev-Shwartz (2012)) *$\sum_i w(i) \ln w(i)$ is 1-strongly-convex w.r.t $\|\cdot\|_1$ over the probability simplex.*

We rearrange ONES (Eq. (7)) as follows,

$$\widetilde{v}_{t+1} = \widetilde{v}_t - \theta \ell_t, \qquad \widetilde{w}_{t+1} = \widetilde{N}_{t+1} e^{\widetilde{v}_{t+1}}, \quad \widetilde{v}_1 = v_1 \in \mathbb{R}^{n+1},$$
$$v_{t+1} = \widetilde{v}_{t+1} - \theta \widehat{\ell}_{t+1}, \quad w_{t+1} = N_{t+1} e^{v_{t+1}},$$

where $\widetilde{N}_{t+1}$ and N_{t+1} represents the normalization coefficients. Note that

$$
\begin{aligned}
\langle \theta \ell_t, w_t - w \rangle &= \theta \left\langle \ell_t - \widehat{\ell}_t, w_t - \widetilde{w}_{t+1} \right\rangle + \langle \theta \ell_t, \widetilde{w}_{t+1} - w \rangle + \left\langle \theta \widehat{\ell}_t, w_t - \widetilde{w}_{t+1} \right\rangle \\
&= \theta \left\langle \ell_t - \widehat{\ell}_t, w_t - \widetilde{w}_{t+1} \right\rangle + \langle \widetilde{v}_t - \widetilde{v}_{t+1}, \widetilde{w}_{t+1} - w \rangle + \langle \widetilde{v}_t - v_t, w_t - \widetilde{w}_{t+1} \rangle ,
\end{aligned}
$$

where

$$
\theta \left\langle \ell_t - \widehat{\ell}_t, w_t - \widetilde{w}_{t+1} \right\rangle \leqslant \theta \left\| \ell_t - \widehat{\ell}_t \right\|_\infty \| w_t - \widetilde{w}_{t+1} \|_1 \leqslant \frac{\theta^2}{2} \left\| \ell_t - \widehat{\ell}_t \right\|_\infty^2 + \frac{1}{2} \| w_t - \widetilde{w}_{t+1} \|_1^2 ,
$$

and

$$
\frac{1}{2} \| w_t - \widetilde{w}_{t+1} \|_1^2 \leqslant \left\langle \widetilde{w}_{t+1}, \ln \frac{\widetilde{w}_{t+1}}{w_t} \right\rangle ,
$$

since $\langle w, \ln w \rangle$ is 1-strongly-convex w.r.t $\|\cdot\|_1$ over the probability simplex according to Lemma 2. Note that

$$
\begin{aligned}
&\langle \widetilde{v}_t - \widetilde{v}_{t+1}, \widetilde{w}_{t+1} - w \rangle + \langle \widetilde{v}_t - v_t, w_t - \widetilde{w}_{t+1} \rangle \\
&\quad = \langle \widetilde{v}_t, w_t - w \rangle - \langle \widetilde{v}_{t+1}, \widetilde{w}_{t+1} - w \rangle - \langle v_t, w_t - \widetilde{w}_{t+1} \rangle \\
&\quad = \left\langle \ln \frac{\widetilde{w}_t}{\widetilde{N}_t}, w_t - w \right\rangle - \left\langle \ln \frac{\widetilde{w}_{t+1}}{\widetilde{N}_{t+1}}, \widetilde{w}_{t+1} - w \right\rangle - \left\langle \ln \frac{w_t}{N_t}, w_t - \widetilde{w}_{t+1} \right\rangle \\
&\quad = \langle \ln \widetilde{w}_t, w_t - w \rangle - \langle \ln \widetilde{w}_{t+1}, \widetilde{w}_{t+1} - w \rangle - \langle \ln w_t, w_t - \widetilde{w}_{t+1} \rangle \\
&\quad = \left\langle w, \ln \frac{\widetilde{w}_{t+1}}{\widetilde{w}_t} \right\rangle - \left\langle w_t, \ln \frac{w_t}{\widetilde{w}_t} \right\rangle - \left\langle \widetilde{w}_{t+1}, \ln \frac{\widetilde{w}_{t+1}}{w_t} \right\rangle ,
\end{aligned}
$$

then

$$
\begin{aligned}
\langle \theta \ell_t, w_t - w \rangle &\leqslant \frac{\theta^2}{2} \left\| \ell_t - \widehat{\ell}_t \right\|_\infty^2 + \left\langle w, \ln \frac{\widetilde{w}_{t+1}}{\widetilde{w}_t} \right\rangle - \left\langle w_t, \ln \frac{w_t}{\widetilde{w}_t} \right\rangle \\
&\leqslant \frac{\theta^2}{2} \left\| \ell_t - \widehat{\ell}_t \right\|_\infty^2 + \left\langle w, \ln \frac{\widetilde{w}_{t+1}}{\widetilde{w}_t} \right\rangle - \frac{1}{2} \| w_t - \widetilde{w}_t \|_1^2
\end{aligned}
$$

according to Lemma 2, and thus,

$$
\begin{aligned}
\sum_{t=1}^{T} \langle \ell_t, w_t - w \rangle &\leqslant \frac{1}{\theta} \sum_{t=1}^{T} \left\langle w, \ln \frac{\widetilde{w}_{t+1}}{\widetilde{w}_t} \right\rangle + \frac{\theta}{2} \sum_{t=1}^{T} \left\| \ell_t - \widehat{\ell}_t \right\|_\infty^2 - \frac{1}{2\theta} \sum_{t=1}^{T} \| w_t - \widetilde{w}_t \|_1^2 \\
&= \frac{1}{\theta} \left\langle w, \ln \frac{w}{w_1} \right\rangle + \frac{\theta}{2} \sum_{t=1}^{T} \left\| \ell_t - \widehat{\ell}_t \right\|_\infty^2 - \frac{1}{2\theta} \sum_{t=1}^{T} \| w_t - \widetilde{w}_t \|_1^2
\end{aligned}
$$

since

$$
\begin{aligned}
\sum_{t=1}^{T} \left\langle w, \ln \frac{\widetilde{w}_{t+1}}{\widetilde{w}_t} \right\rangle &= \left\langle w, \ln \frac{\widetilde{w}_{T+1}}{\widetilde{w}_1} \right\rangle = \left\langle w, \ln \frac{w}{\widetilde{w}_1} \right\rangle - \left\langle w, \ln \frac{w}{\widetilde{w}_{T+1}} \right\rangle \\
&\leqslant \left\langle w, \ln \frac{w}{\widetilde{w}_1} \right\rangle - \frac{1}{2} \| w - \widetilde{w}_{T+1} \|_1^2 \leqslant \left\langle w, \ln \frac{w}{\widetilde{w}_1} \right\rangle = \frac{1}{\theta} \left\langle w, \ln \frac{w}{w_1} \right\rangle .
\end{aligned}
$$

A.3 Proof of Theorem 3

Suppose the game has been played for T rounds, and is in stage m. $M_T \in [4^m, 4^{m+1})$. Denote by T_s the total rounds number have been played in stage s. $T = \sum_{s=1}^{m} T_s$. According to Eq. (16), the dynamic regret upper bound of stage s is

$$
O\left(\alpha \ln \left(2 + \log_2 \sqrt{1 + \frac{2P_{T_s}}{\rho}} \right) 2^s + 2\rho^2 \varrho^2 2^s + 3\varrho \sqrt{\rho \left(\rho + 2P_{T_s} \right) 2^s} \right) \leqslant O\left(\sqrt{(1 + P_T) 2^s} \right) ,
$$

and then

$$
\sum_{s=1}^{m} O\left(\sqrt{(1 + P_T) 2^s} \right) = O\left(\sqrt{(1 + P_T) 2^m} \right) = O\left(\sqrt{(1 + P_T) M_T} \right) .
$$

A.4 Proof of Corollary 1

Let $x_t^{\varphi_t} = x_t^*, \widetilde{x}_t^{\widehat{\varphi}_t} = \widehat{x}_t^*$. Note that

$$\left\|x_t^{\varphi_t} - \widetilde{x}_t^{\widehat{\varphi}_t}\right\|^2 \leqslant \left(\left\|x_t^{\varphi_t} - x_t^{\widehat{\varphi}_t}\right\| + \left\|x_t^{\widehat{\varphi}_t} - \widetilde{x}_t^{\widehat{\varphi}_t}\right\|\right)^2 \leqslant 2\left\|x_t^{\varphi_t} - x_t^{\widehat{\varphi}_t}\right\|^2 + 2L^2 \left\|x_t - \widetilde{x}_t\right\|^2,$$

where $x_t^{\widehat{\varphi}_t} \in \partial \widehat{\varphi}_t\left(x_t\right)$. According to Theorem 1, the dynamic regret upper bound for OMD is

$$\begin{aligned}
&\frac{\rho^2}{2\eta} + \frac{\rho}{\eta} P_T + \eta \sum_{t=1}^{T} \left\|x_t^{\varphi_t} - x_t^{\widehat{\varphi}_t}\right\|^2 + \left(\eta L^2 - \frac{1}{2\eta}\right) \sum_{t=1}^{T} \left\|x_t - \widetilde{x}_t\right\|^2 \\
\leqslant\ &\frac{\rho\left(\rho + 2P_T\right)}{2\eta} + \eta \sum_{t=1}^{T} \left(\sup_{x \in C} \left\|x^{\varphi_t} - x^{\widehat{\varphi}_t}\right\|^2 + 1_{\eta > \frac{1}{\sqrt{2}L}} L^2 \left\|x_t - \widetilde{x}_t\right\|^2\right).
\end{aligned}$$