Minimax Alternating Regret for the Experts Problem and Online Convex Optimization

Mengxiao Zhang
University of Iowa
mengxiao-zhang@uiowa.edu

Abstract

In this paper, we study alternating regret in online convex optimization (OCO), motivated by the success of alternating learning dynamics in two-player games. Although previous works have shown that $o(\sqrt{T})$ alternating regret is achievable under various assumptions on the loss functions and feasible domains, the minimax regret rate has remained open even for the expert problem. In this paper, we resolve this question by showing matching lower and upper bounds for both the expert problem and general OCO.

Somewhat surprisingly, for the d-expert problem, we show that the minimax alternating regret is $\Theta(\log d)$ by providing both a lower bound and an algorithm achieving a matching upper bound. This significantly improves upon the $\mathcal{O}(T^{1/3}\log^{2/3} d)$ established by [HLLZ25]. We further extend our results to general OCO over a d-dimensional compact convex set and prove that the worst-case minimax alternating regret is $\Theta\big(d\log\big(1+\frac{T}{d}\big)\big)$, also significantly improving upon the $\mathcal{O}((d\log T)^{2/3}T^{1/3})$ upper bound and resolving the open problem posed by [CCK$^+$23, HLLZ25]. As a direct implication, our alternating dynamics achieve $\mathcal{O}(\frac{\log d}{T})$ convergence to Nash equilibrium in two-player zero-sum games, matching the state of the art, and give the first such convergence rate to coarse correlated equilibrium in two-player general-sum games.

Technically, our upper bound for the expert problem is achieved by a corrected variant of Hedge. We extend the same corrected-potential argument to continuous action sets to obtain the optimal alternating-regret rate for OCO. For the lower bounds, the construction for the expert problem repeatedly eliminates half of the candidate experts, while the OCO lower bound instance construction replaces this discrete elimination by a more involved multiscale construction on the unit disk.

1 Introduction

Alternating regret is a recently introduced performance measure for online convex optimization (OCO) [CCK$^+$23, HLLZ25], motivated by the success of alternating learning dynamics in two-player games [Tam14, WTP22]. Unlike classical regret, which evaluates each loss only at the action selected before the loss is revealed, alternating regret additionally evaluates the same loss at the learner's next action, selected after observing the loss. This criterion captures the dynamics of alternating play in two-player games: a player responds to the opponent's most recent move, while the same response also serves as its action against the opponent's subsequent update. As shown by [HLLZ25], when both players guarantee alternating regret at most V_T over T rounds, their average strategy profile forms an $\mathcal{O}(V_T/T)$-approximate equilibrium. This connection has motivated growing interest in understanding how small alternating regret can be in OCO.

[CCK$^+$23] initiated the study of alternating regret in adversarial online linear optimization (OLO) and established $o(\sqrt{T})$ regret bounds over several important domains. Specifically, they proposed two algorithms that achieve $\mathcal{O}(T^{1/3}\log^{4/3}(dT))$ alternating regret for the d-expert problem and $\mathcal{O}(\log T)$ alternating regret over the unit ℓ_2 ball, respectively. Subsequently, [HLLZ25] showed that Hedge achieves $\mathcal{O}(T^{1/3}\log^{2/3} d)$ alternating regret for the d-expert problem and extended the study to general OCO with arbitrary bounded convex losses. Together, these results demonstrate that alternating regret can have a substantially better dependence on the horizon than classical regret.

However, the bounds established in these works did not determine the minimax rates, even for the basic expert problem. This is in sharp contrast to classic regret, whose minimax rates are well understood. In particular, they left open whether a polynomial dependence on the horizon T is unavoidable. This motivated the following question posed in [CCK$^+$23, HLLZ25]:

What is the minimax alternating regret for the expert problem and for general online convex optimization?

Contributions. In this paper, we completely resolve this question by establishing matching lower and upper bounds in both settings. Somewhat surprisingly, for the d-expert problem, we show that the minimax alternating regret is $\Theta(\log d)$, independent of the horizon T. For general OCO with bounded convex losses over a d-dimensional compact convex set, we prove that the worst-case minimax alternating regret is $\Theta(d\log(1+\frac{T}{d}))$. Our main results are summarized informally below.

Theorem 1.1 (Informal)**.** *For every $d \geq 2$ and every horizon $T \geq \log_2 d$, the minimax alternating regret for the d-expert problem is $\Theta(\log d)$. For OCO with bounded convex losses, the worst-case minimax alternating regret over d-dimensional compact convex sets is $\Theta(d\log(1+\frac{T}{d}))$.*

More specifically, our contributions are as follows:

- In Section 3, we study the d-expert problem and first establish an $\Omega(\log d)$ lower bound using a halving construction in which the adversary reveals the identity of the best expert one bit at a time. Although this logarithmic lower bound may initially appear weak, we show that it is in fact tight by introducing a corrected variant of Hedge with matching $\mathcal{O}(\log d)$ alternating regret. Importantly, this guarantee is independent of T, revealing a sharp distinction from classical regret and resolving the open problem posed by [CCK$^+$23, HLLZ25]. By applying the alternating-regret-to-equilibrium reduction of [HLLZ25], our result yields $\mathcal{O}(\frac{\log d}{T})$ convergence to Nash equilibrium in two-player zero-sum games, matching the state of the art [RS13, SALS15], and the first such convergence rate to coarse correlated equilibrium in two-player general-sum games. Technically, our algorithm is based on the standard exponential-weights algorithm but importantly modifies the played distribution using carefully designed linear and quadratic correction terms. These corrections cancel the local norm terms that arise in the analysis and allow the resulting potential argument to telescope without accumulating a dependence on T.

- In Section 4, we extend our results to general OCO. We first prove an $\Omega\left(d \log\left(1+\frac{T}{d}\right)\right)$ lower bound using a multiscale geometric construction on the unit two-dimensional disk, organized over a sequence of epochs with geometrically decreasing spatial scales. In each epoch, the construction creates a constant expected loss gap while ensuring that a single final comparator remains nearly as competitive on all previous epochs. A direct-product argument then extends this two-dimensional lower bound to dimension d. To provide a matching upper bound, we extend our algorithm for the d-expert problem to a continuous action space, and we show that the same corrected-potential argument yields a matching alternating-regret bound, resolving the open problem proposed in [HLLZ25].

1.1 Related Works

Online convex optimization. Prediction with expert advice and online convex optimization are classical models for sequential decision making. Weighted Majority, Hedge, and exponential weights are foundational algorithms for the expert problem [LW94, FS97, CBL06], while OCO extends the action space from a finite set of experts to a convex domain [Zin03, Haz16]. The classical objective in this literature is standard regret, and a broad line of work develops sharper guarantees by exploiting additional structure such as curvature or predictability of the loss sequence [HK10, RS13]. Alternating regret studies a different performance measure: the loss f_t is evaluated not only at the action x_t chosen before f_t is revealed, but also at the next action x_{t+1}, which may depend on f_t. Consequently, alternating regret can be substantially smaller than standard regret even when the loss sequence is fully adversarial.

Alternating regret and alternating learning dynamics. Alternating updates have long been used in game solving, most notably in CFR+ and related methods for extensive-form games [Tam14, BMS19]. Their theoretical advantages have also been studied in normal-form and convex-concave zero-sum games [BGP20, WTP22, KWW26]. [WTP22] connected the acceleration produced by alternation to a regret quantity that was subsequently formalized as alternating regret by [CCK$^+$23]. The latter work studied alternating regret against an arbitrary adversarial sequence in online linear optimization and established $o(\sqrt{T})$ guarantees over several important domains. [HLLZ25] extended this adversarial perspective to general OCO with bounded convex losses and showed how alternating-regret guarantees translate directly into convergence guarantees for alternating learning dynamics in two-player zero-sum and general-sum convex-concave games. In contrast to analyses that rely on the particular coupled dynamics of a game, these results isolate a worst-case online-learning property of each player. Our work completes this line by identifying the minimax alternating-regret rates for both the expert problem and general bounded-loss OCO.

Fast convergence in games. The connection between regret minimization and equilibrium computation is classical: no-regret learning yields convergence of time-averaged play to equilibrium, and the usual $\mathcal{O}(\sqrt{T})$ regret guarantee leads to the canonical $\mathcal{O}(1/\sqrt{T})$ convergence rate [FS99]. A large literature obtains faster rates by exploiting the structure of game-generated loss sequences, with optimistic and predictive methods providing a particularly successful route [RS13, SALS15, CP20, DFG21, FAL$^+$22]. For two-player zero-sum games, such methods can achieve total regret independent of T and hence an $\mathcal{O}(1/T)$ average-iterate convergence rate. Alternation provides a distinct route to acceleration: under alternating play, equilibrium error is controlled by the players' average alternating regret [WTP22, HLLZ25]. Recent work further demonstrates that optimism and alternation are genuinely different mechanisms: [LPSS25] show that, without regularization, optimistic Fictitious Play attains constant regret in 2×2 zero-sum games with a unique interior Nash equilibrium, whereas Alternating Fictitious Play can suffer $\Omega(\sqrt{T})$ regret even on Matching Pennies.

Subsequent to this work, a line of work has shown constant-regret guarantees under simultaneous full-information self-play, which yield $\mathcal{O}(1/T)$ convergence to coarse correlated equilibrium in general-sum games with any fixed number of players and actions [LFO26, ALM26, SP26, Ha26], and constant swap-regret

guarantees, which further yield $\mathcal{O}(1/T)$ convergence to correlated equilibrium [Mai26]. Our results show that $\mathcal{O}(\frac{\log d}{T})$ convergence to coarse correlated equilibrium in two-player general-sum games can be obtained via alternation.

Concurrent work. In concurrent work, [TZ26] establish the same minimax alternating-regret rates for experts and bounded-loss OCO, together with the corresponding two-player equilibrium guarantees. Their alternation-aware Hedge update places a centered quadratic loss term inside the exponential weights and determines its center through a scalar optimization. Our update instead applies explicit linear and quadratic corrections to the standard exponential-weights distribution, derived from a variance correction to the log-partition potential. The expert lower bounds also differ: their construction requires a stronger adaptive adversary and uses $T \geq d-1$, whereas our halving construction is oblivious, and applies for $T \geq \log_2 d$. Both OCO lower bounds use multiscale constructions on a two-dimensional disk followed by a product construction.

2 Preliminaries

Notation. For an integer $m \geq 1$, denote by $[m]$ the set $\{1, \dots, m\}$, let $\Delta_m = \{p \in \mathbb{R}_+^m : \sum_{i=1}^m p_i = 1\}$ be the probability simplex, and let e_i denote its i-th standard basis vector. With a slight abuse of notation, given a d-dimensional compact convex set $\Omega \subseteq \mathbb{R}^d$, we also define $\Delta_\Omega \triangleq \{p : \Omega \to \mathbb{R}_+ : \int_\Omega p(x)\,dx = 1\}$ to be the set of probability densities over Ω, and denote by $\mathrm{vol}(\Omega)$ its Lebesgue volume. We use $\mathbf{0}$ for the all-zero vector in an appropriate dimension (or the zero function, when the context is clear). Inner products are denoted by $\langle x, y\rangle$, and $\|\cdot\|_2$ denotes the Euclidean norm. We denote by ∂A the boundary of a set A.

For a distribution $q \in \Delta_m$, a subset $S \subseteq [m]$, and a vector $v \in \mathbb{R}^m$, we use

$$q(S) \triangleq \sum_{i \in S} q_i, \qquad \mathbb{E}_q[v] = \mathbb{E}_{i\sim q}[v_i] \triangleq \sum_{i=1}^m q_i v_i, \qquad \mathrm{Var}_q(v) = \mathrm{Var}_{i\sim q}(v_i) \triangleq \mathbb{E}_{i\sim q}\Big[\big(v_i - \mathbb{E}_q[v]\big)^2\Big].$$

The same notation is used for expectations and variances under a probability density $q \in \Delta_\Omega$, with sums replaced by integrals. For a score vector $R \in \mathbb{R}^m$, let $q[R] \in \Delta_m$ denote the exponential-weights distribution with $q_i[R] \propto \exp(R_i)$. For a score function $R : \Omega \to \mathbb{R}$, we use the same notation $q[R] \in \Delta_\Omega$ for the density $q[R](x) \propto \exp(R(x))$. For a differentiable convex function $F : \mathbb{R}^m \to \mathbb{R}$, its Bregman divergence is $D_F(a, b) \triangleq F(a) - F(b) - \langle \nabla F(b), a - b\rangle$. We use $\widetilde{\mathcal{O}}(\cdot)$ to suppress logarithmic factors.

Problem setting. Let $\mathcal{X}$ be a d-dimensional compact convex set. The interaction between the learner and the environment lasts for T rounds. At each round $t \in [T]$, the learner chooses $x_t \in \mathcal{X}$, while simultaneously the adversary picks a convex function $f_t : \mathcal{X} \to [-1, 1]$ and reveals it to the learner. Let $(\mathcal{F}_t)_{t=0}^T$ denote the learner's filtration, where $\mathcal{F}_0$ is the σ-algebra generated by the initial information and the learner's internal random seed, and, for each $t \in [T]$, $\mathcal{F}_t \triangleq \sigma\big(\mathcal{F}_0, f_1, \dots, f_t\big)$ is the σ-algebra generated by $\mathcal{F}_0$ and the revealed loss functions $f_1, \dots, f_t$. Thus the action x_t must be $\mathcal{F}_{t-1}$-measurable and the adversary must commit to f_t independently of the learner's fresh round-t sampling randomness. After observing f_T, the learner additionally chooses x_{T+1} based on $\mathcal{F}_T$.

In this paper, we consider alternating regret as the performance measure. Specifically, the alternating regret against a comparator $u \in \mathcal{X}$ is defined as $\mathrm{Reg}_T^{\mathrm{alt}}(u) \triangleq \sum_{t=1}^T \big(f_t(x_t) + f_t(x_{t+1}) - 2f_t(u)\big)$, and we define $\mathrm{Reg}_T^{\mathrm{alt}} \triangleq \sup_{u\in\mathcal{X}} \mathrm{Reg}_T^{\mathrm{alt}}(u)$. We adopt the endpoint convention $f_0(x) = f_{T+1}(x) = 0$ for all $x \in \mathcal{X}$ and define the effective loss $h_t = f_{t-1} + f_t$ for $t = 1, \dots, T+1$. Each f_t appears once in $h_t(x_t)$, once in

$h_{t+1}(x_{t+1})$, and twice at the comparator. Consequently, we can equivalently write $\mathrm{Reg}_T^{\mathrm{alt}}(u)$ as follows:

$$\mathrm{Reg}_T^{\mathrm{alt}}(u) = \sum_{t=1}^{T+1} \big(h_t(x_t) - h_t(u)\big). \tag{1}$$

We only assume that the convex losses satisfy $f_t \in [-1, 1]$ and do not assume Lipschitzness, smoothness, or strong convexity.

Expert Problem. In this work, we also consider the expert problem, which is the linear specialization of the OCO protocol. In this case, $\mathcal{X} = \Delta_d$, and the loss function f_t is induced by a loss vector $\ell_t \in [-1, 1]^d$ defined as $f_t(p) = \langle p, \ell_t \rangle$. Following the same convention, we set $\ell_0 = \ell_{T+1} = \mathbf{0}$ and define the effective loss vector $z_t = \ell_{t-1} + \ell_t$ for all $t = 1, \ldots, T + 1$. Thus $h_t(p) = \langle p, z_t \rangle$, and the expert alternating regret with respect to comparator $u \in \Delta_d$ is defined as

$$\mathrm{Reg}_T^{\mathrm{alt,exp}}(u) \triangleq \sum_{t=1}^{T+1} \langle p_t - u, z_t \rangle \,, \tag{2}$$

and we also define $\mathrm{Reg}_T^{\mathrm{alt,exp}} \triangleq \max_{u \in \Delta_d} \mathrm{Reg}_T^{\mathrm{alt,exp}}(u)$.

3 Alternating Regret in the Expert Problem

We begin with the d-expert problem and derive tight lower and upper bounds.

3.1 Lower Bound

In this section, we show that the alternating-regret lower bound for the d-expert problem is $\Omega(\log d)$ even when the adversary is oblivious, as formally stated in the following theorem.

Theorem 3.1. *For every $T \geq \log_2 d$, $d \geq 2$, and every possibly randomized expert algorithm, there is an oblivious sequence $\ell_1, \ldots, \ell_T \in [-1, 1]^d$ for which*

$$\mathbb{E}\big[\mathrm{Reg}_T^{\mathrm{alt,exp}}\big] \geq \lfloor \log_2 d \rfloor.$$

Here, the expectation is taken with respect to the internal randomization of the learner.

Although this lower bound may seem weak because it does not depend on T, we will show in Section 3.2 that this rate is in fact tight. The hard loss sequence construction has a simple interpretation. Before the game begins, the adversary chooses a nested sequence of sets, each of which is a uniformly random half of the preceding set. Each active round reveals one more bit about the final comparator. Conditional on the losses revealed in previous rounds, the learner cannot determine which half will survive. Consequently, the learner assigns, in expectation, only half of its probability mass on the current set to the surviving half. After $\log_2 d$ cuts only one expert remains, and there are no more bits to hide.

Proof. Let $m = \lfloor \log_2 d \rfloor$, and choose an initial set $S_0 \subseteq [d]$ of 2^m experts. To prove the existence of the desired oblivious sequence, first consider the following randomized oblivious construction. Before the game begins, sample a uniformly random permutation π of the elements of S_0, and define

$$S_t = \big\{\pi(1), \ldots, \pi(2^{m-t})\big\}, \qquad t \in [m].$$

Thus, conditional on S_{t-1}, the set S_t is a uniformly random subset of S_{t-1} of cardinality $\frac{1}{2}|S_{t-1}|$. On every active round $t \le m$, define $\ell_t \in [-1,1]^d$ by

$$\ell_{t,i} = \begin{cases} -1, & i \in S_t, \\ +1, & i \notin S_t. \end{cases} \tag{3}$$

All later losses are zero, meaning that $\ell_t = \mathbf{0}$ for $t \in \{m+1, \dots, T\}$. Since the permutation π, and hence the entire loss sequence, is sampled before the learner chooses p_1, this adversary is oblivious.

Recall that $\mathcal{F}_{t-1}$ contains the history immediately after p_t has been chosen and before ℓ_t is revealed. The past losses reveal $S_1, \dots, S_{t-1}$, but they do not reveal the ordering of the elements remaining in S_{t-1}. Moreover, p_t depends only on the past losses and the learner's internal randomization. Therefore, conditional on $\mathcal{F}_{t-1}$, the set S_t remains a uniformly random half of S_{t-1}. It follows that $\mathbb{E}\big[p_t(S_t) \mid \mathcal{F}_{t-1}\big] = \frac{1}{2}p_t(S_{t-1})$.

Finally, let $i^\star = \pi(1)$ denote the unique element of S_m. Because $i^\star \in S_t$ for every $t \in [m]$, we have $\ell_{t,i^\star} = -1$ on every active round. Consequently, $\langle p_t, \ell_t\rangle - \ell_{t,i^\star} = 2\big(1 - p_t(S_t)\big)$. Taking the conditional expectation and using $p_t(S_{t-1}) \le 1$, we obtain that

$$\mathbb{E}\left[\langle p_t, \ell_t\rangle - \ell_{t,i^\star} \mid \mathcal{F}_{t-1}\right] = 2\left(1 - \frac{1}{2}p_t(S_{t-1})\right) = 2 - p_t(S_{t-1}) \ge 1.$$

The next distribution p_{t+1} is chosen after ℓ_t is known, but its loss cannot be smaller than the minimum coordinate of ℓ_t. Therefore,

$$\langle p_{t+1}, \ell_t\rangle - \ell_{t,i^\star} = 2\big(1 - p_{t+1}(S_t)\big) \ge 0.$$

Thus, each active round contributes at least one unit of alternating regret in expectation. Summing over the m active rounds gives $\mathbb{E}_{\pi,\mathrm{alg}}\big[\mathrm{Reg}_T^{\mathrm{alt,exp}}\big] \ge m$. Since this inequality holds after averaging over the finitely many permutations of S_0, there exists a fixed permutation $\bar{\pi}$ such that the corresponding deterministic loss sequence satisfies $\mathbb{E}_{\mathrm{alg}}\big[\mathrm{Reg}_T^{\mathrm{alt,exp}}\big] \ge m$. This loss sequence is fixed before the interaction begins and is therefore oblivious, which proves the claim. □

3.2 Upper Bound

We now turn to the upper bound for the expert problem. Somewhat surprisingly, the $\Omega(\log d)$ lower bound from Section 3.1 is achievable without any dependence on the horizon, revealing a sharp separation from classical regret. Our algorithm is based on exponential weights, but modifies the distribution played on each round by linear and quadratic correction terms. We first give a walkthrough of the algorithm and then explain the design for these corrections from the one-step condition needed for logarithmic alternating regret.

Algorithm walkthrough. Our algorithm maintains a score vector R_t and the corresponding exponential-weights distribution q_t at each round t. At the beginning of round t, the learner knows q_{t-1} and the previously revealed loss ℓ_{t-1}. Starting from q_{t-1}, the learner adjusts the weights using each expert's previous loss relative to the average. Denote by $\bar{\ell}_{t-1} \triangleq \mathbb{E}_{i\sim q_{t-1}}[\ell_{t-1,i}]$ the mean loss at round $t-1$, $\widetilde{\ell}_{t-1,i} \triangleq \ell_{t-1,i} - \bar{\ell}_{t-1}$ the centered loss, and $H_{t-1} \triangleq \mathbb{E}_{i\sim q_{t-1}}[\widetilde{\ell}^2_{t-1,i}]$ its variance under q_{t-1}. The learner then plays

$$p_{t,i} = q_{t-1,i} \cdot \left(1 - 4\eta\widetilde{\ell}_{t-1,i} - 2\eta^2\left(\widetilde{\ell}^2_{t-1,i} - H_{t-1}\right)\right), \tag{4}$$

where $\eta > 0$ is the learning rate. Specifically, (4) incorporates two correction terms. The linear correction increases the weights of experts with below-average losses in the previous round and decreases those with above-average losses. The quadratic correction further decreases weights when the squared deviation exceeds H_{t-1} and increases them when it is smaller. Both corrections have zero mean under q_{t-1}, so the resulting

weights sum to one. With the choice of $\eta = \frac{1}{10}$, we show in Proposition A.1 in Appendix A that the weights are also nonnegative, making p_t a valid distribution. After observing ℓ_t, the algorithm updates $R_t = R_{t-1} - (\ell_{t-1} + \ell_t)$, and lets q_t be the exponential-weights distribution induced by R_t. The terminal distribution p_{T+1} is computed in the same way from the known state (q_T, ℓ_T) after ℓ_T is revealed. The complete pseudocode is shown in Algorithm 1.

Algorithm 1: Corrected AltHedge

Input: Number of experts d and learning rate $\eta > 0$
Set $\ell_0 = R_0 = \mathbf{0} \in \mathbb{R}^d$, and let q_0 be uniform on $[d]$.
for $t = 1, \dots, T$ **do**
Compute

$$\bar{\ell}_{t-1} = \mathbb{E}_{i\sim q_{t-1}}[\ell_{t-1,i}], \qquad \widetilde{\ell}_{t-1,i} = \ell_{t-1,i} - \bar{\ell}_{t-1}, \qquad H_{t-1} = \mathbb{E}_{i\sim q_{t-1}}[\widetilde{\ell}^2_{t-1,i}].$$

Compute $p_t \in \Delta_d$ according to (4).
Play p_t and observe $\ell_t \in [-1, 1]^d$.
Set $R_t = R_{t-1} - (\ell_t + \ell_{t-1})$ and $q_t = q[\eta R_t]$.

At first sight, the two correction terms in (4) may appear specialized. We next explain in the remainder of this section the reason for the choice of the two corrections by comparing this update with an ideal distribution that has access to the current loss. For the analysis, we use the endpoint convention $\ell_{T+1} = \mathbf{0}$ and set $R_{T+1} = R_T - \ell_T$. Recall that $z_t = \ell_{t-1} + \ell_t$, so that $R_t = -\sum_{\tau=1}^{t} z_\tau$ for all $t \in [T + 1]$.

The ideal post-update iterate. To understand the design of p_t, we first consider the distribution that would be used if the current effective loss z_t were already known, as in the standard Be-the-Leader argument. Define the log-partition function and its gradient by

$$F(R) \triangleq \frac{1}{\eta} \log\left(\frac{1}{d}\sum_{i=1}^{d} e^{\eta R_i}\right), \qquad q[\eta R] \triangleq \nabla F(R). \tag{5}$$

By construction, $q_t = q[\eta R_t]$. Since $R_t = R_{t-1} - z_t$, convexity of F gives $\langle q_t, z_t\rangle \le F(R_{t-1}) - F(R_t)$. Therefore, we have for every expert $i \in [d]$,

$$\sum_{t=1}^{T+1} \langle q_t - e_i, z_t\rangle \le -F(R_{T+1}) + R_{T+1,i} \le \frac{\log d}{\eta}, \tag{6}$$

where the last inequality follows from $F(R) \ge \max_i R_i - (\log d)/\eta$. Thus, if the learner could play the post-update distribution q_t, the desired logarithmic bound would follow immediately.

However, the caveat is that q_t is not $\mathcal{F}_{t-1}$ measurable and depends on $z_t = \ell_{t-1} + \ell_t$, and hence on the unrevealed loss ℓ_t. At the beginning of round t, the latest available exponential-weights iterate is q_{t-1}. If the learner plays this iterate, then $\langle q_{t-1}, z_t\rangle = F(R_{t-1}) - F(R_t) + D_F(R_t, R_{t-1})$, and consequently

$$\sum_{t=1}^{T+1} \langle q_{t-1} - e_i, z_t\rangle \le \frac{\log d}{\eta} + \sum_{t=1}^{T+1} D_F(R_t, R_{t-1}). \tag{7}$$

The additional Bregman-divergence sum is the standard Hedge stability cost. In general, this term is of order $\eta \sum_t \mathrm{Var}_{q_{t-1}}(z_t)$ and may contribute $\mathcal{O}(\eta T)$. Therefore, the natural $\mathcal{F}_{t-1}$-measurable choice q_{t-1} recovers only the classical $\mathcal{O}(\frac{\log d}{\eta} + \eta T)$ guarantee.

Unlike in the standard expert problem, however, the learner is not completely ignorant of the effective loss z_t when selecting p_t: its component ℓ_{t-1} is already $\mathcal{F}_{t-1}$-measurable. We will then see how this property

can be utilized to design a specific *potential function* so that the $\mathcal{F}_{t-1}$-measurable distribution p_t enjoys the same telescoping bound as the ideal post-update iterate.

A corrected-potential criterion. The preceding derivation suggests the following sufficient condition. It is enough to find a correction sequence C_t satisfying $C_0 = C_{T+1} = 0$ and

$$\langle p_t, z_t \rangle \leq F(R_{t-1}) - F(R_t) + C_t - C_{t-1} \tag{8}$$

on every round t. Indeed, summing (8) makes both the log-partition terms and the corrections telescope. Using $F(R_0) = 0$, we obtain $\sum_{t=1}^{T+1} \langle p_t, z_t \rangle \leq -F(R_{T+1})$. Combining this with $F(R_{T+1}) \geq \max_i R_{T+1,i} - (\log d)/\eta$ gives the desired regret bound. Thus, (8) is the one-step inequality that we aim to establish.

More precisely, for each round t, we aim to construct a concave function $G : [0,1] \to \mathbb{R}$ satisfying $G(0) = F(R_{t-1}) - C_{t-1}$, $G(1) = F(R_t) - C_t$, and $G'(0) = -\langle p_t, z_t \rangle$, where p_t depends only on information available before ℓ_t is revealed. Concavity then gives $G(1) - G(0) \leq G'(0) = -\langle p_t, z_t \rangle$, which rearranges to (8). The use of the left-endpoint derivative is important because its coefficients can depend only on the pre-update state and are thus $\mathcal{F}_{t-1}$ measurable, whereas the right endpoint depends on the unrevealed loss ℓ_t.

To construct such a function, we consider the following family of corrected potentials, which will determine both the correction and the played distribution. For $\alpha > 0$, define

$$U_\alpha(R, \ell) \triangleq F(R) - \alpha\eta \operatorname{Var}_{q[\eta R]}(\ell), \tag{9}$$

where $R \in \mathbb{R}^d$ represents a score vector and $\ell \in \mathbb{R}^d$ represents a single-round loss vector.

Intuition for the potential function. We next explain why the potential in (9) takes this particular form. Recall that the log-partition term $F(R)$ has positive curvature, which gives rise to the positive Bregman-divergence term in (7). Concretely, for an arbitrary effective-loss direction z, define

$$R(\theta) \triangleq R_{t-1} - \theta z, \qquad \theta \in [0,1]. \tag{10}$$

Along this score path, a direct calculation shows that

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} F(R(\theta)) = \eta \operatorname{Var}_{q[\eta R(\theta)]}(z) \geq 0.$$

The reason for introducing the negative variance term in (9) is then to offset this positive curvature. To see why this correction is natural, define $\ell(\theta) \triangleq \ell_{t-1} - \theta \cdot z$. For the realized direction $z = z_t = \ell_{t-1} + \ell_t$, this path joins the known loss ℓ_{t-1} to $-\ell_t$ and stays in $[-1,1]^d$. Since variance is unchanged by negation, $U_\alpha(R_t, -\ell_t) = U_\alpha(R_t, \ell_t)$, so these endpoints give the potential difference needed in (8). For now, suppose that $q[\eta R(\theta)]$ were held fixed along the path. Then we have that

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \operatorname{Var}_q(\ell(\theta)) = 2 \operatorname{Var}_q(z).$$

Hence the correction $-\alpha\eta \operatorname{Var}_q(\ell(\theta))$ contributes a negative curvature $-2\alpha\eta \operatorname{Var}_q(z)$. For $G_\alpha(\theta) \triangleq U_\alpha(R(\theta), \ell(\theta))$, this fixed-$q$ calculation would therefore give the curvature to be $(1 - 2\alpha)\eta \operatorname{Var}_q(z)$. Thus, any $\alpha > 1/2$ would reverse the positive curvature of F.

Moreover, the derivative of the corrected potential at the known left endpoint also naturally induces a candidate distribution. By the chain rule, we have that the derivative of G_α at $\theta = 0$ is equal to

$$G'_\alpha(0) = -\left\langle \nabla_R U_\alpha(R_{t-1}, \ell_{t-1}) + \nabla_\ell U_\alpha(R_{t-1}, \ell_{t-1}), z \right\rangle.$$

Accordingly, define the coefficient vector

$$\widetilde{p}_t^{(\alpha)} \triangleq \nabla_R U_\alpha(R_{t-1}, \ell_{t-1}) + \nabla_\ell U_\alpha(R_{t-1}, \ell_{t-1}).$$

Since this vector depends only on (R_{t-1}, ℓ_{t-1}), it is $\mathcal{F}_{t-1}$-measurable. It also automatically has unit mass. Indeed, for every $c \in \mathbb{R}$, the potential satisfies $U_\alpha(R + c\mathbf{1}, \ell) = U_\alpha(R, \ell) + c$ and $U_\alpha(R, \ell + c\mathbf{1}) = U_\alpha(R, \ell)$. Differentiating these identities with respect to c gives $\langle \nabla_R U_\alpha(R, \ell), \mathbf{1} \rangle = 1$ and $\langle \nabla_\ell U_\alpha(R, \ell), \mathbf{1} \rangle = 0$, and hence $\sum_i \widetilde{p}_{t,i}^{(\alpha)} = 1$. These are the two desired properties for our choice of p_t and it remains only to choose α and η so that $\widetilde{p}_t^{(\alpha)}$ is coordinatewise nonnegative.

Concrete choices of α and η. Because $q[\eta R(\theta)]$ also varies along the path, the previous fixed-$q[\eta R(\theta)]$ curvature calculation above is only heuristic. However, with the help of the boundedness of the loss interpolation $\ell(\theta)$ and a more involved calculation, we can indeed show in Proposition A.2 in Appendix A that the complete second-order derivative of $U_\alpha(R(\theta), \ell(\theta))$ is upper bounded as follows:

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} U_\alpha(R(\theta), \ell(\theta)) \leq \left[(1 - 2\alpha)\eta + 8\alpha\eta^2 + 3\alpha\eta^3\right] \mathrm{Var}_{q[\eta R(\theta)]}(z). \tag{11}$$

Therefore, choosing $\alpha = 2$ and $\eta = 1/10$ makes the coefficient on the right-hand side strictly negative. We then fix these two values and denote U_2 by U. For notational convenience, however, we use η in the formulas below to represent the fixed value $1/10$, so that the dependence of the algorithm and the potential on the learning rate remains explicit. Thus, we define

$$U(R, \ell) \triangleq F(R) - 2\eta \,\mathrm{Var}_{q[\eta R]}(\ell). \tag{12}$$

Set $C_t \triangleq 2\eta \,\mathrm{Var}_{q_t}(\ell_t)$. Then we have that $U(R_t, \ell_t) = F(R_t) - C_t$. We also have $C_0 = C_{T+1} = 0$ as $\ell_0 = \ell_{T+1} = \mathbf{0}$. Moreover, (11) shows that $G_2(\theta)$ is concave on $[0, 1]$.

Now we compute $G_2'(0)$ and verify that this indeed induces a valid distribution. Based on this choice of potential, direct differentiation gives that

$$G_2'(0) = -\sum_{i=1}^{d} q_{t-1,i} \left[1 - 4\eta\widetilde{\ell}_{t-1,i} - 2\eta^2 \left(\widetilde{\ell}_{t-1,i}^2 - H_{t-1}\right)\right] z_i.$$

The coefficient vector in $-G_2'(0)$ is exactly the distribution p_t in (4). As shown above, its coordinates sum to one, while Proposition A.1 establishes their nonnegativity at $\eta = 1/10$. Hence $p_t \in \Delta_d$.

The following lemma summarizes the two properties needed in the corrected-potential argument.

Lemma 3.2. *Fix a round t, a score vector $R_{t-1} \in \mathbb{R}^d$, and a loss vector $\ell_{t-1} \in [-1, 1]^d$. For any $z \in \mathbb{R}^d$ satisfying $\ell_{t-1} - z \in [-1, 1]^d$, let $R(\theta)$ be as in* (10) *and let $\ell(\theta)$ be the loss interpolation introduced above. If $\eta = 1/10$, then:*

1. *G_2 is concave on $[0, 1]$;*
2. *$G_2'(0) = -\langle p_t, z \rangle$, where p_t is the $\mathcal{F}_{t-1}$-measurable distribution in* (4)*.*

The complete proof of Lemma 3.2 is deferred to Appendix A. Applying Lemma 3.2 to the realized direction $z = z_t$ and using $U(R_t, -\ell_t) = U(R_t, \ell_t)$, we know that

$$\langle p_t, z_t \rangle \leq U(R_{t-1}, \ell_{t-1}) - U(R_t, \ell_t) = F(R_{t-1}) - F(R_t) + C_t - C_{t-1}, \tag{13}$$

which is exactly the target inequality (8) that we want.

Regret guarantee. Having derived the correction and the played distribution, we now state the resulting alternating-regret guarantee.

Theorem 3.3. *For every adaptive sequence $\ell_1, \dots, \ell_T \in [-1,1]^d$, Algorithm 1 with $\eta = 1/10$ satisfies $\mathrm{Reg}_T^{\mathrm{alt,exp}} \le 10 \log d$.*

Proof. Summing (13) over $t = 1, \dots, T+1$ yields $\sum_{t=1}^{T+1} \langle p_t, z_t \rangle \le U(R_0, \ell_0) - U(R_{T+1}, \ell_{T+1})$. Since $R_0 = \ell_0 = \mathbf{0}$, we have $U(R_0, \ell_0) = 0$. Moreover, $\ell_{T+1} = \mathbf{0}$, so $U(R_{T+1}, \ell_{T+1}) = F(R_{T+1})$. Therefore, $\sum_{t=1}^{T+1} \langle p_t, z_t \rangle \le -F(R_{T+1})$. Using the effective-loss representation of alternating regret,

$$\mathrm{Reg}_T^{\mathrm{alt,exp}} = \sum_{t=1}^{T+1} \langle p_t, z_t \rangle + \max_{i \in [d]} R_{T+1,i} \le -F(R_{T+1}) + \max_{i \in [d]} R_{T+1,i} \le \frac{\log d}{\eta}.$$

Substituting $\eta = 1/10$ completes the proof. □

To our knowledge, Theorem 3.3 gives the first $\mathcal{O}(\log d)$ alternating-regret guarantee, which is independent of the horizon T. Together with the $\Omega(\log d)$ lower bound in Theorem 3.1, it establishes the minimax alternating regret as $\Theta(\log d)$ and improves the previous $\mathcal{O}(T^{1/3} \log^{2/3} d)$ upper bound of [HLLZ25]. By applying the reduction of [HLLZ25], the logarithmic alternating-regret bound yields an $\mathcal{O}(\frac{\log d}{T})$ convergence rate for alternating learning in two-player games. Specifically, in two-player zero-sum games, our result yields $\mathcal{O}(\frac{\log d}{T})$ convergence to Nash equilibrium. This matches the average-iterate convergence rate achieved by optimistic multiplicative-weight update under simultaneous play [RS13, SALS15]. Moreover, in two-player general-sum games, it gives the first $\mathcal{O}(\frac{\log d}{T})$ convergence rate to coarse correlated equilibrium. The following corollary follows directly from the reduction of [HLLZ25].

Corollary 3.4. *Consider a two-player normal-form game in which the players have d_x and d_y actions, respectively, and losses in $[-1,1]$. Suppose both players use Algorithm 1 with $\eta = 1/10$ in the alternating learning dynamic of [HLLZ25]. Then the empirical distribution assigning mass $1/(2T)$ to each of the $2T$ indexed profiles $\{(x_t, y_t), (x_{t+1}, y_t)\}_{t \in [T]}$ is an $\varepsilon_{\mathrm{CCE}}$-approximate coarse correlated equilibrium with*

$$\varepsilon_{\mathrm{CCE}} = \mathcal{O}\left(\frac{\log(d_x d_y)}{T}\right).$$

If the game is zero-sum, the time-averaged strategies form an $\varepsilon_{\mathrm{NE}}$-approximate Nash equilibrium with

$$\varepsilon_{\mathrm{NE}} = \mathcal{O}\left(\frac{\log(d_x d_y)}{T}\right).$$

4 Alternating Regret for Online Convex Optimization

Now we shift our focus to the more general OCO problem. In Section 4.1, we first show the lower-bound construction. Compared with the expert case, the main additional difficulty is that convexity allows the learner to hedge continuously between candidate comparators and a more careful geometric construction is required to force a loss gap despite this interpolation. We then give a matching upper bound in Section 4.2 by extending the finite exponential-weights distribution to the continuous setting.

4.1 Lower Bound

Our lower-bound construction can be viewed as a geometric analogue of the expert lower bound. In the expert problem, at each stage we keep a uniformly random half of the surviving experts, with all choices

made before play. A linear loss can assign a small loss to every expert in the surviving subset and a large loss to every expert in the other subset, so the learner incurs a constant loss in expectation relative to any expert that remains. Repeating this elimination over logarithmically many stages gives the logarithmic lower bound. However, this argument does not extend directly to a fixed-dimensional convex domain. If the candidate comparators are represented by points in a convex set, an arbitrary assignment of small losses to one subset and large losses to another may not be realizable by a convex function.

We overcome this issue on the two-dimensional unit disk by providing a different hard instance construction. Fix a boundary point w on the two-dimensional unit disk, and let u be a small counterclockwise rotation of w satisfying $1 - \langle w, u\rangle = \delta > 0$. We use two convex losses $\mathsf{H}_w^{(1)}$ and $\mathsf{H}_{w,\delta}^{(2)}$. The first one $\mathsf{H}_w^{(1)}$ mildly favors staying at w, while the second one $\mathsf{H}_{w,\delta}^{(2)}$ strongly favors moving from w to u. Their values at the two relevant points are

	w	u
$\mathsf{H}_w^{(1)}$	0	$\delta/2$
$\mathsf{H}_{w,\delta}^{(2)}$	$1/2$	$0.$

The second loss is sampled with probability $p = \delta/(1+\delta)$, and the first loss is sampled otherwise. This choice ensures that, before the loss is revealed, every point in the disk has expected loss at least $p/2$. Thus the learner cannot avoid the uncertainty by choosing an intermediate point between w and u.

This gives a two-point hard instance at a single spatial scale δ. To amplify the lower bound, we repeat the construction over a sequence of epochs with geometrically decreasing scales $\delta_1, \delta_2, \ldots, \delta_J$. In epoch $j \in [J]$, we use the corresponding losses $\mathsf{H}_{w_j}^{(1)}$ and $\mathsf{H}_{w_j,\delta_j}^{(2)}$ for $n_j = \Theta(1/\delta_j)$ rounds. If no $\mathsf{H}_{w_j,\delta_j}^{(2)}$ loss appears, the reference point remains at w_j; otherwise we move the reference point to a nearby boundary point w_{j+1}. This creates a constant expected loss gap in each epoch. However, these gaps cannot be summed directly because regret is measured against a single fixed comparator. We therefore decrease the scales geometrically, so that all later movements are much smaller than the movement at the current epoch. As a result, each epoch still contributes a positive expected loss gap against the final-epoch reference point. Hence the constant gaps from different epochs can be accumulated against one fixed comparator. Since the epoch lengths n_j grow geometrically, only $\Theta(\log T)$ epochs fit within T rounds, yielding an $\Omega(\log T)$ lower bound in two dimensions. Finally, taking a product of $\Theta(d)$ such disks and assigning a separate time segment to each coordinate pair yields $\Omega\left(d \log(1 + \frac{T}{d})\right)$.

Formally, we have the following theorem and we defer the complete proof to Appendix B.

Theorem 4.1. *There exists a universal constant $c > 0$ such that, for every $d \geq 2$ and every horizon $T \geq 1$, there exists a d-dimensional compact convex set $\mathcal{X} \subseteq \mathbb{R}^d$ with the following property: for every possibly randomized learner, there exists an oblivious sequence of continuous convex loss functions $f_1, \ldots, f_T : \mathcal{X} \to [0,1]$ such that*

$$\mathbb{E}[\mathrm{Reg}_T^{\mathrm{alt}}] \geq c \cdot d \log\left(1 + \frac{T}{d}\right), \tag{14}$$

where the expectation is over the learner's internal randomness.

4.2 Upper Bound: a Continuous Correction Algorithm

We now turn to the algorithm design for obtaining a matching upper bound. Our proposed algorithm, Algorithm 2, is essentially the continuous counterpart of Algorithm 1. The score vector R_t is replaced by a score function $R_t : \mathcal{X} \to \mathbb{R}$, the exponential-weights distribution is replaced by the density $q_t = q[\eta R_t]$, and finite expectations are replaced by integrals over $\mathcal{X}$. Apart from these changes, the two correction terms in (15) are identical to those in Algorithm 1.

Algorithm 2: Continuous Corrected AltHedge

Input: a d-dimensional compact convex set $\mathcal{X}$ and learning rate $\eta > 0$

Set $f_0 = R_0 = 0$ and $q_0(x) = 1/\operatorname{vol}(\mathcal{X})$.

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

Compute

$$\begin{aligned} \bar{f}_{t-1} &\triangleq \int_{\mathcal{X}} f_{t-1}(x) q_{t-1}(x)\, dx, \\ \widetilde{f}_{t-1}(x) &\triangleq f_{t-1}(x) - \bar{f}_{t-1}, \\ H_{t-1} &\triangleq \int_{\mathcal{X}} \widetilde{f}_{t-1}(x)^2 q_{t-1}(x)\, dx. \end{aligned}$$

Compute $p_t \in \Delta_{\mathcal{X}}$ according to

$$p_t(x) \triangleq q_{t-1}(x)\left[1 - 4\eta \widetilde{f}_{t-1}(x) - 2\eta^2\big(\widetilde{f}_{t-1}(x)^2 - H_{t-1}\big)\right]. \tag{15}$$

Play $x_t = \begin{cases} \displaystyle\int_{\mathcal{X}} x p_t(x)\, dx, & \text{(Option I)}, \\ X_t, \quad X_t \sim p_t, & \text{(Option II)}, \end{cases}$ and observe $f_t : \mathcal{X} \to [-1, 1]$.

Set $R_t = R_{t-1} - (f_t + f_{t-1})$, and set $q_t = q[\eta R_t]$.

As in the expert case, both correction terms in (15) integrate to zero under q_{t-1}, and p_t is a valid probability density when $\eta = \frac{1}{10}$, as formally verified in Proposition C.1. After constructing the corrected density p_t, there are two natural ways to turn it into an action. Option I plays its mean, $x_t = \int_{\mathcal{X}} x p_t(x) dx \in \mathcal{X}$, while Option II samples $x_t \sim p_t$. The first choice gives a deterministic strategy and a deterministic regret guarantee, but requires computing the mean of p_t, which is generally an integration problem. The second avoids this computation and only requires sampling from p_t, but its regret guarantee holds only in expectation.

In particular, for Option II, sampling is particularly convenient. Since R_t is concave, q_t is log-concave. Given the exact values of $\bar{f}_{t-1}$ and H_{t-1}, a sample from p_t can be obtained by rejection sampling from q_{t-1} with constant acceptance probability. Thus Option II incurs only a constant rejection-sampling overhead beyond sampling from the underlying continuous exponential-weights density. A more concrete discussion is deferred to Appendix C. By contrast, implementing Option I requires additionally estimating the mean of p_t, for example using repeated samples. Our regret analysis below considers the exact versions of both options.

For the analysis, we use the same endpoint convention as in the expert case: set $f_{T+1} = 0$ and $R_{T+1} = R_T - f_T$. We compute $\bar{f}_T$, $\widetilde{f}_T$, H_T, and p_{T+1} from (q_T, f_T) using the same formulas as above. Under Option I, set $x_{T+1} = \int_{\mathcal{X}} x p_{T+1}(x)\, dx$; under Option II, sample $x_{T+1} \sim p_{T+1}$. The guarantee of Algorithm 2 is formally stated below. Both options achieve the minimax rate from Theorem 4.1.

Theorem 4.2. *For every d-dimensional compact convex set $\mathcal{X}$ and every adaptive sequence of convex losses $f_1, \dots, f_T : \mathcal{X} \to [-1, 1]$, Algorithm 2 with $\eta = 1/10$ satisfies*

$$\begin{aligned} \operatorname{Reg}_T^{\mathrm{alt}} &\le \mathcal{O}\left(d \log\left(1 + \frac{T}{d}\right)\right) && \textit{under Option I}, \\ \mathbb{E}[\operatorname{Reg}_T^{\mathrm{alt}}] &\le \mathcal{O}\left(d \log\left(1 + \frac{T}{d}\right)\right) && \textit{under Option II}, \end{aligned}$$

where the expectation in the second bound is over the learner's internal randomness.

Theorem 4.2 significantly improves the $\widetilde{\mathcal{O}}(d^{2/3}T^{1/3})$ alternating-regret guarantee of [HLLZ25] for general bounded convex losses. Together with Theorem 4.1, it characterizes the minimax alternating regret. Similar to the expert case, the connection between alternating regret and alternating learning dynamics established by [HLLZ25] immediately yields improved convergence rates in convex games. In particular, using Option I gives deterministic convergence guarantees.

Corollary 4.3. *Consider a two-player game with action sets $\mathcal{X}$ and $\mathcal{Y}$ of dimensions d_x and d_y, respectively, and losses in $[-1, 1]$ that are convex in each player's own action. Suppose both players use Option I of Algorithm 2 with $\eta = 1/10$ in the alternating learning dynamic of [HLLZ25]. Then the empirical distribution assigning mass $1/(2T)$ to each of the $2T$ indexed profiles $\{(x_t, y_t), (x_{t+1}, y_t)\}_{t\in[T]}$ is an $\varepsilon_{\mathrm{CCE}}$-approximate coarse correlated equilibrium with*

$$\varepsilon_{\mathrm{CCE}} = \mathcal{O}\left(\frac{d_x \log(1 + \frac{T}{d_x}) + d_y \log(1 + \frac{T}{d_y})}{T}\right).$$

If the game is convex-concave and zero-sum, the time-averaged strategies form an $\varepsilon_{\mathrm{NE}}$-approximate Nash equilibrium with

$$\varepsilon_{\mathrm{NE}} = \mathcal{O}\left(\frac{d_x \log(1 + \frac{T}{d_x}) + d_y \log(1 + \frac{T}{d_y})}{T}\right).$$

4.3 Proof of Theorem 4.2

The proof is almost the same as the expert analysis in Section 3.2. Recall that the expert proof uses the log-partition potential (5) and the corrected potential (12). Their continuous counterparts are obtained by replacing the uniform average over the d experts by the normalized Lebesgue integral over $\mathcal{X}$. Specifically, for a score function $R : \mathcal{X} \to \mathbb{R}$, with a slight abuse of notation, we define

$$F(R) \triangleq \frac{1}{\eta} \log\left(\frac{1}{\mathrm{vol}(\mathcal{X})} \int_{\mathcal{X}} \exp(\eta R(x))\, dx\right), \tag{16}$$

$$U(R, f) \triangleq F(R) - 2\eta \operatorname{Var}_{x\sim q[\eta R]}(f(x)). \tag{17}$$

Thus F is the continuous analogue of (5), while U is the continuous analogue of (12).

The first ingredient is the direct continuous counterpart of the one-step corrected-potential inequality (13). It shows that the same potential decrease controls the loss averaged under the corrected density.

Lemma 4.4. *Define $h_t \triangleq f_{t-1} + f_t$ for $t \in [T + 1]$. Then, for every $t \in [T + 1]$,*

$$\int_{\mathcal{X}} p_t(x) h_t(x)\, dx \le U(R_{t-1}, f_{t-1}) - U(R_t, f_t). \tag{18}$$

The proof uses the same interpolation as Lemma 3.2. The only additional analytic step is to justify differentiation under the integral. Other than this, the derivative calculation is identical to the expert case. The complete proof is given in Appendix C.

After telescoping (13), the expert proof compares the terminal log-partition potential with the best expert through $\max_i R_{T+1,i} \le F(R_{T+1}) + (\log d)/\eta$. An individual point $u \in \mathcal{X}$ has zero volume, so this discrete comparison has no literal continuous analogue. The following lemma replaces it by comparing u with a small scaled copy of $\mathcal{X}$ around it.

Lemma 4.5. *For every $u \in \mathcal{X}$ and every $\delta \in (0, 1)$,*

$$F(R_{T+1}) \ge R_{T+1}(u) - 4T\delta - \frac{d}{\eta} \log \frac{1}{\delta}. \tag{19}$$

This is the only additional ingredient beyond the expert proof. The proof considers $(1-\delta)u + \delta\mathcal{X}$, whose volume is $\delta^d \operatorname{vol}(\mathcal{X})$, and uses convexity and boundedness to compare the cumulative loss throughout this set with the loss at u. The complete proof is deferred to Appendix C.

We are now ready to prove Theorem 4.2.

Proof of Theorem 4.2. We prove the result for Option II. The guarantee for Option I follows from the same argument together with Jensen's inequality.

Exactly as in the proof of Theorem 3.3, summing (18) over $t = 1, \dots, T+1$ and using $U(R_0, f_0) = 0$ and $f_{T+1} = 0$ gives $\sum_{t=1}^{T+1} \int_{\mathcal{X}} p_t(x) h_t(x)\, dx \le -F(R_{T+1})$. Under Option II, p_t is determined by $f_1, \dots, f_{t-1}$, and $x_t \sim p_t$. Therefore, we have $\mathbb{E}[h_t(x_t) \mid f_1, \dots, f_t] = \int_{\mathcal{X}} p_t(x) h_t(x)\, dx$. Hence, by the tower property,

$$\mathbb{E}\left[\sum_{t=1}^{T+1} h_t(x_t)\right] \le -\mathbb{E}[F(R_{T+1})].$$

For each realized loss sequence, fix any $u \in \mathcal{X}$. By Lemma 4.5, for every $\delta \in (0,1)$, $-F(R_{T+1}) \le -R_{T+1}(u) + 4T\delta + (d/\eta)\log(1/\delta)$. Moreover, by the definition of the terminal score, $R_{T+1}(u) = -2\sum_{t=1}^{T} f_t(u)$. Taking the infimum over $u \in \mathcal{X}$ and combining these inequalities with the endpoint identity $\sum_{t=1}^{T+1} h_t(x_t) = \sum_{t=1}^{T}(f_t(x_t) + f_t(x_{t+1}))$ gives

$$\mathbb{E}[\mathrm{Reg}_T^{\mathrm{alt}}] \le 4T\delta + \frac{d}{\eta}\log\frac{1}{\delta}. \tag{20}$$

When $T \ge 2d$, choose $\delta = d/T$ and recall that $\eta = 1/10$. Substituting these choices into (20) gives $\mathbb{E}[\mathrm{Reg}_T^{\mathrm{alt}}] \le 4d + 10d\log(T/d) = \mathcal{O}(d\log(1+T/d))$. When $T < 2d$, boundedness of the losses gives the deterministic bound $\mathrm{Reg}_T^{\mathrm{alt}} \le 4T = \mathcal{O}(d\log(1+T/d))$. This proves the claimed bound for Option II. Finally, under Option I, convexity of h_t and Jensen's inequality give $h_t(x_t) \le \int_{\mathcal{X}} p_t(x) h_t(x)\, dx$ at every round. Repeating the same argument without taking expectations therefore gives $\mathrm{Reg}_T^{\mathrm{alt}} \le \mathcal{O}(d\log(1+T/d))$ under Option I. □

5 Conclusion

In this paper, we study the minimax alternating regret in the expert problem and online convex optimization. For the d-expert problem, somewhat surprisingly, we achieve $\mathcal{O}(\log d)$ alternating regret, independent of the horizon T, with a matching lower bound. For d-dimensional OCO with bounded convex losses, we establish the minimax rate $\Theta\left(d\log\left(1+\frac{T}{d}\right)\right)$. These results significantly improve the previous $T^{1/3}$-dependent upper bounds and resolve the minimax rates in both settings. Our upper bounds are based on a corrected-potential argument for Hedge and its continuous analogue, while the lower bounds use a halving construction for experts and a multiscale geometric construction for OCO. As immediate consequences, the improved alternating-regret guarantees also yield faster convergence rates for alternating learning dynamics in games. In particular, they give the first $\mathcal{O}(\frac{\log d}{T})$ convergence rate to coarse correlated equilibrium in two-player general-sum games. Interesting directions for future work include obtaining similarly sharp alternating-regret guarantees under partial-information feedback and extending the approach to games with more than two players.

A Omitted Details in Section 3

This appendix verifies the feasibility of the corrected distribution and provides the derivative calculation used in Lemma 3.2. We first verify that the correction in (4) indeed defines a probability distribution.

Proposition A.1. *For every $t = 1, \dots, T+1$, the vector p_t defined in* (4) *with $\eta = 1/10$ belongs to Δ_d.*

Proof. Both correction terms in (4) are centered under q_{t-1}, so $\sum_i p_{t,i} = 1$. Moreover, $|\widetilde{\ell}_{t-1,i}| \le 2$ and $H_{t-1} \ge 0$. Hence, for $\eta = 1/10$, the multiplicative correction factor is at least $1 - 8\eta - 8\eta^2 > 0$. Therefore every coordinate of p_t is nonnegative, and $p_t \in \Delta_d$. □

We next establish the curvature calculation used to choose the variance correction. The proposition below gives both derivatives of the parametric corrected potential and the upper bound on its curvature used in Section 3.2.

Proposition A.2. *Fix a round t, a score vector $R_{t-1} \in \mathbb{R}^d$, a loss vector $\ell_{t-1} \in [-1,1]^d$, and $z \in \mathbb{R}^d$ such that $\ell_{t-1} - z \in [-1,1]^d$. For $\theta \in [0,1]$, let $R(\theta) \triangleq R_{t-1} - \theta z$ and $\ell(\theta) \triangleq \ell_{t-1} - \theta z$. For $\alpha > 0$, define $U_\alpha(R, \ell) \triangleq F(R) - \alpha\eta \operatorname{Var}_{q[\eta R]}(\ell)$ and $G_\alpha(\theta) \triangleq U_\alpha(R(\theta), \ell(\theta))$.*

Let $\mathbb{E}_\theta[\cdot]$ denote expectation under $q[\eta R(\theta)]$, and denote the centered variables by $\widetilde{z}(\theta) \triangleq z - \mathbb{E}_\theta[z]$ and $\widetilde{\ell}(\theta) \triangleq \ell(\theta) - \mathbb{E}_\theta[\ell(\theta)]$. Then

$$G'_\alpha(\theta) = -\mathbb{E}_\theta[z] + 2\alpha\eta\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)] + \alpha\eta^2\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)^2], \tag{21}$$

and

$$\begin{aligned} G''_\alpha(\theta) = (1-2\alpha)\eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2] &- 4\alpha\eta^2\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)] \\ &- \alpha\eta^3\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)^2] + \alpha\eta^3\mathbb{E}_\theta[\widetilde{z}(\theta)^2]\mathbb{E}_\theta[\widetilde{\ell}(\theta)^2] \\ &+ 2\alpha\eta^3\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)]^2. \end{aligned} \tag{22}$$

In particular, we have

$$G''_\alpha(\theta) \le \left[(1-2\alpha)\eta + 8\alpha\eta^2 + 3\alpha\eta^3\right] \operatorname{Var}_{q[\eta R(\theta)]}(z). \tag{23}$$

Proof. Fix a round t. Throughout the proof, we keep the round index only in R_{t-1} and ℓ_{t-1}, and omit it from the quantities defined along the interpolation. Denote the distribution and its expectation operator by $q(\theta) \triangleq q[\eta R(\theta)]$ and $\mathbb{E}_\theta[\cdot] \triangleq \mathbb{E}_{i\sim q(\theta)}[\cdot]$.

We first record a differentiation rule that will be used repeatedly. Define the scalar $\bar{z}(\theta) \triangleq \mathbb{E}_\theta[z]$ and the centered vector $\widetilde{z}(\theta) \in \mathbb{R}^d$ by $\widetilde{z}_i(\theta) \triangleq z_i - \bar{z}(\theta)$. Since $R'_i(\theta) = -z_i$, differentiating the exponential-weights distribution gives $q'_i(\theta) = -\eta q_i(\theta)\widetilde{z}_i(\theta)$. Consequently, for any differentiable indexed family $\phi(\theta) = (\phi_i(\theta))_{i=1}^d$,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[\phi(\theta)] = \sum_i q'_i(\theta)\phi_i(\theta) + \sum_i q_i(\theta)\phi'_i(\theta) = \mathbb{E}_\theta[\phi'(\theta)] - \eta\mathbb{E}_\theta[\widetilde{z}(\theta)\phi(\theta)]. \tag{24}$$

Products and powers inside $\mathbb{E}_\theta[\cdot]$ are interpreted coordinatewise. For example, $\mathbb{E}_\theta[\widetilde{z}(\theta)^2] = \sum_i q_i(\theta)\widetilde{z}_i(\theta)^2$ and $\mathbb{E}_\theta[\widetilde{z}(\theta)\phi(\theta)] = \sum_i q_i(\theta)\widetilde{z}_i(\theta)\phi_i(\theta)$ are scalars.

We next derive the centered quantities needed below. Define the scalar $\bar{\ell}(\theta) \triangleq \mathbb{E}_\theta[\ell(\theta)]$ and the centered vector $\widetilde{\ell}(\theta) \in \mathbb{R}^d$ by $\widetilde{\ell}_i(\theta) \triangleq \ell_i(\theta) - \bar{\ell}(\theta)$. Thus $z = \widetilde{z}(\theta) + \bar{z}(\theta)\mathbf{1}$ and $\ell(\theta) = \widetilde{\ell}(\theta) + \bar{\ell}(\theta)\mathbf{1}$.

Applying (24) with $\phi(\theta) = z$, and noting that z does not depend on θ, gives $\bar{z}'(\theta) = -\eta\mathbb{E}_\theta[\widetilde{z}(\theta)z]$. Since $\mathbb{E}_\theta[\widetilde{z}(\theta)] = 0$, we obtain $\bar{z}'(\theta) = -\eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2]$. Therefore, for every coordinate i, $\widetilde{z}'_i(\theta) = \eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2]$.

Next apply (24) with $\phi(\theta) = \ell(\theta)$. Since $\ell'(\theta) = -z$, we have $\bar{\ell}'(\theta) = -\bar{z}(\theta) - \eta\mathbb{E}_\theta[\widetilde{z}(\theta)\ell(\theta)]$. Using $\ell(\theta) = \widetilde{\ell}(\theta) + \bar{\ell}(\theta)\mathbf{1}$ and $\mathbb{E}_\theta[\widetilde{z}(\theta)] = 0$ gives $\mathbb{E}_\theta[\widetilde{z}(\theta)\ell(\theta)] = \mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)]$. Hence $\bar{\ell}'(\theta) = -\bar{z}(\theta) - \eta\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)]$. Differentiating $\widetilde{\ell}_i(\theta) = \ell_i(\theta) - \bar{\ell}(\theta)$ then yields

$$\widetilde{\ell}_i'(\theta) = -\widetilde{z}_i(\theta) + \eta\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)].$$

The derivatives of the log-partition term follow immediately. Since $\frac{\mathrm{d}}{\mathrm{d}\theta}F(R(\theta)) = -\bar{z}(\theta)$, the identity for $\bar{z}'(\theta)$ gives $\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}F(R(\theta)) = \eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2] = \eta\,\mathrm{Var}_{i\sim q(\theta)}(z_i)$.

We now differentiate the variance term. Let $H(\theta) \triangleq \mathrm{Var}_{q[\eta R(\theta)]}(\ell(\theta)) = \mathbb{E}_\theta[\widetilde{\ell}(\theta)^2]$. Applying (24) and the expression for $\widetilde{\ell}_i'(\theta)$ gives

$$H'(\theta) = -2\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)] - \eta\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)^2].$$

We can now obtain the first derivative of the corrected potential. Since $G_\alpha(\theta) = F(R(\theta)) - \alpha\eta H(\theta)$, combining the expressions for $\frac{\mathrm{d}}{\mathrm{d}\theta}F(R(\theta))$ and $H'(\theta)$ gives

$$G_\alpha'(\theta) = -\bar{z}(\theta) + 2\alpha\eta\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)] + \alpha\eta^2\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)^2].$$

To differentiate $H'(\theta)$ once more, we apply (24) to the two expectations above. For the first term, using the expressions for $\widetilde{z}_i'(\theta)$ and $\widetilde{\ell}_i'(\theta)$ derived above gives

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)] = \mathbb{E}_\theta\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\big(\widetilde{z}(\theta)\widetilde{\ell}(\theta)\big)\right] - \eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)] = -\mathbb{E}_\theta[\widetilde{z}(\theta)^2] - \eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)].$$

Here the remaining terms vanish because $\mathbb{E}_\theta[\widetilde{z}(\theta)] = \mathbb{E}_\theta[\widetilde{\ell}(\theta)] = 0$.

Similarly, applying (24) to $\phi(\theta) = \widetilde{z}(\theta)\widetilde{\ell}(\theta)^2$ gives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)^2] &= \mathbb{E}_\theta\left[\widetilde{z}'(\theta)\widetilde{\ell}(\theta)^2 + 2\widetilde{z}(\theta)\widetilde{\ell}(\theta)\widetilde{\ell}'(\theta)\right] - \eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)^2] \\ &= \eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2]\mathbb{E}_\theta[\widetilde{\ell}(\theta)^2] - 2\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)] + 2\eta\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)]^2 - \eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)^2].\end{aligned}$$

Differentiating $H'(\theta)$ and substituting the two identities above first gives

$$\begin{aligned}H''(\theta) &= -2\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)] - \eta\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)^2] \\ &= -2\left(-\mathbb{E}_\theta[\widetilde{z}(\theta)^2] - \eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)]\right) \\ &\quad - \eta\left(\eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2]\mathbb{E}_\theta[\widetilde{\ell}(\theta)^2] - 2\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)] + 2\eta\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)]^2 - \eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)^2]\right).\end{aligned}$$

Collecting terms, we obtain that

$$\begin{aligned}H''(\theta) &= 2\mathbb{E}_\theta[\widetilde{z}(\theta)^2] + 4\eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)] \\ &\quad + \eta^2\Big(\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)^2] - \mathbb{E}_\theta[\widetilde{z}(\theta)^2]\mathbb{E}_\theta[\widetilde{\ell}(\theta)^2] - 2\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)]^2\Big).\end{aligned}$$

Recall that $\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}F(R(\theta)) = \eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2]$. Since $G_\alpha(\theta) = F(R(\theta)) - \alpha\eta H(\theta)$, we therefore have

$$\begin{aligned}G_\alpha''(\theta) &= \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}F(R(\theta)) - \alpha\eta H''(\theta) \\ &= \eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2] - \alpha\eta\Big(2\mathbb{E}_\theta[\widetilde{z}(\theta)^2] + 4\eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)] \\ &\quad + \eta^2\big(\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)^2] - \mathbb{E}_\theta[\widetilde{z}(\theta)^2]\mathbb{E}_\theta[\widetilde{\ell}(\theta)^2] - 2\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)]^2\big)\Big).\end{aligned}$$

Collecting terms gives

$$\begin{aligned} G''_\alpha(\theta) = (1-2\alpha)\eta\mathbb{E}_\theta[\widetilde{z}(\theta)^2] - 4\alpha\eta^2\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)] \\ - \alpha\eta^3\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)^2] + \alpha\eta^3\mathbb{E}_\theta[\widetilde{z}(\theta)^2]\mathbb{E}_\theta[\widetilde{\ell}(\theta)^2] \\ + 2\alpha\eta^3\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)]^2. \end{aligned} \tag{25}$$

It remains to upper bound the curvature. Since $\ell(\theta) \in [-1,1]^d$, we have $|\widetilde{\ell}_i(\theta)| \le 2$ and $\mathbb{E}_\theta[\widetilde{\ell}(\theta)^2] = \mathrm{Var}_{i\sim q(\theta)}(\ell_i(\theta)) \le 1$. By Cauchy–Schwarz, $\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)]^2 \le \mathbb{E}_\theta[\widetilde{z}(\theta)^2]$, and $|\mathbb{E}_\theta[\widetilde{z}(\theta)^2\widetilde{\ell}(\theta)]| \le 2\mathbb{E}_\theta[\widetilde{z}(\theta)^2]$. Applying these bounds to (25) and dropping its nonpositive term gives (23). □

We finally use the preceding calculation to prove the two properties stated in Lemma 3.2: concavity of the corrected potential along the alternating path and identification of its left derivative with the played distribution.

Proof of Lemma 3.2. For $\alpha = 2$, Proposition A.2 gives $G''_2(\theta) \le (-3\eta + 16\eta^2 + 6\eta^3)\,\mathrm{Var}_{q[\eta R(\theta)]}(z)$. At $\eta = 1/10$, the coefficient is negative, so G_2 is concave on $[0,1]$.

It remains to identify the derivative at $\theta = 0$. Setting $\alpha = 2$ in (21) gives

$$G'_2(\theta) = -\mathbb{E}_\theta[z] + 4\eta\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)] + 2\eta^2\mathbb{E}_\theta[\widetilde{z}(\theta)\widetilde{\ell}(\theta)^2].$$

Using $\widetilde{z}_i(\theta) = z_i - \mathbb{E}_\theta[z]$ and collecting the coefficient of each z_i, we can rewrite this as

$$G'_2(\theta) = -\sum_{i=1}^{d} q[\eta R(\theta)]_i \left[1 - 4\eta\widetilde{\ell}_i(\theta) - 2\eta^2\left(\widetilde{\ell}_i(\theta)^2 - \mathbb{E}_\theta[\widetilde{\ell}(\theta)^2]\right)\right] z_i.$$

At $\theta = 0$, $q[\eta R(0)] = q_{t-1}$, $\widetilde{\ell}_i(0) = \widetilde{\ell}_{t-1,i}$, and $\mathbb{E}_0[\widetilde{\ell}(0)^2] = H_{t-1}$. Therefore $G'_2(0) = -\langle p_t, z\rangle$ by (4), which proves the lemma. □

B Omitted Details in Section 4.1

We first establish the lower bound on a two-dimensional disk. The construction uses epochs at geometrically decreasing spatial scales. In each epoch, the learner incurs a constant expected loss gap relative to a reference point determined by the losses observed up to that epoch. The reference point continues to move in later epochs, but these future movements form a summable series and therefore change the loss on any earlier epoch by only a small constant. Since the epoch lengths grow geometrically, a horizon of length T contains $\Theta(\log T)$ such scales.

Lemma B.1 (Two-dimensional disk lower bound)**.** *Let $\mathcal{D} = \{x \in \mathbb{R}^2 : \|x\|_2 \le 1\}$. There exists a universal constant $c_0 > 0$ such that, for every $T \ge 5$, there is an oblivious distribution P_T over sequences of convex losses $f_1, \dots, f_T : \mathcal{D} \to [0,1]$ for which every learner satisfies*

$$\mathbb{E}_{f_{1:T}\sim\mathsf{P}_T}\left[\mathrm{Reg}_T^{\mathrm{alt}}\right] \ge c_0 \log T. \tag{26}$$

If the learner is randomized, the expectation also includes its internal randomness.

Proof. We first describe the multiscale construction. Fix $w_1 = e_1 = (1,0) \in \partial\mathcal{D}$ and, for $\alpha \in \mathbb{R}$, let $M_\alpha \triangleq \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$ be the counterclockwise rotation matrix. For $w \in \partial\mathcal{D}$ and $\delta \in (0, 1/8]$, define two

loss functions

$$\mathsf{H}^{(1)}_{w}(x) \triangleq \frac{1-\langle w,x\rangle}{2}, \tag{27}$$

$$\mathsf{H}^{(2)}_{w,\delta}(x) \triangleq \frac{1}{2}\max\left\{1-\frac{1-\langle w,x\rangle}{\delta},0\right\}. \tag{28}$$

The first function is affine, while the second is the maximum of two affine functions. Hence both are convex. Moreover, since $0 \le 1-\langle w,x\rangle \le 2$ on $\mathcal{D}$, both losses take values in $[0,1]$.

In our instance construction, we use several variables defined as follows:

$$\delta_1 \triangleq \frac{1}{8},\ \ r \triangleq \frac{1}{4096},\ \ \delta_{j+1} \triangleq r\delta_j,\ \ p_j \triangleq \frac{\delta_j}{1+\delta_j},\ \ n_j \triangleq \left\lfloor \frac{1}{2p_j} \right\rfloor,\ \ \alpha_j \triangleq \arccos(1-\delta_j). \tag{29}$$

This choice makes the cost of later movements in (39) smaller than $3/80$, while the gap in (35) exceeds $1/20$. Let $\tau_0 = 0$ and $\tau_j \triangleq \sum_{k=1}^{j}(n_k+1)$. Epoch j consists of the n_j rounds $\mathcal{I}_j \triangleq \{\tau_{j-1}+1,\dots,\tau_{j-1}+n_j\}$ followed by the zero-loss separator $f_{\tau_j} \equiv 0$, which separates the paired-loss contributions of consecutive epochs.

Now we define our loss function at each epoch. Conditioned on the current boundary point w_j, the losses in epoch $\mathcal{I}_j$ are sampled independently according to the following probability:

$$f_t = \begin{cases} \mathsf{H}^{(2)}_{w_j,\delta_j}, & \text{with probability } p_j, \\ \mathsf{H}^{(1)}_{w_j}, & \text{with probability } 1-p_j. \end{cases} \tag{30}$$

Let N_j be the number of $\mathsf{H}^{(2)}_{w_j,\delta_j}$ loss functions in epoch j. At the end of the epoch, we rotate w_j by α_j if $N_j \ge 1$ and keep the direction otherwise. Specifically, we define w_{j+1} as follows:

$$w_{j+1} \triangleq \begin{cases} w_j, & N_j = 0, \\ M_{\alpha_j} w_j, & N_j \ge 1. \end{cases} \tag{31}$$

All Bernoulli variables used in (30) can be sampled before the interaction begins. Once these variables are fixed, the entire sequence $w_1, w_2, \dots$ and hence the entire loss sequence are fixed as well. Thus the resulting random environment is oblivious.

We next compare the learner's loss with the reference points generated by this construction. For epoch j, define the learner's paired loss $L_j \triangleq \sum_{t\in\mathcal{I}_j}\big(f_t(x_t)+f_t(x_{t+1})\big)$ and for any $v \in \mathcal{D}$, define its (doubled) comparator loss on this epoch by $C_j(v) \triangleq 2\sum_{t\in\mathcal{I}_j} f_t(v)$. The zero-loss separator contributes to neither quantity.

Fix an epoch j and condition on the history before this epoch. For any $\mathcal{F}_{t-1}$-measurable action $x_t \in \mathcal{D}$, let $s = 1-\langle w_j, x_t\rangle$. If $s \le \delta_j$, then the choice $p_j = \delta_j/(1+\delta_j)$ gives that

$$\mathbb{E}[f_t(x_t) \mid \mathcal{F}_{t-1}] = p_j \mathsf{H}^{(2)}_{w_j,\delta_j}(x_t) + (1-p_j)\mathsf{H}^{(1)}_{w_j}(x_t) = \frac{p_j}{2} + \frac{s}{2}\left(1-p_j-\frac{p_j}{\delta_j}\right) = \frac{p_j}{2}.$$

If $s \ge \delta_j$, then the $\mathsf{H}^{(1)}_{w_j}$ loss alone gives $\mathbb{E}[f_t(x_t) \mid \mathcal{F}_{t-1}] \ge \frac{(1-p_j)\delta_j}{2} = \frac{p_j}{2}$. Since every loss is nonnegative, $f_t(x_{t+1}) \ge 0$. Summing over the epoch yields

$$\mathbb{E}[L_j] \ge \frac{n_j p_j}{2}. \tag{32}$$

Now define the reference point after epoch j by $v_j \triangleq w_{j+1}$. If $N_j = 0$, then $v_j = w_j$ and every loss in the epoch is $\mathsf{H}^{(1)}_{w_j}$, so $C_j(v_j) = 0$. If $N_j \geq 1$, then $v_j = M_{\alpha_j} w_j$ and, by the definition of α_j, we know that

$$1 - \langle w_j, v_j \rangle = 1 - \cos \alpha_j = \delta_j.$$

Hence $\mathsf{H}^{(2)}_{w_j,\delta_j}(v_j) = 0$, while each $\mathsf{H}^{(1)}_{w_j}(v_j)$ contributes δ_j to the doubled comparator loss. Therefore, pathwise, we have that

$$C_j(v_j) = (n_j - N_j)\delta_j \mathbf{1}\{N_j \geq 1\}. \tag{33}$$

Let $a_j \triangleq (1-p_j)^{n_j}$. Since $N_j \sim \mathrm{Binomial}(n_j, p_j)$, (33) shows that

$$\mathbb{E}[C_j(v_j)] = \mathbb{E}[n_j\delta_j(1-a_j) - \delta_j N_j] = n_j\delta_j\big((1-p_j) - a_j\big).$$

Combining the above identity with (32) and using $\delta_j = p_j/(1-p_j)$ gives

$$\mathbb{E}[L_j - C_j(v_j)] \geq \frac{n_j p_j}{2} - n_j\delta_j\big((1-p_j) - a_j\big) = \frac{n_j p_j}{2}\left(\frac{2a_j}{1-p_j} - 1\right). \tag{34}$$

Because $\delta_j \leq 1/8$, we have $p_j \leq 1/9$. Moreover,

$$n_j p_j = p_j \left\lfloor \frac{1}{2p_j} \right\rfloor \in \left[\frac{1}{2} - p_j, \frac{1}{2}\right] \subseteq \left[\frac{1}{3}, \frac{1}{2}\right].$$

Also, using $n_j p_j \leq 1/2$ and $\log(1-p) \geq -p/(1-p)$ for $p \in (0,1)$, we know that

$$\frac{a_j}{1-p_j} = (1-p_j)^{n_j - 1} \geq \exp\left(-\frac{(n_j-1)p_j}{1-p_j}\right) \geq e^{-n_j p_j}.$$

Since $u \mapsto u(2e^{-u} - 1)/2$ is decreasing on $[1/3, 1/2]$, for any $j \geq 1$,

$$\mathbb{E}[L_j - C_j(v_j)] \geq \frac{1}{4}\big(2e^{-1/2} - 1\big) > \frac{1}{20}. \tag{35}$$

Thus, in every epoch, the learner's expected paired loss exceeds the loss of the reference point v_j by a universal constant.

Next, we bound the expected loss of the reference point v_j in each epoch k that is before j. Fix $k \leq j$ and define the angle between v_k and v_j to be $\beta_{k\leftrightarrow j}$. Since every update in (31) is a forward rotation, $\beta_{k\leftrightarrow j}$ is at most

$$\beta_{k\leftrightarrow j} \leq \sum_{\ell=k+1}^{\infty} \alpha_\ell = \sum_{\ell=k+1}^{\infty} \arccos(1-\delta_\ell) \leq \sum_{\ell=k+1}^{\infty} \sqrt{3\delta_\ell} = \frac{\sqrt{3r\delta_k}}{1-\sqrt{r}} \leq \frac{4}{3}\sqrt{3r\delta_k}, \tag{36}$$

where the first equality is by the definition of α_ℓ, the second inequality uses the elementary bound $\arccos(1-\delta) \leq \sqrt{3\delta}$ for $\delta \in (0, \frac{1}{8}]$, the second equality follows from $\delta_\ell = r^{\ell-k}\delta_k$, and the final inequality uses $1/(1-\sqrt{r}) \leq 4/3$, which holds since $r = 1/4096$.

We now bound how much the later movement from v_k to v_j can increase the loss of v_j on epoch k. If $N_k = 0$, then $v_k = w_k$ and all losses in the epoch are $\mathsf{H}^{(1)}_{w_k}$. Therefore, using the fact that $1 - \cos z \leq z^2/2$ gives that

$$C_k(v_j) - C_k(v_k) = 2n_k \cdot \left(\mathsf{H}^{(1)}_{w_k}(v_j) - \mathsf{H}^{(1)}_{w_k}(v_k)\right) = n_k(1 - \cos\beta_{k\leftrightarrow j}) \leq \frac{n_k\beta^2_{k\leftrightarrow j}}{2} \leq \frac{8rn_k\delta_k}{3} \leq \frac{3}{2}r, \tag{37}$$

where the second inequality uses (36) and the last inequality is because $n_k \le \frac{9}{16\delta_k}$ as $n_k = \lfloor (1+\delta_k)/(2\delta_k) \rfloor$ and $\delta_k \le 1/8$.

If $N_k \ge 1$, then $v_k = M_{\alpha_k} w_k$. Since all subsequent rotations are in the same direction and $\alpha_k + \beta_{k \leftrightarrow j} < \pi$, we know that

$$1 - \langle w_k, v_j \rangle \ge 1 - \langle w_k, v_k \rangle = \delta_k,$$

meaning that every $\mathsf{H}^{(2)}_{w_k,\delta_k}$ loss function remains zero at v_j. Here, $\alpha_k + \beta_{k \leftrightarrow j} < \pi$ follows from $\alpha_k \le \sqrt{3\delta_k}$, (36), $\delta_k \le 1/8$, and $r = 1/4096$. Thus only the $\mathsf{H}^{(1)}_{w_k}$ losses can increase. Using the fact that $\cos\alpha - \cos(\alpha+\beta) \le \beta(\alpha+\beta)$ for $\alpha, \beta \ge 0$, together with (36), we obtain that

$$\begin{aligned} C_k(v_j) - C_k(v_k) &= (n_k - N_k)\,\langle w_k, v_k - v_j \rangle \\ &\le n_k\big(\cos\alpha_k - \cos(\alpha_k + \beta_{k\leftrightarrow j})\big) \\ &\le n_k \beta_{k\leftrightarrow j}(\alpha_k + \beta_{k\leftrightarrow j}) \le \frac{9}{4}\sqrt{r} + 3r, \end{aligned} \tag{38}$$

where the last inequality uses $\alpha_k \le \sqrt{3\delta_k}$, $n_k \le 9/(16\delta_k)$, $r = 1/4096$, and (36). Combining (37) and (38), and again using $r = 1/4096$, gives the pathwise bound

$$C_k(v_j) - C_k(v_k) \le \frac{9}{4}\sqrt{r} + 3r < \frac{3}{80}. \tag{39}$$

Further combining (35) with (39) shows that for every $k \le j$,

$$\mathbb{E}[L_k - C_k(v_j)] \ge \mathbb{E}[L_k - C_k(v_k)] + \mathbb{E}[C_k(v_k) - C_k(v_j)] \ge \frac{1}{20} - \frac{3}{80} = \frac{1}{80}. \tag{40}$$

We now relate this calculation to the actual hindsight optimizer for the entire prefix. As

$$2 \min_{u \in \mathcal{D}} \sum_{t=1}^{\tau_j} f_t(u) \le 2 \sum_{t=1}^{\tau_j} f_t(v_j) = \sum_{k=1}^{j} C_k(v_j),$$

and the learner's paired loss through round τ_j is $\sum_{k=1}^{j} L_k$ because all separators are zero, we know that

$$\mathrm{Reg}^{\mathrm{alt}}_{\tau_j} \ge \sum_{k=1}^{j} \big(L_k - C_k(v_j)\big). \tag{41}$$

Taking expectations and applying (40) yields

$$\mathbb{E}[\mathrm{Reg}^{\mathrm{alt}}_{\tau_j}] \ge \frac{j}{80}. \tag{42}$$

It remains to count how many complete epochs fit in the horizon. From the definition of n_k and the fact that $\delta_k \le 1/8$, we have $n_k + 1 \le \delta_k^{-1}$. Hence, the total length of the first j epochs satisfies

$$\tau_j = \sum_{k=1}^{j} (n_k + 1) \le \sum_{k=1}^{j} \delta_k^{-1} \le \frac{2}{\delta_j},$$

where the last inequality uses $\delta_k = \delta_j r^{k-j}$ and $r < 1/2$.

In particular, since $\tau_1 = 5$, for every $T \geq 5$, let J be the largest index satisfying $\tau_J \leq T$. Then, we have

$$T < \tau_{J+1} \leq \frac{2}{\delta_{J+1}} = 16r^{-J}.$$

Taking logarithms and using $J \geq 1$ gives

$$\log T \leq \log 16 + J \log(1/r) \leq J \log(16/r), \tag{43}$$

meaning that $J \geq \frac{\log T}{\log(16/r)}$. We then run these J complete epochs and set all remaining losses to zero. Since the additional zero losses do not change the alternating regret, (42) and (43) imply that

$$\mathbb{E}[\mathrm{Reg}_T^{\mathrm{alt}}] \geq \frac{J}{80} \geq \frac{1}{80 \log(16/r)} \log T.$$

□

We are now ready to prove Theorem 4.1. The proof lifts Lemma B.1 to higher dimensions by a direct-product construction. For horizons that are too short to invoke the disk lemma on every two-dimensional coordinate block, we use a separate coordinatewise construction.

Proof of Theorem 4.1. Fix $d \geq 2$, and let $m \triangleq \lfloor d/2 \rfloor$. We first construct an oblivious distribution over loss sequences and then use an averaging argument to obtain a single deterministic loss sequence. We consider two regimes.

Long-horizon regime: $T \geq 5m$. Recall that $\mathcal{D} = \{z \in \mathbb{R}^2 : \|z\|_2 \leq 1\}$. We choose the d-dimensional set as follows

$$\mathcal{X} = \begin{cases} \mathcal{D}^m, & d = 2m, \\ \mathcal{D}^m \times [-1, 1], & d = 2m + 1. \end{cases}$$

Equivalently, for every $x \in \mathcal{X}$ and $k \in [m]$, the coordinate pair (x_{2k-1}, x_{2k}) belongs to $\mathcal{D}$; when d is odd, the last coordinate x_d belongs to $[-1, 1]$ and will not be used.

Let $H \triangleq \lfloor T/m \rfloor$. Since $T \geq 5m$, we have $H \geq 5$. Therefore, by Lemma B.1, there exists an oblivious distribution P_H over sequences $(g_1, \ldots, g_H)$ of convex loss functions $g_s : \mathcal{D} \to [0, 1]$ such that every learner on $\mathcal{D}$ satisfies

$$\mathbb{E}_{g_{1:H} \sim \mathsf{P}_H} \left[\sum_{s=1}^{H} \big(g_s(z_s) + g_s(z_{s+1}) \big) - 2 \min_{z \in \mathcal{D}} \sum_{s=1}^{H} g_s(z) \right] \geq c_0 \log H. \tag{44}$$

Fix one such distribution P_H. We then partition the first mH rounds into m consecutive segments, where segment $k \in [m]$ consists of rounds

$$\mathcal{J}_k \triangleq \{(k-1)H + 1, \ldots, kH\}.$$

Independently for every $k \in [m]$, sample a sequence $(g_{k,1}, \ldots, g_{k,H}) \sim \mathsf{P}_H$, where each $g_{k,s} : \mathcal{D} \to [0, 1]$. All these sequences are sampled before the interaction begins without the knowledge of the learner's decision. For the s-th round of segment k, namely $t = (k-1)H + s$, define

$$f_t(x) \triangleq g_{k,s}\big(x_{2k-1}, x_{2k}\big). \tag{45}$$

For all remaining rounds $t > mH$, set $f_t \equiv 0$. Since $(x_{2k-1}, x_{2k}) \in \mathcal{D}$ for every $x \in \mathcal{X}$, each f_t is well defined, convex, and takes values in $[0, 1]$.

We next show that the alternating regret decomposes across the m segments. For every $k \in [m]$ and $s \in [H+1]$, define

$$z_{k,s} \triangleq \big(x_{(k-1)H+s,\,2k-1}, x_{(k-1)H+s,\,2k}\big) \in \mathcal{D}.$$

Thus $z_{k,s}$ is the pair of coordinates $(2k-1, 2k)$ used by the learner at the corresponding time. For each realized sequence, let

$$z_k^\star \in \arg\min_{z \in \mathcal{D}} \sum_{s=1}^{H} g_{k,s}(z).$$

Define a single comparator $u^\star \in \mathcal{X}$ where $(u^\star_{2k-1}, u^\star_{2k}) = z_k^\star$ for all $k \in [m]$ and, when d is odd, set $u^\star_d = 0$. Since the losses in segment k depend only on coordinates $(2k-1, 2k)$, the minimization over the product domain separates, and hence we have

$$\min_{u \in \mathcal{X}} \sum_{t=1}^{T} f_t(u) = \sum_{k=1}^{m} \min_{z \in \mathcal{D}} \sum_{s=1}^{H} g_{k,s}(z). \tag{46}$$

Together with (45), this gives

$$\mathrm{Reg}_T^{\mathrm{alt}} = \sum_{k=1}^{m} \left[\sum_{s=1}^{H} \big(g_{k,s}(z_{k,s}) + g_{k,s}(z_{k,s+1})\big) - 2 \min_{z \in \mathcal{D}} \sum_{s=1}^{H} g_{k,s}(z) \right], \tag{47}$$

where the losses after round mH are zero and therefore make no contribution.

To lower bound the expectation of each term in (47), fix any segment k and condition on the information available before this segment, namely $\mathcal{F}_{(k-1)H}$. During segment k, $(g_{k,1}, \ldots, g_{k,H})$ is independent of $\mathcal{F}_{(k-1)H}$ and is distributed according to P_H. Therefore, Lemma B.1, applied conditionally on $\mathcal{F}_{(k-1)H}$, yields

$$\mathbb{E}\left[\sum_{s=1}^{H} \big(g_{k,s}(z_{k,s}) + g_{k,s}(z_{k,s+1})\big) - 2 \min_{z \in \mathcal{D}} \sum_{s=1}^{H} g_{k,s}(z) \,\middle|\, \mathcal{F}_{(k-1)H} \right] \geq c_0 \log H.$$

Taking expectations and summing over $k \in [m]$ in (47) gives

$$\mathbb{E}[\mathrm{Reg}_T^{\mathrm{alt}}] \geq c_0 m \log H. \tag{48}$$

Finally, for every $d \geq 2$, we have $m = \lfloor d/2 \rfloor \geq d/3$. Since $T/m \geq 5$, we have $H = \lfloor \frac{T}{m} \rfloor \geq \frac{T}{2m} \geq \frac{T}{d}$. Moreover, $H \geq 5$ implies $\log H \geq \frac{1}{2} \log\big(1 + \frac{T}{d}\big)$. Combining these inequalities with (48) yields $\mathbb{E}[\mathrm{Reg}_T^{\mathrm{alt}}] \geq \frac{c_0}{6} d \log\big(1 + \frac{T}{d}\big)$.

Short-horizon regime: $T < 5m$. When $T < 5m$, the instance construction is from the standard OCO lower bound. Choose $\mathcal{X} = [-1, 1]^d$ and let $K \triangleq \min\{T, d\}$. Draw independent Rademacher signs $\sigma_1, \ldots, \sigma_K$ before play and define $f_t(x) \triangleq \frac{1 + \sigma_t x^{(t)}}{2}$ for all $t \in [K]$, where $x^{(t)}$ denotes the t-th coordinate of x, distinguishing it from the iterate x_t, and $f_t \equiv 0$ for $t > K$.

Since x_t is chosen before σ_t is revealed, we have $\mathbb{E}[f_t(x_t) \mid \mathcal{F}_{t-1}] = 1/2$, while $f_t(x_{t+1}) \geq 0$. On the other hand, define the fixed comparator $u^\star \in [-1,1]^d$ as $u_i^\star = -\sigma_i$ for $i \in [K]$ and $u_i^\star = 0$ for $i > K$. In this case, $f_t(u^\star) + f_{t+1}(u^\star) = 0$ for all t. Hence $\mathbb{E}[\mathrm{Reg}_T^{\mathrm{alt}}] \geq \frac{K}{2}$. Moreover, as $T < 5m \leq 5d/2$, we have

$$K = d \min\left\{\frac{T}{d}, 1\right\} \geq \frac{d}{2}\log\left(1 + \frac{T}{d}\right),$$

where we use $\min\{x, 1\} \geq \log(1+x)/2$ for $x \in [0, 5/2]$.

Thus, in both regimes, there exists a universal constant $c \triangleq \min\left\{\frac{c_0}{6}, \frac{1}{4}\right\} > 0$ and an oblivious distribution over deterministic loss sequences such that $\mathbb{E}[\mathrm{Reg}_T^{\mathrm{alt}}] \geq cd\log\left(1 + \frac{T}{d}\right)$. Since the expectation is an average over deterministic loss sequences, for the fixed learner there exists at least one sequence in the support whose expected alternating regret is at least the same quantity. This finishes the proof. □

C Omitted Details in Section 4.2

This appendix verifies the feasibility of the continuous correction and proves the two auxiliary lemmas used in the proof of Theorem 4.2. The corrected-potential calculation is the continuous counterpart of Proposition A.2. We therefore reuse that calculation whenever possible and only verify the additional analytic step needed to pass from finite expectations to integrals.

We first verify that the corrected density in (15) is well defined and record the rejection sampling representation discussed in Section 4.2.

Proposition C.1. *For $\eta = 1/10$, every density p_t defined in* (15) *belongs to $\Delta_{\mathcal{X}}$. Moreover, if $\Gamma_t(x) \triangleq 1 - 4\eta\widetilde{f}_{t-1}(x) - 2\eta^2\big(\widetilde{f}_{t-1}(x)^2 - H_{t-1}\big)$, then $0 < \Gamma_t(x) < 2$ for every $x \in \mathcal{X}$.*

Proof. By the definitions of $\bar{f}_{t-1}$ and H_{t-1},

$$\int_{\mathcal{X}} \widetilde{f}_{t-1}(x) q_{t-1}(x)\, dx = 0,$$
$$\int_{\mathcal{X}} (\widetilde{f}_{t-1}(x)^2 - H_{t-1}) q_{t-1}(x)\, dx = 0.$$

Hence $\int_{\mathcal{X}} p_t(x)\, dx = 1$. Since $f_{t-1}(x) \in [-1,1]$, $|\widetilde{f}_{t-1}(x)| \leq 2$ and $H_{t-1} \leq 1$. Thus, at $\eta = 1/10$, $\Gamma_t(x) \geq 1 - 8\eta - 8\eta^2 > 0$ and $\Gamma_t(x) \leq 1 + 8\eta + 2\eta^2 < 2$. Therefore p_t is a probability density. □

Remark C.2. Proposition C.1 also gives a simple way to sample from p_t. Recall that $p_t(x) = q_{t-1}(x)\Gamma_t(x)$ and $0 < \Gamma_t(x) < 2$. Given the exact values of $\bar{f}_{t-1}$ and H_{t-1}, draw $X \sim q_{t-1}$ and accept it with probability $\Gamma_t(X)/2$. The density of an accepted proposal is proportional to $q_{t-1}(x)\Gamma_t(x)/2 = p_t(x)/2$. Moreover, $\mathbb{E}_{q_{t-1}}[\Gamma_t] = \int_{\mathcal{X}} p_t(x)\, dx = 1$, so the overall acceptance probability is $1/2$. Therefore, conditioned on acceptance, X has density p_t, and an exact sample from p_t requires two proposals from q_{t-1} in expectation.

We next prove Lemma 4.4. The only new step relative to Appendix A is to verify the moving-expectation identity for the continuous exponential-weights density. Once this identity is established, the derivative calculation in Proposition A.2 applies without change.

Proof of Lemma 4.4. Fix $t \in [T+1]$ and define $R(\theta) \triangleq R_{t-1} - \theta h_t$ and $f(\theta) \triangleq f_{t-1} - \theta h_t$ for $\theta \in [0,1]$. The loss path $f(\theta) = (1-\theta) f_{t-1} - \theta f_t$ takes values in $[-1,1]$. Let $q_\theta \triangleq q[\eta R(\theta)]$. Since $R(\theta)$, h_t, and $f(\theta)$ are uniformly bounded over $\theta \in [0,1]$ and $x \in \mathcal{X}$, differentiation under the integral is justified. Differentiating the exponential-weights density gives

$$\nabla_\theta q_\theta(x) = -\eta q_\theta(x)\left(h_t(x) - \mathbb{E}_{q_\theta}[h_t]\right).$$

Consequently, for any bounded differentiable family ϕ_θ with $\nabla_\theta \phi_\theta(x)$ uniformly bounded over (θ, x), we have

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_{q_\theta}[\phi_\theta] = \mathbb{E}_{q_\theta}[\nabla_\theta \phi_\theta] - \eta \mathbb{E}_{q_\theta}\left[(h_t - \mathbb{E}_{q_\theta}[h_t])\,\phi_\theta\right]. \tag{49}$$

Equation (49) is exactly the continuous counterpart of (24). To make the correspondence with Proposition A.2 explicit, the objects in that proposition are replaced as follows: z is replaced by the function h_t, $q[\eta R(\theta)]$ is replaced by the density q_θ, finite expectations are replaced by integrals under q_θ, and $\widetilde{z}_i(\theta)$ is replaced by $\widetilde{h}_\theta(x) \triangleq h_t(x) - \mathbb{E}_{q_\theta}[h_t]$. Similarly, define $\widetilde{f}_\theta(x) \triangleq f(\theta, x) - \mathbb{E}_{q_\theta}[f(\theta)]$.

Every subsequent step in the proof of Proposition A.2 uses only the moving-expectation identity, differentiation of the affine interpolation, and algebraic identities for centered quantities. Therefore the same calculation gives, for $G_\alpha(\theta) \triangleq F(R(\theta)) - \alpha\eta \operatorname{Var}_{q_\theta}(f(\theta))$,

$$G'_\alpha(\theta) = -\mathbb{E}_{q_\theta}[h_t] + 2\alpha\eta \mathbb{E}_{q_\theta}[\widetilde{h}_\theta \widetilde{f}_\theta] + \alpha\eta^2 \mathbb{E}_{q_\theta}[\widetilde{h}_\theta \widetilde{f}_\theta^2],$$

and

$$G''_\alpha(\theta) \le \left[(1-2\alpha)\eta + 8\alpha\eta^2 + 3\alpha\eta^3\right] \operatorname{Var}_{q_\theta}(h_t).$$

These are precisely (21) and (23), respectively, with the above continuous substitutions. Setting $\alpha = 2$ and $\eta = 1/10$ gives $G''_2(\theta) \le 0$, so G_2 is concave.

The interpolation also has the required endpoints. Indeed, $R(1) = R_{t-1} - h_t = R_t$ and $f(1) = f_{t-1} - h_t = -f_t$. Since variance is unchanged by negation, we have $G_2(0) = U(R_{t-1}, f_{t-1})$ and $G_2(1) = U(R_t, -f_t) = U(R_t, f_t)$. The same identities hold for $t = T+1$ under the endpoint convention in Section 4.2.

It remains to identify the left derivative. Setting $\alpha = 2$ in the first-derivative formula above gives

$$G'_2(\theta) = -\mathbb{E}_{q_\theta}[h_t] + 4\eta \mathbb{E}_{q_\theta}[\widetilde{h}_\theta \widetilde{f}_\theta] + 2\eta^2 \mathbb{E}_{q_\theta}[\widetilde{h}_\theta \widetilde{f}_\theta^2].$$

As in the last step of the proof of Lemma 3.2, collecting the coefficient of $h_t(x)$ rewrites this derivative as

$$G'_2(\theta) = -\int_{\mathcal{X}} q_\theta(x) \left[1 - 4\eta \widetilde{f}_\theta(x) - 2\eta^2 \left(\widetilde{f}_\theta(x)^2 - \mathbb{E}_{q_\theta}[\widetilde{f}_\theta^2]\right)\right] h_t(x)\, dx.$$

At $\theta = 0$, we have $q_\theta = q_{t-1}$, $\widetilde{f}_\theta(x) = \widetilde{f}_{t-1}(x)$, and $\mathbb{E}_{q_{t-1}}[\widetilde{f}_\theta^2] = H_{t-1}$. Hence, by (15), $G'_2(0) = -\int_{\mathcal{X}} p_t(x) h_t(x)\, dx$. Concavity therefore gives $G_2(1) - G_2(0) \le G'_2(0)$, which is exactly (18). □

Finally, we prove Lemma 4.5. Unlike the corrected-potential calculation above, this step is genuinely new in the continuous setting. A single comparator has zero volume, so we compare it with a scaled copy of the domain around it.

Proof of Lemma 4.5. Fix $u \in \mathcal{X}$ and $\delta \in (0,1)$, and let $A_u \triangleq (1-\delta)u + \delta\mathcal{X}$. Since $\mathcal{X}$ is d-dimensional, $\operatorname{vol}(A_u) = \delta^d \operatorname{vol}(\mathcal{X})$. For any $a \in \mathcal{X}$ and $x = (1-\delta)u + \delta a \in A_u$, convexity and the range $f_t \in [-1,1]$ give $f_t(x) \le (1-\delta) f_t(u) + \delta f_t(a) \le f_t(u) + 2\delta$. Because $R_{T+1} = -2\sum_{t=1}^T f_t$, it follows that $R_{T+1}(x) \ge R_{T+1}(u) - 4T\delta$ for every $x \in A_u$.

Restricting the log-partition integral to A_u gives

$$F(R_{T+1}) \ge \frac{1}{\eta} \log\left(\frac{\operatorname{vol}(A_u)}{\operatorname{vol}(\mathcal{X})} \exp\big(\eta(R_{T+1}(u) - 4T\delta)\big)\right) = R_{T+1}(u) - 4T\delta - \frac{d}{\eta}\log\frac{1}{\delta},$$

which proves (19). □

AI Methodology Statement

The author used GPT-5.6 to assist with writing and exploring proof strategies. Specifically, the author developed the lower bound for the experts problem (Section 3.1). In response to the author's prompts, GPT-5.6 generated an initial version of Algorithm 1 (Section 3.2), which the author subsequently simplified substantially. The author also substantially revised the motivation and exposition in Section 3.2 to make the algorithm's design easier to understand. GPT-5.6 also helped develop the lower-bound proof for online convex optimization (Section 4.1) from the experts construction. The author then rewrote and reorganized the proof. The author developed the online convex optimization algorithm (Section 4.2) by extending the experts algorithm to a continuous decision space. GPT-5.6 proposed a more complicated alternative that was not used in the paper. The author has checked the arguments and takes full responsibility for the accuracy, originality, and integrity of the paper.

References

[ALM26] Omar Abbadi, Rida Laraki, and Panayotis Mertikopoulos. Constant regret in general games via higher-order optimism. *arXiv preprint arXiv:2609.04113*, 2026.

[BGP20] James P. Bailey, Gauthier Gidel, and Georgios Piliouras. Finite regret and cycles with fixed-step-size via alternating gradient descent-ascent. In *Proceedings of the 33rd Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 391–407, 2020.

[BMS19] Neil Burch, Matej Moravčík, and Martin Schmid. Revisiting CFR+ and alternating updates. *Journal of Artificial Intelligence Research*, 64:429–443, 2019.

[CBL06] Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[CCK+23] Volkan Cevher, Ashok Cutkosky, Ali Kavis, Georgios Piliouras, Stratis Skoulakis, and Luca Viano. Alternation makes the adversary weaker in two-player games. In *Advances in Neural Information Processing Systems 36*, 2023.

[CP20] Xi Chen and Binghui Peng. Hedging in games: Faster convergence of external and swap regrets. In *Advances in Neural Information Processing Systems 33*, pages 18990–18999, 2020.

[DFG21] Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. In *Advances in Neural Information Processing Systems 34*, pages 27604–27616, 2021.

[FAL+22] Gabriele Farina, Ioannis Anagnostides, Haipeng Luo, Chung-Wei Lee, Christian Kroer, and Tuomas Sandholm. Near-optimal no-regret learning dynamics for general convex games. In *Advances in Neural Information Processing Systems 35*, pages 39076–39089, 2022.

[FS97] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[FS99] Yoav Freund and Robert E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1–2):79–103, 1999.

[Ha26] Junsoo Ha. A horizon-independent regret bound for Optimistic Hedge in general-sum games. *arXiv preprint arXiv:2609.22839*, 2026.

[Haz16] Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016.

[HK10] Elad Hazan and Satyen Kale. Extracting certainty from uncertainty: Regret bounded by variation in costs. *Machine Learning*, 80(2–3):165–188, 2010.

[HLLZ25] Soumita Hait, Ping Li, Haipeng Luo, and Mengxiao Zhang. Alternating regret for online convex optimization. In *Proceedings of the 38th Annual Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 2632–2633, 2025.

[KWW26] Jonas E. Katona, Xiuyuan Wang, and Andre Wibisono. A symplectic analysis of alternating mirror descent. *Journal of Machine Learning Research*, 27(44):1–61, 2026.

[LFO26] Mingyang Liu, Gabriele Farina, and Asuman Ozdaglar. Constant individual regret in general games. *arXiv preprint arXiv:2608.31166*, 2026.

[LPSS25] John Lazarsfeld, Georgios Piliouras, Ryann Sim, and Stratis Skoulakis. Optimism without regularization: Constant regret in zero-sum games. In *Advances in Neural Information Processing Systems 38*, 2025.

[LW94] Nick Littlestone and Manfred K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

[Mai26] Tung Mai. Constant swap regret in general-sum games via two-scale higher-order optimism. *arXiv preprint arXiv:2609.16751*, 2026.

[RS13] Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. *Advances in Neural Information Processing Systems*, 26, 2013.

[SALS15] Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E. Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems 28*, 2015.

[SP26] Ashkan Soleymani and Georgios Piliouras. Multiplicative optimism for constant regret in games. *arXiv preprint arXiv:2609.21976*, 2026.

[Tam14] Oskari Tammelin. Solving large imperfect information games using CFR+. *arXiv preprint arXiv:1407.5042*, 2014.

[TZ26] Yixin Tao and Weiqiang Zheng. Optimal alternating regret for online learning and games. *arXiv preprint arXiv:2608.24731*, 2026.

[WTP22] Andre Wibisono, Molei Tao, and Georgios Piliouras. Alternating mirror descent for constrained min-max games. In *Advances in Neural Information Processing Systems 35*, 2022.

[Zin03] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning*, pages 928–936, 2003.