Optimal anytime regret for two experts

Nicholas J. A. Harvey*, Christopher Liaw*, Edwin A. Perkins†, Sikander Randhawa*
*Department of Computer Science, University of British Columbia
Email: {nickhar,cvliaw,srand}@cs.ubc.ca
†Department of Mathematics, University of British Columbia
Email: perkins@math.ubc.ca

***Abstract*—The multiplicative weights method is an algorithm for the problem of prediction with expert advice. It achieves the optimal regret asymptotically if the number of experts is large, and the time horizon is known in advance. Optimal algorithms are also known if there are exactly two, three or four experts, and the time horizon is known in advance.**

In the anytime setting, where the time horizon is *not* known in advance, algorithms can be obtained by the "doubling trick", but they are not optimal, let alone practical. No minimax optimal algorithm was previously known in the anytime setting, regardless of the number of experts.

We design the first minimax optimal algorithm for minimizing regret in the anytime setting. We consider the case of two experts, and prove that the optimal regret is $\gamma\sqrt{t}/2$ at all time steps t, where γ is a natural constant that arose 35 years ago in studying fundamental properties of Brownian motion. The algorithm is designed by considering a continuous analogue of the regret problem, which is solved using ideas from stochastic calculus.

This is the extended abstract of the paper. The full paper can be found in [arXiv:2002.08994].

***Keywords*-regret; online learning; stochastic calculus**

I. Introduction

We study the problem of prediction with expert advice, whose origin can be traced back to the 1950s [1]. The problem is a sequential game between an adversary and an algorithm as follows. There are n actions, which are called "experts". At each time step, the algorithm computes a distribution over the experts, then randomly chooses an expert according to that distribution; concurrently, the adversary chooses a cost for each expert, with knowledge of the algorithm's distribution but not its random choice. The cost of each expert is then revealed to the algorithm, and the algorithm incurs the cost that its chosen expert incurred. The goal is to design an algorithm whose expected *regret* is small. That is, the goal is to minimize the difference between the algorithm's expected total cost and the total cost of the best expert. This problem and its variants have been a key component in many results in theoretical computer science; see, e.g., [2].

The most well-known algorithm for the experts problem is the celebrated multiplicative weights update algorithm (MWU) [3], [4]. In the fixed-time setting (where a time horizon T is known in advance), MWU suffers a regret of $\sqrt{(T/2)\ln n}$ at time T, where n is the number of experts [5], [6]. This bound on the regret of MWU is known to be tight for any even n [7]. It is also known [5] that $\sqrt{(T/2)\ln n}$ is asymptotically optimal for any algorithm as $n, T \to \infty$, in an appropriate sense. Hence, MWU is a minimax optimal[1] algorithm as $n, T \to \infty$. Interestingly, MWU is *not* optimal for small values of n. For $n = 2$, Cover [8] observed decades earlier that a natural dynamic programming formulation of the problem leads to a simple analysis showing that the minimax optimal regret is $\sqrt{T/2\pi}$.

For some applications, the time horizon T is not known in advance; examples include any sort of online tasks (e.g., online learning), or tasks requiring convergence over time (e.g., convergence to equilibria). An alternative model, more suited to those scenarios, is the *anytime setting*[2], in which algorithms are not given T but must bound the regret *for all* T. Yet another model is to assume that T is random with a known distribution [9]. For example, the *geometric horizon setting* of Gravin, Peres, and Sivan [10] assumes that T is a geometric random variable. In this setting, they gave the optimal algorithm for two and three experts. They also propose a conjecture on the relationship between the fixed-time and the geometric horizon settings that could lead to optimal bounds for all n.

Our focus is the anytime setting. One can convert algorithms for the fixed-time setting to the anytime setting by the well-known "doubling trick" [5, §4.6]. This involves restarting the fixed-time horizon algorithm every power-of-two steps with new parameters. If the fixed-time algorithm has regret $O(T^c)$ at time T for some $c \in (0, 1)$ then the doubling trick yields an algorithm with regret $O(t^c)$ at time t for every $t \geq 1$. Although this reduction is conceptually simple,

[1]This means that the algorithm minimizes the maximum, over all adversaries, of the regret.

[2]Other authors have referred to this setting as an "unknown time horizon" or "bounds that hold uniformly over time".

restarting the algorithm and discarding its state is clearly wasteful and probably not very practical.

In lieu of the doubling trick, one can use variants of MWU with a dynamic step size; see, e.g., [11, §2.3], [12, Theorem 1], [13, §2.5], [14, Corollary 5.5]. This is a much more elegant and practical approach and is even simpler to implement. However, the analysis is more difficult than for MWU, and is rarely taught. It is known that, with an appropriate choice of step sizes, MWU can guarantee[3] a regret of $\sqrt{t \ln n}$ for all $t \geq 1$ and all $n \geq 2$ (see [13, Theorem 2.4] or [15, Proposition 2.1]). Until our work it was unknown, for every n, whether $\sqrt{t \ln n}$ is the minimax anytime regret.

Results and techniques: This work considers the anytime setting with $n = 2$ experts. We show that the optimal regret is $\frac{\gamma}{2}\sqrt{t}$, where $\gamma \approx 1.30693$ is a fundamental constant that arises in the study of Brownian motion [16]. (Note that $\gamma/2 \approx 0.653 < 0.833 \approx \sqrt{\ln 2}$.) It is not a priori obvious why this fundamental constant should play a role in both Brownian motion and regret. Nevertheless, some connections are known. For example, in the fixed-time setting, the optimal algorithms for $n \in \{2, 3, 4\}$ (see [10]) and the optimal lower bound for $n \to \infty$ all involve properties of random walks. Since Brownian motion is a continuous limit of random walks, a connection between anytime regret and Brownian motion is plausible.

Our techniques to analyze the optimal anytime regret are a significant departure from previous work on regret minimization. First, we define a continuous-time analogue of the problem which expresses the regret as a stochastic integral. This allows us to utilize tools from stochastic calculus to arrive at a potential function whose derivative gives the optimal *continuous*-time algorithm. Remarkably, the optimal *discrete*-time algorithm is the *discrete* derivative of the same potential function. We do note that some prior work on regret minimization have also made use of discrete derivatives of other potential functions [17], [18].

The potential function that we derive involves a "confluent hypergeometric function". Such functions often arise in solutions to differential equations, and are useful in discrete mathematics [19, §5.5]. In addition, they appear to be inherent to our problem since they also arise in the matching lower bound.

Application: An interesting application of our results is to a problem in probability theory that does not involve regret at all. Let $(X_t)_{t \geq 0}$ be a standard random walk. Then $\mathrm{E}[\,|X_\tau|\,] \leq \gamma\,\mathrm{E}[\,\sqrt{\tau}\,]$ for every stopping time τ; moreover, the constant γ cannot be improved.[4] This result is originally due to Davis [20, Eq. (3.8)], who proved it first for Brownian motion and later derived the result for random walks (via the Skorokhod embedding). We give a new derivation of Davis' result from our results in Subsection II-D.

Related work: The minimax regret for the experts problem has been well-studied in the fixed-time horizon setting. For two experts the minimax regret was shown to be $\sqrt{T/2\pi}$ by Cover in 1965 [8]. It has been known for twenty years that $\sqrt{T \ln(n)/2}$ is the minimax regret as $n \to \infty$ [5], [6]. Building on the work of Gravin et al. [10], it has recently been shown that the minimax regret is $\sqrt{8T/9\pi}$ for three experts [21] and $\sqrt{\pi T/8}$ for four experts [22]. The anytime setting is not as well understood. In the two-experts setting, Luo and Schapire [9] demonstrate that, if the time horizon T is chosen by an adversary and unknown to the algorithm then the algorithm may be forced to incur regret at least $\sqrt{T/\pi}$. This exceeds the minimax regret of $\sqrt{T/2\pi}$ if T is known to the algorithm a priori, which indicates that the adversary has more power to force regret when it is allowed to select the time horizon.

Recently, interactions between algorithms in discrete and continuous *time* have been fruitful in other lines of work, e.g., [23], [24], [25], [26], [27], [28], [29]. There is also a line of work that makes connections between the experts problem (in the finite-time horizon and geometric-time horizon setting) and PDEs [22], [30], [31], [32], [33], [34]. There is also work connecting regret minimization to option pricing [35] and to the Black-Scholes formula [36], which is based on Brownian motion and stochastic calculus. Intuitively, stochastic calculus is a crucial tool to optimally hedge against future costs, which we exploit too.

Our work crucially uses stopping times for Brownian motion hitting a time-dependent boundary. Such techniques have also been used for non-adversarial bandits to approximate Gittins indices (see, e.g., [37]).

II. Discussion of results and techniques

A. Formal problem statement

The previous section informally described the model as involving an algorithm that randomly selects an expert. Here we will instead describe the algorithm as being deterministic but selecting a distribution over experts. The latter definition is consistent with the viewpoint of online convex optimization (see, e.g., [13], [38]). We will mention below some subtleties that arise when the algorithm makes random selections.

[3] It can be shown, by modifying arguments of [7], that this is the optimal anytime analysis for MWU with step sizes $c/\sqrt{t}$.

[4] At first glance, the inequality may seem to contradict the law of the iterated logarithm. However, we remark that if $\tau := \inf\{t > 0 : |X_t| \geq c\sqrt{t \ln \ln t}\}$ for some $c \in (0, \sqrt{2})$ then $\mathrm{E}[\,\sqrt{\tau}\,] = \infty$ (despite τ being a.s. finite) and the inequality is trivial.

Let n denote the number of experts. There is a deterministic algorithm $\mathcal{A}$, and a deterministic adversary $\mathcal{B}$ that knows $\mathcal{A}$. For each integer $t \geq 1$, there is a prediction task that is said to occur at time t. In this task, $\mathcal{A}$ picks a probability distribution $x_t \in [0,1]^n$, and $\mathcal{B}$ picks a cost vector $\ell_t \in [0,1]^n$. The coordinate $\ell_{t,j}$ denotes the cost of the j^{th} expert at time t.

After x_t is chosen the vector ℓ_t is revealed, so x_t depends on $\ell_1, \ldots, \ell_{t-1}$ (and implicitly $x_1, \ldots, x_{t-1}$). The vector ℓ_t depends on $\mathcal{A}$ and on $\ell_1, \ldots, \ell_{t-1}$ (and implicitly $x_1, \ldots, x_t$, since $\mathcal{A}$ is deterministic and known to $\mathcal{B}$). The game can end whenever $\mathcal{B}$ wishes, or continue forever. Since $\mathcal{A}$ is deterministic and known to $\mathcal{B}$, the entire sequence of interactions, including the ending time, can be predetermined by $\mathcal{B}$.

The cost incurred by the algorithm at time t is the inner product $\langle x_t, \ell_t \rangle$. This may be thought of as the "expected cost" of the algorithm, although the algorithm is actually deterministic. The total expected cost of the algorithm up to time t is $\sum_{i=1}^{t} \langle x_i, \ell_i \rangle$. For $j \in [n]$, the total cost of the j^{th} expert up to time t is $L_{t,j} = \sum_{i=1}^{t} \ell_{i,j}$. The regret at time t of algorithm $\mathcal{A}$ against adversary $\mathcal{B}$ is the difference between the algorithm's total expected cost and the total cost of the best expert, i.e.,

$$\text{Regret}(n, t, \mathcal{A}, \mathcal{B}) \;=\; \sum_{i=1}^{t} \langle x_i, \ell_i \rangle \;-\; \min_{j \in [n]} L_{t,j}.$$

Anytime setting: This work focuses on the anytime setting. In this setting, one may view the algorithm as running forever, with the goal of minimizing, for *all* t, the regret normalized by $\sqrt{t}$. Alternatively, one may view the game as ending at a time chosen by the adversary, and the algorithm must minimize the regret at that ending time. (It does not matter whether the adversary chooses the ending time in advance or dynamically, since $\mathcal{A}$ and $\mathcal{B}$ are deterministic so all interactions are predetermined.) These two views are equivalent because the algorithm cannot distinguish between them.

Formally, we will design an algorithm which achieves the infimum in the following expression.

$$\begin{aligned} &\text{AnytimeNormRegret}(n) \\ &\qquad := \inf_{\mathcal{A}} \sup_{\mathcal{B}} \sup_{t \geq 1} \frac{\text{Regret}(n, t, \mathcal{A}, \mathcal{B})}{\sqrt{t}}. \end{aligned} \tag{II.1}$$

The minimax anytime regret is unknown even in the case of $n = 2$. The best known bounds at present are

$$\begin{aligned} 0.564 \;\approx\; \sqrt{1/\pi} \;\leq\; &\text{AnytimeNormRegret}(2) \\ &\leq\; \sqrt{\ln 2} \;\approx\; 0.833. \end{aligned} \tag{II.2}$$

The lower bound, due to [9], demonstrates a gap between the anytime setting and the fixed-time setting, where the optimal normalized regret is $\sqrt{1/2\pi}$ [8]. Our main result is that $\text{AnytimeNormRegret}(2) = \gamma/2 \approx 0.653$ and consequently neither inequality in (II.2) is tight.

As mentioned above, MWU with a dynamic step size shows that $\text{AnytimeNormRegret}(n) \leq \sqrt{\ln n}$ for all values of n at least 2 [13, §2.5]. The lower bound $\liminf_{n\to\infty} \text{AnytimeNormRegret}(n)/\sqrt{\ln n} \geq \sqrt{1/2}$ follows from the bound in the fixed-time setting [5]. Thus, the upper bound is loose by at most a factor $\sqrt{2}$.

Remark on randomization: Several alternative formulations of the problem arise if $\mathcal{A}$ selects a single expert $I_t \in [n]$ randomly at each time t, and the adversary chooses an ending time τ. The differences relate to the power of the adversary $\mathcal{B}$. The most powerful adversary has ℓ_t depending on $I_1, \ldots, I_t$, in which case it is easy to design $\mathcal{B}$ with $\text{Regret}(n, t, \mathcal{A}, \mathcal{B}) = \Omega(t)$. Another interesting possibility is if the cost vector ℓ_t and the event $\tau = t$ are determined by $I_1, \ldots, I_{t-1}$. (This is analogous to the "non-oblivious opponent" of [11, §4.1].) In this case, one can design an adversary $\mathcal{B}$ for which $\text{E}\left[\frac{\text{Regret}(n,\tau,\mathcal{A},\mathcal{B})}{\sqrt{\tau \log\log \tau}}\right] = \Omega(1)$; this is a consequence of the law of the iterated logarithm.[5] A third possibility is if ℓ_t and τ depend only on $\mathcal{A}$ and not its random choices $I_1, I_2, \ldots$. (This is analogous to the "oblivious opponent" of [11, §4.1].) The expected regret in this model is identical to the regret in the deterministic model described at the start of this section. We favour this third model because its minimax regret has the familiar bound $\Theta(\sqrt{t})$, and it is consistent with the online convex optimization literature. It is intriguing that in the anytime setting, a non-oblivious adversary has more power than an oblivious adversary. In contrast, the two adversaries have the same power in the fixed time setting [11, §4.1].

B. Statement of results

To state our results, we require two definitions. Recall that the imaginary error function is defined as $\text{erfi}(x) \;=\; \frac{2}{\sqrt{\pi}} \int_0^x e^{z^2}\, \mathrm{d}z$. Next, we define

$$M_0(x) \;=\; e^x - \sqrt{\pi x}\, \text{erfi}(\sqrt{x}). \tag{II.3}$$

The function M_0 is an example of a confluent hypergeometric function, a very broad class of special functions that includes, e.g., Bessel functions and Laguerre polynomials. Our analysis makes use of a few elementary properties of these functions. A key

[5] The asserted lower bound holds for "reasonable" algorithms that never place more than half the weight on the worst expert. We believe that the lower bound holds for all algorithms but have not worked out the details.

constant used in this paper is γ, which is defined to be the smallest[6] positive root[7] of $M_0(x^2/2)$, i.e.,

$$\gamma \;:=\; \min\left\{\, x > 0 \,:\, M_0(x^2/2) = 0 \,\right\}. \tag{II.4}$$

Numerically, $\gamma \approx 1.30693\ldots$.

Theorem II.1 (Main result). In the anytime setting with two experts, the minimax normalized regret (over deterministic algorithms $\mathcal{A}$ and adversaries $\mathcal{B}$) is

$$\mathrm{AnytimeNormRegret}(2) \;=\; \frac{\gamma}{2}. \tag{II.5}$$

The proof of this theorem has two parts: an upper bound, in Section III, which exhibits an optimal algorithm, and a lower bound, in Section IV, which exhibits an optimal randomized adversary. The algorithm is very short, and it appears below in Algorithm 1.

One might imagine that some form of duality theory is involved in our matching upper and lower bounds. Indeed, if the costs are in $\{0,1\}$ one may write $\mathrm{AnytimeNormRegret}(2)$ as the value of an infinite-dimensional linear program, although we do not explicitly adopt this viewpoint. Instead, γ arises in our lower bound as the maximizer in (IV.3), whereas γ arises in our upper bound as the optimizer of a certain boundary condition for a PDE (see Section V). Nevertheless, our algorithm and our lower bound can be seen as constructing feasible primal and dual solutions, respectively, to the aforementioned linear program.

Comparison to existing techniques: A duality viewpoint is adopted by Gravin et al. [10] in the fixed-time and geometric horizon settings using von Neumann's minimax theorem. Their dual problem is characterized by properties of random walks, which allows one to determine the optimal dual value directly without reference to the primal. It is conceivable that some form of von Neumann's minimax theorem can be applied for the anytime setting, although it is unclear due to the appearance of the supremum and $1/\sqrt{t}$ in (II.5). Our results of Section IV may be viewed as using random walks to construct feasible dual solutions of value $\gamma/2 - \epsilon \;\; \forall \epsilon > 0$, but it is *not obvious* that these solutions converge to the optimal dual value.

The only way we know of to prove optimality of those dual solutions is to construct an algorithm whose regret is $\gamma\sqrt{t}/2$. This is the more challenging part of this paper, which we discuss in Sections III. Interestingly, unlike some previous work, we explicitly obtain an optimal algorithm for costs in $[0,1]$, not just for costs in $\{0,1\}$.

[6]In fact, γ is the *unique* positive root.

[7]The *roots* of certain confluent hypergeometric functions have appeared in studying some natural phenomena of Brownian motion; for some examples see [39], [20], [40], [16].

Remark. Our lower bound can be strengthened to show that, for any algorithm $\mathcal{A}$,

$$\sup_{\mathcal{B}} \limsup_{t\geq 1} \frac{\mathrm{Regret}(2,t,\mathcal{A},\mathcal{B})}{\sqrt{t}} \;\geq\; \frac{\gamma}{2}.$$

In particular, even if $\mathcal{A}$ is granted a "warm-up" period during which its regret is ignored, an adversary can still force it to incur large regret afterwards. More details can be found in the full version of this paper.

The algorithm's description and analysis relies heavily on a function $R\colon \mathbb{R}_{\geq 0} \times \mathbb{R} \to \mathbb{R}$ defined by

$$R(t,g) = \begin{cases} 0 & (t=0) \\ \frac{g}{2} + \kappa\sqrt{t} M_0\left(g^2/2t\right) & (t>0 \;\&\; g \leq \gamma\sqrt{t}) \\ \frac{\gamma\sqrt{t}}{2} & (t>0 \;\&\; g \geq \gamma\sqrt{t}) \end{cases}, \tag{II.6}$$

where $\kappa = \frac{1}{\sqrt{2\pi}\,\mathrm{erfi}(\gamma/\sqrt{2})}$ and M_0 as defined in (II.3). The function R may seem mysterious at first, but in fact arises naturally from the solution to a stochastic calculus problem; this is briefly discussed in Section V and more details can be found in the full version of the paper. In our usage of this function, t will correspond to the time and g will correspond to the *gap* between (i.e., absolute difference of) the total loss for the two experts. One may verify that R is continuous on $\mathbb{R}_{>0} \times \mathbb{R}$ because the second and third cases agree on the curve $\left\{\, (t, \gamma\sqrt{t}) \,:\, t > 0 \,\right\}$ since γ satisfies $M_0(\gamma^2/2) = 0$. We next define a function p to be

$$p(t,g) = \tfrac{1}{2}\big(R(t,g+1) - R(t,g-1)\big). \tag{II.7}$$

This is the discrete derivative of R at time t and gap g. The algorithm constructs its distribution x_t so that $p(t,g)$ is the probability mass assigned to the expert with the greatest accumulated loss so far. It is shown later that $p(t,g) \in [0,1/2]$ whenever $t \geq 1$ and $g \geq 0$ so that p is indeed a probability and the algorithm is well defined. We remark that $p(t,0) = 1/2$ (Lemma III.3) for all $t \geq 1$ so the algorithm places equal mass on both experts when their cumulative losses are equal.

C. Techniques

Lower Bound: The common approach to prove lower bounds in the experts problem is to consider a random adversary that changes the gap by ± 1 at each step. In the fixed-time setting, the adversary has no control over the time horizon; it is known to both the adversary and the algorithm beforehand. The adversary in the anytime setting has the additional power to choose the time horizon, without informing the algorithm, and therefore it is not surprising that an adversary using a fixed time horizon does not provide a good anytime lower bound.

Algorithm 1 The algorithm achieving the minimax anytime regret for two experts. At each time step, each expert incurs a cost in the interval $[0,1]$, so the cost vector ℓ_t lies in $[0,1]^2$. Here, p is the function defined by (II.7).

Initialize $L_0 \leftarrow \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$.
for $t = 1, 2, \ldots$ **do**
  Swap indices so that $L_{t-1,1} \geq L_{t-1,2}$.
  The current gap is $g_{t-1} \leftarrow L_{t-1,1} - L_{t-1,2}$.
  Set $x_t \leftarrow \left[\, p(t, g_{t-1}),\ 1 - p(t, g_{t-1}) \,\right]$.
  ▷ Observe vector ℓ_t and incur cost $\langle\, x_t,\, \ell_t \,\rangle$.
  $L_t \leftarrow L_{t-1} + \ell_t$
end for

To obtain the optimal lower bound, we allow the adversary to select a *random time*, τ, as the time horizon. As a first step, let us view the regret as a discrete stochastic process. To analyze this stochastic process, we use an elementary identity known as Tanaka's Formula for random walks, which allows us to write the regret process as $\mathrm{Regret}(t) = Z_t + g_t/2$ where Z_t is a martingale with $Z_0 = 0$ and g_t is the current gap at time t. When τ is a sufficiently "nice" *stopping time*[8], the *Optional Stopping Theorem* (OST) yields $\mathrm{E}\,[\,Z_\tau\,] = Z_0 = 0$. (This step is trivial in the fixed-time and geometric horizon settings since they involve stopping times that are always nice.)

In particular, we consider adversaries that select τ as the first time that the gap g_t exceeds some time dependent boundary $f(t)$[9]. Applying the OST, one might expect that $\mathrm{E}\,[\,\mathrm{Regret}(\tau)\,] = \mathrm{E}\,[\,g_\tau\,]/2 \geq \mathrm{E}\,[\,f(\tau)\,]/2$. Unfortunately, such an argument must involve additional assumptions; otherwise the adversary could just select the boundary $f(t)$ to be arbitrarily large, and the resulting regret lower bound would violate known upper bounds.

The issue lies in the fact that the OST requires certain conditions on the martingale and stopping time. First observe that it is *not* sufficient for the stopping time to be *almost surely finite*. (Otherwise, one could use a boundary $f(t) = \Theta(\sqrt{t \ln \ln t})$ and the law of the iterated logarithm [41] to prove lower bounds that contradict the $O(\sqrt{t})$ upper bound of Cover or MWU.) It is tempting to fix this by considering only boundaries where $\mathrm{E}\,[\,\tau\,] < \infty$. However, this restriction makes the adversaries much too weak. Suppose that we consider boundaries of the form $f(t) = c\sqrt{t}$, as this would be in harmony with the known $\Theta(\sqrt{t})$ regret bounds. To ensure that $\mathrm{E}\,[\,\tau\,] < \infty$, it is known [42], [39] that choosing $c < 1$ is necessary and sufficient. Unfortunately this would yield a regret lower bound of $\sqrt{t}/2$, which is trivial since the algorithm can easily be forced to have regret $1/2$ at time $t = 1$. Therefore, we must relax the restriction that $\mathrm{E}\,[\,\tau\,] < \infty$.

Fortunately there is a strengthening of the OST with a weaker and somewhat surprising hypothesis that leads to optimal results in our setting. We show that the optimal adversary chooses a stopping time to satisfy this weak hypothesis. This strengthened OST states: if Z_t is a martingale with bounded increments (i.e. $\sup_{t \geq 0} |Z_{t+1} - Z_t| \leq K$ for some $K > 0$) and τ is a stopping time satisfying $\mathrm{E}\left[\,\sqrt{\tau}\,\right] < \infty$, then $\mathrm{E}\,[\,Z_\tau\,] = 0$. The crucial detail is to bound the *expected square root* of τ. This result is stated formally in Theorem IV.2. It remains to choose as large a boundary as possible such that the associated stopping time of hitting the boundary satisfies $\mathrm{E}\,[\,\sqrt{\tau}\,] < \infty$. Using classical results of Breiman [39] and Greenwood and Perkins [40], we show that the optimal choice of c is γ.

Upper Bound: Our analysis of Algorithm 1, to prove the upper bound in Theorem II.1, uses a deceptively simple argument where R defined in (II.6) acts as a potential function. Specifically, we show that the change in regret from time $t-1$ with gap g_{t-1} to time t with gap g_t is at most $R(t, g_t) - R(t-1, g_{t-1})$. This implies that $\max_g R(t, g)$ is an upper bound on the regret at time t. The analysis has a number of key features. First, note that the potential function R is bivariate; it depends on both the *time* t as well as the *state* g_t. To deal with this bivariate potential, we use a tool known as the discrete Itô formula. This formula allows us to relate the regret to the potential R, while elegantly handling changes to both time and state. In fact, the potential R turns out to be an extremely tight approximation to the actual regret. Previously, there have been several works that make use of bivariate potentials (e.g. [43], [17]). However, to the best of our knowledge, our work is the first to use the discrete Itô formula in the setting of regret minimization.

The function R and the use of discrete Itô do not come "out of thin air"; they come from considering a continuous-time analogue of the problem. This continuous viewpoint brings a wealth of analytical tools that do not exist (or are more cumbersome) in the discrete setting. As discussed in the lower bound section above, in discrete-time it is natural to assume the gap process evolves as a reflected random walk. In order to formulate the continuous-time problem, we assume that the continuous adversary evolves the gap between the best and worst expert as a reflected Brownian motion (the continuous-time analogue of a random walk). Using this adversary, the continuous-time regret becomes a stochastic integral.

[8] Intuitively, a stopping time must make the decision that now is the time to stop without knowledge of future random bits.

[9] Note that $\tau = \min\{\, t \geq 0 : g_t \geq f(t) \,\}$ is a stopping time.

The most natural way to analyze an integral is to use the fundamental theorem of calculus (FTC). However, the continuous-time regret is defined by a stochastic integral so the FTC cannot be applied[10]. However there is a stochastic analog of the FTC, namely the (continuous) Itô formula, which we state in Theorem V.3. We use it to provide an insightful decomposition of the continuous-time regret. In particular, this decomposition suggests that the algorithm should satisfy an analytic condition known as the *backwards heat equation*. A key resulting idea is: if the algorithm satisfies the backward heat equation, then there is a natural potential function that upper bounds the regret of the algorithm. This enables a systematic approach to obtain an explicit continuous-time algorithm and a potential function that bounds the continuous algorithm's regret. To go back to the discrete setting, using the *same* potential function, we replace applications of Itô's formula with the discrete Itô formula. Remarkably, this leads to *exactly* the same regret bound as the continuous setting.

D. Application

As mentioned in Section I, the following theorem of Davis can be proven as a corollary of our techniques. Intriguingly, the proof involves regret, despite the fact that regret does not appear in the theorem statement.

Theorem II.2 (Davis [20]). Let $(X_t)_{t\geq 0}$ be a standard random walk. Then $\mathrm{E}\,[\,|X_\tau|\,] \leq \gamma\,\mathrm{E}\,[\,\sqrt{\tau}\,]$ for every stopping time τ; moreover, the constant γ cannot be improved.

Proof: We begin by proving the first assertion. Suppose that $\mathrm{Regret}(T)$ is the regret process when Algorithm 1 is used against a random adversary. As discussed in Subsection II-C and (IV.2), we can write the regret process as $\mathrm{Regret}(T) = Z_T + g_T/2$ where Z_T is a martingale and g_T evolves as a reflected random walk.[11] Moreover, if τ is a stopping time satisfying $\mathrm{E}\,[\,\sqrt{\tau}\,] < \infty$, then $\mathrm{E}\,[\,Z_\tau\,] = 0$ (see Theorem IV.2).

The upper bound in Theorem II.1 asserts that $\gamma\sqrt{T}/2 \geq \mathrm{Regret}(T) = Z_T + g_T/2$ for any fixed $T \geq 0$. Hence, $\gamma\,\mathrm{E}\,[\,\sqrt{\tau}\,]/2 \geq \mathrm{E}\,[\,g_\tau\,]/2$. Replacing g_τ with $|X_\tau|$ (since both g_t and $|X_t|$ are reflected random walks), the proof of the first assertion is complete.

The fact that no constant smaller than γ is possible is a direct consequence of the results of Breiman [39] and Greenwood and Perkins [40] as mentioned in Subsection II-C (see also Section IV or [20]). ∎

[10]The integrator is reflected Brownian motion, which is almost surely not of bounded variation.

[11]Equality holds because Algorithm 1 sets $p(t,0) = 1/2$.

E. An expression for the regret involving the gap

In our two-expert prediction problem, the most important scenario restricts each cost vector ℓ_t to be either $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ or $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$. That is, at each time step, some expert incurs cost 1 and the other expert incurs no cost. This restricted scenario is equivalent to the condition $g_t - g_{t-1} \in \{\pm 1\}\ \ \forall t \geq 1$, where $g_t \coloneqq |L_{t,1} - L_{t,2}|$ is the gap at time t. To prove the optimal lower bound it suffices to consider this restricted scenario. In this version of the paper, we prove the optimal upper bound assuming the restricted scenario. In the full version of the paper, we extend the analysis to all cost vectors by a concavity argument. In the remainder of the paper, we assume the restricted scenario.

We now present an expression, valid for any algorithm, that emphasizes how the regret depends on the *change* in the gap. This expression will be useful in proving both the upper and lower bounds. Henceforth we write $\mathrm{Regret}(t) \coloneqq \mathrm{Regret}(2, t, \mathcal{A}, \mathcal{B})$ where $\mathcal{A}$ and $\mathcal{B}$ are usually implicit from the context.

Proposition II.3. Assume the restricted setting in which $g_t - g_{t-1} \in \{\pm 1\}$ for every $t \geq 1$. When $g_{t-1} \neq 0$, let p_t denote the probability mass assigned by the algorithm to the "worst expert", i.e., if $L_{t-1,1} \geq L_{t-1,2}$ then $p_t = x_{t,1}$ and otherwise $p_t = x_{t,2}$. The quantity p_t may depend arbitrarily on $\ell_1, \ldots, \ell_{t-1}$. Then

$$\begin{aligned}\mathrm{Regret}(T) \;&=\; \sum_{t=1}^{T} p_t \cdot (g_t - g_{t-1}) \cdot \mathbf{1}[g_{t-1} \neq 0] \\ &\quad + \sum_{t=1}^{T} \langle\, x_t,\, \ell_t \,\rangle \cdot \mathbf{1}[g_{t-1} = 0].\end{aligned} \tag{II.8}$$

Furthermore, assume that if $g_{t-1} = 0$, then $p_t = x_{t,1} = x_{t,2} = 1/2$. In this case

$$\mathrm{Regret}(T) \;=\; \sum_{t=1}^{T} p_t \cdot (g_t - g_{t-1}). \tag{II.9}$$

Remark. Observe that (II.9) is a discrete analog of a Riemann–Stieltjes integral of p with respect to g. If $(g_t)_{t\geq 0}$ is a random process, then (II.9) is called a discrete stochastic integral. In the specific case that $(g_t)_{t\geq 0}$ is a reflected random walk (the absolute value of a standard random walk), then (II.8) is the Doob decomposition [44, Theorem 10.1] of the regret process $\big(\mathrm{Regret}(t)\big)_{t\geq 0}$, i.e., the first sum is a martingale and the second sum is an increasing predictable process.

Proof: Define $\Delta_{\mathrm{R}}(t) = \mathrm{Regret}(t) - \mathrm{Regret}(t-1)$. The total cost of the best expert at time t is $L_t^* \coloneqq \min\{L_{t,1}, L_{t,2}\}$. The change in regret at time t is the cost incurred by the algorithm minus the change in the total cost of the best expert, so $\Delta_{\mathrm{R}}(t) = \langle\, x_t,\, \ell_t \,\rangle - (L_t^* - L_{t-1}^*)$.

Case 1: $g_{t-1} \neq 0$**:** In this case, the best expert at time $t-1$ remains a best expert at time t. If the worst expert incurs cost 1, then the algorithm incurs cost p_t and the best expert incurs cost 0, so $\Delta_{\mathrm{R}}(t) = p_t$ and $g_t - g_{t-1} = 1$. Otherwise, the best expert incurs cost 1 and the algorithm incurs cost $1 - p_t$, so $\Delta_{\mathrm{R}}(t) = -p_t$ and $g_t - g_{t-1} = -1$. For either choice of cost, we have $\Delta_{\mathrm{R}}(t) = p_t \cdot (g_t - g_{t-1})$.

Case 2: $g_{t-1} = 0$**:** Both experts are best, but one incurs no cost, so $L_t^* = L_{t-1}^*$ and $\Delta_{\mathrm{R}}(t) = \langle\, x_t,\, \ell_t \,\rangle$.

The above two cases prove (II.8). For the last assertion, we have that $\langle\, x_t,\, \ell_t \,\rangle = 1/2 = p_t \cdot (g_t - g_{t-1})$ whenever $g_{t-1} = 0$. Hence, we can collapse the two sums in (II.8) into one to get (II.9). ∎

III. Upper bound

In this section, we prove the upper bound in Theorem II.1 via a sequence of simple steps. We remind the reader that for simplicity, we will assume that the gap changes by ± 1 at each step, which corresponds to each loss vector ℓ_t being either $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ or $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$. The analysis can be extended to general loss vectors in $[0,1]^2$ through the use of concavity arguments. The details of this extension can be found in the full version of the paper.

The proof in this section uses the potential function R which, as explained in Subsection II-C, is defined via continuous-time arguments in Section V. Moreover, the structure of the proof is heavily inspired by the proof in the continuous setting. Finally, we remark that the analysis of this section uses the potential function in a modular way[12], and could conceivably be used to analyze other algorithms (e.g., MWU).

Moving forward, we will need a few observations about the functions R and p, which were defined in equations (II.6) and (II.7).

Lemma III.1. For any $t > 0$, $R(t, g)$ is concave and non-decreasing in g.

The proof of Lemma III.1 is a calculus exercise and appears in the full version. As a consequence, we can easily get the maximum value of $R(t, g)$ for any t.

Lemma III.2. For any $t > 0$, $R(t, g) \leq \gamma\sqrt{t}/2$.

Proof: $R(t, g)$ is constant for $g \geq \gamma\sqrt{t}$ and Lemma III.1 shows that $R(t, g)$ is non-decreasing in g. Hence $\max_g R(t, g) \leq R(t, \gamma\sqrt{t}) = \gamma\sqrt{t}/2$. ∎

In the definition of the prediction task, the algorithm must produce a probability vector x_t. Recalling the definition of x_t in Algorithm 1, it is not a priori clear whether x_t is indeed a probability vector. We now verify that it is, since Lemma III.3 implies that $p(t, g) \in [0, 1/2]$ for all t, g.

Lemma III.3. Fix $t \geq 1$. Then

1) $p(t, 0) = 1/2$;
2) $p(t, g)$ is non-increasing in g; and
3) $p(t, g) \geq 0$.

Proof: For the first assertion, we have

$$\begin{aligned} & p(t,0) \\ &= \frac{1}{2}(R(t,1) - R(t,-1)) \\ &= \frac{1}{2}\left(\frac{1}{2} + \kappa\sqrt{t}M_0(1/2t) + \frac{1}{2} - \kappa\sqrt{t}M_0(1/2t)\right) \\ &= \frac{1}{2}. \end{aligned}$$

For the second equality, we used that $1 \leq \gamma \leq \gamma\sqrt{t}$ for all $t \geq 1$. The second assertion follows from concavity of R, which is asserted in Lemma III.1. The final assertion holds because R is non-decreasing in g, which is also asserted in Lemma III.1. ∎

A. Analysis when gap increments are ± 1

In this subsection we prove the upper bound of Theorem II.1 for a restricted class of adversaries (that nevertheless capture the core of the problem).

Theorem III.4. Let $\mathcal{A}$ be the algorithm described in Algorithm 1. For any adversary $\mathcal{B}$ such that each cost vector ℓ_t is either $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ or $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, we have

$$\sup_{t \geq 1} \frac{\mathrm{Regret}(2, t, \mathcal{A}, \mathcal{B})}{\sqrt{t}} \;\leq\; \frac{\gamma}{2}.$$

Our analysis will rely on an identity known as the discrete Itô formula, which is the discrete analogue of Itô's formula from stochastic analysis (see Theorem V.3). To make this connection (in addition to future connections) more apparent, we define the discrete derivatives of a function f to be

$$\begin{aligned} f_g(t,g) &= \frac{f(t,g+1) - f(t,g-1)}{2}, \\ f_t(t,g) &= f(t,g) - f(t-1,g), \\ f_{gg}(t,g) &= \big(f(t,g+1) + f(t,g-1)\big) - 2f(t,g). \end{aligned}$$

It was remarked earlier that $p(t, g)$ is the discrete derivative of R, and this is because

$$p(t,g) \;=\; R_g(t,g). \tag{III.1}$$

Lemma III.5 (Discrete Itô formula). Let $g_0, g_1, \ldots$ be any sequence of real numbers (not necessarily random) satisfying $|g_t - g_{t-1}| = 1$. Then for any function f and

[12] Our analysis may also be viewed as an amortized analysis. With this viewpoint, the algorithm incurs amortized regret at most $\frac{\gamma}{2}(\sqrt{t} - \sqrt{t-1}) \approx \gamma/4\sqrt{t}$ at each time step t.

any fixed time $T \geq 1$, we have

$$f(T,g_T) - f(0,g_0) \;=\; \sum_{t=1}^{T} f_g(t,g_{t-1}) \cdot (g_t - g_{t-1}) + \sum_{t=1}^{T} \left(\frac{1}{2} f_{gg}(t,g_{t-1}) + f_t(t,g_{t-1}) \right). \tag{III.2}$$

This lemma is a small generalization of [44, Example 10.9] to accommodate a bivariate function f that depends on t. The proof is essentially identical and can be found in the full version.

Now we show how the regret has a formula similar to (III.2). Recall that Lemma III.3(1) guarantees $p(t,0) = 1/2$, i.e., $x_t = [1/2, 1/2]$. Hence, (II.9) gives

$$\mathrm{Regret}(T) = \sum_{t=1}^{T} p(t,g_{t-1}) \cdot (g_t - g_{t-1}) \tag{III.3}$$

where $g_0 = 0$ and $g_t \geq 0$ for all $t \geq 1$.

Key technical step: The following is the most non-obvious step of the proof. We will apply discrete Itô to (III.3), taking $f = R$. Since $p = R_g = f_g$, observe that the main difference between (III.2) and (III.3) is the absence of $\frac{1}{2} f_{gg}(t,g_{t-1}) + f_t(t,g_{t-1})$ in (III.3). In the continuous setting, we will see that a key idea is to try to obtain a solution satisfying $(\frac{1}{2}\partial_{gg} + \partial_t) f = 0$; this is the well-known backwards heat equation. In the discrete setting, by a remarkable stroke of luck, we have the following analogous property.

Lemma III.6 (Discrete backwards heat inequality). $\frac{1}{2} R_{gg}(t,g) + R_t(t,g) \geq 0$ for all $t \geq 1$ and $g \geq 0$.

Proof (of Theorem III.4): Apply Lemma III.5 to the function R and the sequence $g_0, g_1, \ldots$ of (integer) gaps produced by the adversary $\mathcal{B}$. Then, for any time $T \geq 0$,

$$\begin{aligned} & R(T,g_T) - R(0,g_0) \\ &= \sum_{t=1}^{T} R_g(t,g_{t-1}) \cdot (g_t - g_{t-1}) \\ &+ \sum_{t=1}^{T} \left(\frac{1}{2} R_{gg}(t,g_{t-1}) + R_t(t,g_{t-1}) \right) \\ &\geq \sum_{t=1}^{T} p(t,g_{t-1}) \cdot (g_t - g_{t-1}) \\ &= \mathrm{Regret}(T) \end{aligned}$$

Here, the first equality is by Lemma III.5, the inequality is by (III.1) and Lemma III.6, the last equality is (III.3). Since $g_0 = 0$ and $R(0,0) = 0$, applying Lemma III.2 shows that $\mathrm{Regret}(T) \leq R(T,g_T) \leq \gamma\sqrt{T}/2$. ∎

The reader at this point may be wondering why γ is the right constant to appear in the analysis. In Section V, we will define the function R specifically to obtain γ in the preceding analysis. In the next section, our matching lower bound will prove that γ is indeed the right constant.

IV. Lower bound

The main result of this section is the following theorem, which implies the lower bound in Theorem II.1.

Theorem IV.1. For any algorithm $\mathcal{A}$ and any $\epsilon > 0$, there exists an adversary $\mathcal{B}_\epsilon$ such that

$$\sup_{t \geq 1} \frac{\mathrm{Regret}(2,t,\mathcal{A},\mathcal{B}_\epsilon)}{\sqrt{t}} \;\geq\; \frac{\gamma - \epsilon}{2}. \tag{IV.1}$$

It is common in the literature for regret lower bounds to be proven by random adversaries (e.g., [11, Theorem 3.7]). We also consider a random adversary, but the novelty is the use of a non-trivial stopping time at which it can be shown that the regret is large.

A random adversary: Suppose an adversary produces a sequence of cost vectors $\ell_1, \ell_2, \ldots \in \{0,1\}^2$ as follows. For all $t \geq 1$,

- If $g_{t-1} > 0$ then ℓ_t is randomly chosen to be one of the vectors $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ or $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, uniformly and independent of $\ell_1, \ldots, \ell_{t-1}$. Thus $g_t - g_{t-1}$ is uniform in $\{\pm 1\}$.
- If $g_{t-1} = 0$ then $\ell_t = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ if $x_{t,1} \geq 1/2$, and $\ell_t = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ if $x_{t,2} > 1/2$. In both cases $g_t = 1$.

As remarked above, the process $(g_t)_{t \geq 0}$ has the same distribution as the absolute value of a standard random walk (which is also known as a reflected random walk).

We now obtain from (II.8) a lower bound on the regret of any algorithm against this adversary. The adversary's behavior when $g_{t-1} = 0$ ensures that $\langle\, x_t,\, \ell_t \,\rangle \geq 1/2$, showing that

$$\mathrm{Regret}(T) \;\geq\; \underbrace{\sum_{t=1}^{T} p_t\,(g_t - g_{t-1}) \cdot \mathbf{1}[g_{t-1} \neq 0]}_{\text{martingale}} + \frac{1}{2} \underbrace{\sum_{t=1}^{T} \mathbf{1}[g_{t-1} = 0]}_{\text{local time}} \qquad \forall T \in \mathbb{N}.$$

(Equality holds if the algorithm sets $x_t = [1/2, 1/2]$ whenever $g_{t-1} = 0$.) The first sum is a martingale indexed by T. (This holds because $g_t - g_{t-1}$ has conditional expectation 0 when $g_{t-1} \neq 0$, and $\mathbf{1}[g_{t-1} \neq 0] = 0$ when $g_{t-1} = 0$.) The second sum is called the local time of the random walk. Using Tanaka's formula [44, Ex. 10.8], the local time can be written as $\sum_{t=1}^{T} \mathbf{1}[g_{t-1} = 0] = g_t - Z'_t$ where Z'_t is

a martingale with uniformly bounded increments and $Z'_0 = 0$. Thus, combining the two martingales, we have

$$\text{Regret}(t) \;\geq\; Z_t + \frac{g_t}{2} \qquad \forall t \in \mathbb{Z}_{\geq 0}, \tag{IV.2}$$

where Z_t is a martingale with uniformly bounded increments and $Z_0 = 0$.

Intuition for a stopping time: Optional stopping theorems assert that, under some hypotheses, the expected value of a martingale at a stopping time equals the value at the start. Using such a theorem, at a stopping time τ it would hold that $\mathrm{E}[\,\text{Regret}(\tau)\,] \geq \mathrm{E}[\,g_\tau\,]/2$ (under some hypotheses on τ and Z). Thus it is natural to design a stopping time τ that maximizes $\mathrm{E}[\,g_\tau\,]$ and satisfies the hypotheses. We know from (II.2) that the optimal anytime regret at time t is $\Theta(\sqrt{t})$, so one reasonable stopping time would be

$$\tau(c) \;:=\; \min\left\{\, t > 0 \,:\, g_t \geq c\sqrt{t} \,\right\}$$

for some constant c yet to be determined. If $\tau(c)$ and Z satisfy the hypotheses of the optional stopping theorem, then it will hold that $\mathrm{E}[\,\text{Regret}(\tau(c))\,] \geq \frac{c}{2}\,\mathrm{E}[\,\sqrt{\tau(c)}\,]$. From this, it follows, fairly easily, that $\text{AnytimeNormRegret}(2) \geq c/2$; this will be argued more carefully later.

An optional stopping theorem: The optional stopping theorems appearing in standard references require one of the following hypotheses: (i) τ is almost surely bounded, or (ii) $\mathrm{E}[\,\tau\,]$ is bounded and the martingale has bounded increments, or (iii) the martingale is almost surely bounded and τ is almost surely finite. See, e.g., [41, Theorem 4.8.5], [44, Theorem 10.11], [45, Theorem II.57.4], or [46, Theorem 10.10]. These will not suffice for our purposes. For example, condition (ii) is the only useful hypothesis for our setting. It is known [39], [42] that $\mathrm{E}[\,\tau(c)\,] < \infty$, with $\tau(c)$ as above, if and only if $c < 1$; this yields a weak lower bound on the regret. Instead, we require the following theorem, which has a weaker hypothesis (due to the square root). We are unable to find a reference for this theorem, although it is presumably folklore. A proof is provided in the full version.

Theorem IV.2. Let Z_t be a martingale and $K > 0$ a constant such that $|Z_t - Z_{t-1}| \leq K$ almost surely for all t. Let τ be a stopping time. If $\mathrm{E}[\,\sqrt{\tau}\,] < \infty$ then $\mathrm{E}[\,Z_\tau\,] = \mathrm{E}[\,Z_0\,]$.

Optimizing the stopping time: Since the martingale Z_t defined above has bounded increments, Theorem IV.2 may be applied so long as $\mathrm{E}[\,\sqrt{\tau(c)}\,] < \infty$, in which case the preceding discussion yields $\text{AnytimeNormRegret}(2) \geq c/2$. We reiterate that the condition $\mathrm{E}[\,\sqrt{\tau(c)}\,] < \infty$ is a stronger assumption than $\tau(c)$ being almost surely finite. So it remains to determine

$$\sup\{\, c \geq 0 \,:\, \mathrm{E}[\,\sqrt{\tau(c)}\,] < \infty \,\}, \tag{IV.3}$$

where $\tau(c)$ is the first time at which a standard random walk crosses the two-sided boundary $\pm c\sqrt{t}$. We will use the following result, in which M is the confluent hypergeometric function defined as follows. For $a, b \in \mathbb{R}$ with $b \notin \mathbb{Z}_{\leq 0}$,

$$M(a, b, z) = \sum_{n=0}^{\infty} \frac{(a)_n z^n}{(b)_n n!}, \tag{IV.4}$$

where $(x)_n := \prod_{i=0}^{n-1}(x+i)$ is the Pochhammer symbol. We note that $M_0(x) = M(-1/2, 1/2, x)$ (see the appendix in the full version).

Theorem IV.3 (Breiman [39], Theorem 2). Let $c > 1$ and $a < 0$ be such that c is the smallest positive root of the function $x \mapsto M(a, 1/2, x^2/2)$. Then there exists a constant K such that $\Pr[\,\tau(c) > u\,] \sim K u^a$.[13]

Recall the definition of γ in (II.4). For intuition, let us apply Theorem IV.3 with $c = \gamma$, which is defined so that it is the (unique) root for $a = -1/2$ (see Eq. (II.4) and recall $M_0(x) = M(-1/2, 1/2, x)$). It follows that

$$\begin{aligned}
\mathrm{E}\left[\,\sqrt{\tau(\gamma)}\,\right] \;&=\; \int_0^\infty \Pr\left[\,\sqrt{\tau(\gamma)} > s\,\right] \mathrm{d}s \\
&=\; \int_0^\infty \Pr\left[\,\tau(\gamma) > s^2\,\right] \mathrm{d}s \\
&\sim\; K \int_0^\infty s^{-1}\,\mathrm{d}s,
\end{aligned}$$

by Theorem IV.3. This integral is infinite, so Theorem IV.2 cannot be applied to $\tau(\gamma)$. However, the integral is on the cusp of being finite. By slightly decreasing a below $-1/2$, and slightly modifying c to be the new root, we should obtain a finite integral, showing that $\mathrm{E}[\,\sqrt{\tau(c)}\,]$ is finite.

Proof (of Theorem IV.1): Fix any $\epsilon > 0$ that is sufficiently small. Consider the random adversary and the stopping times $\tau(c)$ described above. In the full version, we show that there exists $a_\epsilon \in (-1, -1/2)$ and $c_\epsilon \geq \gamma - \epsilon$ such that c_ϵ is the unique positive root of $z \mapsto M(a_\epsilon, 1/2, z^2/2)$. As in the above calculations, Theorem IV.3 shows that

$$\begin{aligned}
\mathrm{E}\left[\,\sqrt{\tau(c_\epsilon)}\,\right] \;&=\; \int_0^\infty \Pr\left[\,\tau(c_\epsilon) > s^2\,\right] \mathrm{d}s \\
&\sim\; K \int_0^\infty s^{2a_\epsilon}\,\mathrm{d}s \;<\; \infty,
\end{aligned} \tag{IV.5}$$

since $a_\epsilon < -1/2$. It follows that $\tau(c_\epsilon)$ is almost surely finite, and therefore $\text{Regret}(\tau(c_\epsilon))$ and $g_{\tau(c_\epsilon)}$

[13] This means that $\lim_{u\to\infty} \frac{\Pr[\,\tau(c) > u\,]}{K u^a} = 1$.

are almost surely well defined. Applying Theorem IV.2 to the martingale Z_t appearing in (IV.2), we obtain that

$$\mathrm{E}\left[\,\mathrm{Regret}(\tau(c_\epsilon))\,\right] \;\geq\; \frac{1}{2}\,\mathrm{E}\left[\,g_{\tau(c_\epsilon)}\,\right] \;=\; \frac{1}{2}\,\mathrm{E}\left[\,c_\epsilon\sqrt{\tau(c_\epsilon)}\,\right].$$

By the probabilistic method, there exists a finite sequence of cost vectors $\ell_1, \ldots, \ell_t$ (depending on $\mathcal{A}$ and ϵ) for which the regret of $\mathcal{A}$ at time t is at least $c_\epsilon\sqrt{t}/2$. The adversary $\mathcal{B}_\epsilon$ (which knows $\mathcal{A}$) provides this sequence of cost vectors to algorithm $\mathcal{A}$, thereby proving (IV.1). ∎

V. A continuous-time analogue of Algorithm 1

This section sketches how the potential function R defined in (II.6) arises naturally as a solution of a stochastic calculus problem.

A. Defining the continuous regret problem

Continuous time regret problem: The continuous regret problem is inspired by (II.9). Notice that, when the adversary chooses cost vectors in $\{\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\}$, the sequence of gaps $g_0, g_1, g_2, \ldots$ live in the support of a reflected random walk. The goal in the discrete case is to find an algorithm p that bounds the regret over all possible sample paths of a reflected random walk. In continuous time it is natural to consider a stochastic integral with respect to reflected Brownian motion, denoted $|B_t|$, instead. Our goal now is to find a continuous-time algorithm whose regret is small for almost all reflected Brownian motion paths.

Definition V.1 (Continuous Regret). Let $p\colon \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \to [0,1]$ be a continuous function that satisfies $p(t,0) = 1/2$ for every $t > 0$. Let B_t be a standard one-dimensional Brownian motion. Then, the *continuous regret* of p with respect to B is the stochastic integral

$$\mathrm{ContRegret}(T, p, B) \;=\; \int_0^T p(t, |B_t|)\,\mathrm{d}\,|B_t|\,. \quad \text{(V.1)}$$

In this definition we may think of p as a continuous-time algorithm and B as a continuous-time adversary. In this section, we sketch the following result.

Theorem V.2. There exists a continuous-time algorithm p^* such that, almost surely,

$$\mathrm{ContRegret}(T, p^*, B) \;\leq\; \frac{\gamma\sqrt{T}}{2} \qquad \forall T \in \mathbb{R}_{\geq 0}. \quad \text{(V.2)}$$

B. Itô's formula and the backwards heat equation

Since $\mathrm{ContRegret}(T)$ evolves as a stochastic integral with respect to a semi-martingale[14] (namely reflected Brownian motion), Itô's lemma provides an insightful decomposition. The following statement of Itô's lemma is a specialization of [47, Theorem IV.3.3] for the special case of reflected Brownian motion.

[14] A semi-martingale is a stochastic process that can written as the sum of a local martingale and a process of finite variation.

Notation: Until now, we have used g as the second parameter to p and R to denote the gap. Henceforth, to be more consistent with the usual notation in the literature, we use x instead of g. We use $C^{1,2}$ to denote the class of bivariate functions that are continuously differentiable in their first argument and twice continuously differentiable in their second argument.

Theorem V.3 (Itô's formula). Let $f\colon \mathbb{R}_{\geq 0} \times \mathbb{R} \to \mathbb{R}$ be $C^{1,2}$. Then, almost surely,

$$\begin{aligned} f(T, |B_T|) - f(0, |B_0|) \;&=\; \int_0^T \partial_x f(t, |B_t|)\,\mathrm{d}\,|B_t| \\ &+\; \int_0^T \Big[\underbrace{\partial_t f(t, |B_t|) + \tfrac{1}{2}\partial_{xx} f(t, |B_t|)}_{=:\overset{*}{\Delta} f(t, |B_t|)}\Big]\,\mathrm{d}t. \end{aligned} \quad \text{(V.3)}$$

The integrand of the second integral is an important quantity arising in PDEs and stochastic processes (see, e.g., [48, pp. 263]). We denote it by $\overset{*}{\Delta} f(t,x) := \partial_t f(t,x) + \frac{1}{2}\partial_{xx} f(t,x)$.

Applying Itô's formula to the continuous regret: By pattern matching (V.1) and (V.3), it is natural to assume that $p = \partial_x f$ for a function f that is $C^{1,2}$ with $f(0,0) = 0$, $\partial_x f \in [0,1]$, and $\partial_x f(t,0) = 1/2$; the latter two conditions are needed for Definition V.1 to be applicable. Itô's formula then yields

$$\begin{aligned} \mathrm{ContRegret}(T, p = \partial_x f, B) \;&=\; \int_0^T \partial_x f(t, |B_t|)\,\mathrm{d}\,|B_t| \\ &=\; f(T, |B_T|) - \int_0^T \overset{*}{\Delta} f(t, |B_t|)\,\mathrm{d}t. \end{aligned} \quad \text{(V.4)}$$

At this point a useful idea arises: as a thought experiment, suppose that $\overset{*}{\Delta} f = 0$. Then the second integral vanishes, and we have the appealing expression $\mathrm{ContRegret}(T, p, B) = f(T, |B_T|)$. Moreover, since f is a deterministic function, the right-hand side depends only on $|B_T|$ rather than the entire Brownian path $B|_{[0,T]}$. Thus, the same must be true of the left-hand side; in other words, the algorithm has *path independent regret*. Our supposition that led to these attractive consequences was only that $\overset{*}{\Delta} f = 0$, which turns out to be a well studied condition.

Definition V.4. Let $f\colon \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{R}$ be a $C^{1,2}$ function. If $\overset{*}{\Delta} f(t,x) = 0$ for all $(t,x) \in \mathbb{R}_{>0} \times \mathbb{R}$ then we say that f satisfies the *backward heat equation*. A synonymous statement is that f is *space-time harmonic*.

The following proposition summarizes the preceding discussion.

Proposition V.5. Let $f : \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{R}$ be a $C^{1,2}$ function that satisfies $\overset{*}{\Delta} f = 0$ everywhere with $f(0,0) = 0$. Let $p = \partial_x f$. Then,

$$\int_0^T p(t, |B_t|) \, \mathrm{d}\, |B_t| \;=\; f(T, |B_T|). \tag{V.5}$$

Given the PDE $\overset{*}{\Delta} f = 0$, the remaining task to define an algorithm is to prescribe boundary conditions. One natural boundary condition comes from comparing (V.4) with Definition V.1: we require $p(t,0) = \partial_x f(t,0) = 1/2$ for all $t > 0$. The second boundary condition is less obvious but intuitively, one may consider putting zero mass on the worst expert if the gap is significantly large. Quantitatively, we add the boundary condition $p(t, \alpha\sqrt{t}) = \partial_x f(t, \alpha\sqrt{t}) = 0$ where $\alpha > 0$ is a parameter to be optimized. This states that if the gap between the two experts is larger than $\alpha\sqrt{t}$ at time t then the algorithm places no mass on the worst expert at time t.

In the full version of the paper, we show that the optimal choice of α is precisely γ (recall (II.4) for the definition of γ). Solving the PDE yields the strategy

$$p(t,x) = \frac{1}{2} \cdot \left(1 - \frac{\operatorname{erfi}(x/\sqrt{2t})}{\operatorname{erfi}(\gamma/\sqrt{2})}\right)$$

and the potential function

$$f(t,x) = \frac{x}{2} + \kappa\sqrt{t} \cdot M_0\left(\frac{x^2}{2t}\right)$$

where κ is as defined in (II.6). Proposition V.5 then asserts that $\int_0^T \partial_x f(t, |B_t|) \, \mathrm{d}|B_t| = f(T, |B_T|)$ almost surely. It is a straightforward exercise to show that, for any fixed $t > 0$, $f(t,x)$ is maximized at $x = \gamma\sqrt{t}$; hence $f(T, |B_T|) \le \gamma\sqrt{T}/2$. At this point, there is a minor snag: $p = \partial_x f$ may be negative so it may not define a valid algorithm (see Definition V.1). This is resolved by considering the algorithm $\max\{p, 0\}$, which is non-negative; in fact, $\max\{p,0\} = \partial_x R$ where R is as defined in (II.6). In the full version of the paper, we show that using $\max\{p,0\}$ for the algorithm allows us to prove Theorem V.2.

Acknowledgment

The authors thank Hu Fu, Bruce Shepherd, Taylor Lundy, Victor Portella, and Paul Liu for valuable discussions and feedback. The authors also thank the anonymous reviewers for valuable feedback.

References

[1] J. Hannan, "Approximation to Bayes risk in repeated play," *Contributions to the Theory of Games*, vol. 3, pp. 97–139, 1957.

[2] S. Arora, E. Hazan, and S. Kale, "The multiplicative weights update method: a meta-algorithm and applications," *Theory of Computing*, vol. 8, no. 1, 2012.

[3] N. Littlestone and M. K. Warmuth, "The weighted majority algorithm," *Information and computation*, vol. 108, no. 2, pp. 212–261, 1994.

[4] V. G. Vovk, "Aggregating strategies," *Proc. of Computational Learning Theory, 1990*, 1990.

[5] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire, and M. K. Warmuth, "How to use expert advice," *Journal of the ACM (JACM)*, vol. 44, no. 3, pp. 427–485, 1997.

[6] N. Cesa-Bianchi, "Analysis of two gradient-based algorithms for on-line regression," *Journal of Computer and System Sciences*, vol. 59, no. 3, pp. 392–411, 1999.

[7] N. Gravin, Y. Peres, and B. Sivan, "Tight Lower Bounds for Multiplicative Weights Algorithmic Families," in *Proceedings of ICALP 2017*, vol. 80, 2017, pp. 48:1–48:14.

[8] T. M. Cover, "Behavior of sequential predictors of binary sequences," in *Proceedings of the 4th Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*. Publishing House of the Czechoslovak Academy of Sciences, Prague, 1965.

[9] H. Luo and R. E. Schapire, "Towards minimax online learning with unknown time horizon," in *Proceedings of ICML*, 2014.

[10] N. Gravin, Y. Peres, and B. Sivan, "Towards optimal algorithms for prediction with expert advice," in *Proceedings of SODA*. SIAM, 2016, pp. 528–547.

[11] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge University Press, 2006.

[12] Y. Nesterov, "Primal-dual subgradient methods for convex problems," *Mathematical Programming*, vol. 120, no. 1, pp. 221–259, 2009.

[13] S. Bubeck, "Introduction to online optimization," December 2011, unpublished.

[14] H. Fang, N. J. A. Harvey, V. S. Portella, and M. P. Friedlander, "Online mirror descent and dual averaging: keeping pace in the dynamic case," in *Proceedings of ICML*, 2020.

[15] S. Gerchinovitz, "Prediction of individual sequences and prediction in the statistical framework: some links around sparse regression and aggregation techniques," Ph.D. dissertation, Université Paris-Sud, 2011.

[16] E. Perkins, "On the Hausdorff dimension of the Brownian slow points," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 64, pp. 369–399, 1983.

[17] H. Luo and R. E. Schapire, "Achieving all with no parameters: AdaNormalHedge," in *Proceedings of The 28th Conference on Learning Theory*, vol. 40, 2015, pp. 1286–1304.

[18] ——, "A drifting-games analysis for online learning and applications to boosting," in *Advances in Neural Information Processing Systems*, 2014, pp. 1368–1376.

[19] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*. Addison-Wesley, 1994.

[20] B. Davis, "On the L_p norms of stochastic integrals and other martingales," *Duke Math. J*, vol. 43, no. 4, pp. 697–704, 1976.

[21] Y. Abbasi-Yadkori, P. L. Bartlett, and V. Gabillon, "Near minimax optimal players for the finite-time 3-expert prediction problem," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 3033–3042.

[22] E. Bayraktar, I. Ekren, and X. Zhang, "Finite-time 4-expert prediction problem," *Communications in Partial Differential Equations*, pp. 1–44, 2020.

[23] S. Abbasi-Zadeh, N. Bansal, G. Guruganesh, A. Nikolov, R. Schwartz, and M. Singh, "Sticky Brownian rounding and its applications to constraint satisfaction problems," *arXiv preprint arXiv:1812.07769*, 2018.

[24] S. Bubeck, M. B. Cohen, Y. T. Lee, J. R. Lee, and A. Mądry, "k-server via multiscale entropic regularization," in *Proceedings of STOC*. ACM, 2018.

[25] S. Bubeck, R. Eldan, and J. Lehec, "Sampling from a log-concave distribution with projected Langevin Monte Carlo," *Discrete Comput. Geom.*, vol. 59, no. 4, Jun. 2018.

[26] G. Calinescu, C. Chekuri, M. Pál, and J. Vondrák, "Maximizing a monotone submodular function subject to a matroid constraint," *SIAM Journal on Computing*, vol. 40, no. 6, pp. 1740–1766, 2011.

[27] C. Chekuri, T. Jayram, and J. Vondrák, "On multiplicative weight updates for concave and submodular function maximization," in *Proceedings of the Conference on Innovations in Theoretical Computer Science (ITCS)*, 2015, pp. 201–210.

[28] J. Diakonikolas and L. Orecchia, "The approximate duality gap technique: A unified theory of first-order methods," *SIAM Journal on Optimization*, vol. 29, no. 1, pp. 660–689, 2019.

[29] R. Eldan and A. Naor, "Krivine diffusions attain the Goemans–Williamson approximation ratio," *arXiv preprint arXiv:1906.10615*, 2019.

[30] E. Bayraktar, I. Ekren, and Y. Zhang, "On the asymptotic optimality of the comb strategy for prediction with expert advice," *arXiv preprint arXiv:1902.02368*, 2019.

[31] N. Drenska, "A PDE approach to a prediction problem involving randomized strategies," Ph.D. dissertation, New York University, 2017.

[32] N. Drenska and R. V. Kohn, "Prediction with expert advice: A PDE perspective," *Journal of Nonlinear Science*, vol. 30, no. 1, pp. 137–173, 2020.

[33] V. A. Kobzar, R. V. Kohn, and Z. Wang, "New potential-based bounds for prediction with expert advice," *arXiv preprint arXiv:1911.01641*, 2019.

[34] ——, "New potential-based bounds for the geometric-stopping version of prediction with expert advice," *arXiv preprint arXiv:1912.03132*, 2019.

[35] P. M. DeMarzo, I. Kremer, and Y. Mansour, "Online trading algorithms and robust option pricing," in *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*. ACM, 2006, pp. 477–486.

[36] J. D. Abernethy, R. M. Frongillo, and A. Wibisono, "Minimax option pricing meets black-scholes in the limit," in *Proceedings of the 44th Symposium on Theory of Computing Conference*. ACM, 2012, pp. 1029–1040.

[37] M. Brezzi and T. L. Lai, "Optimal learning and experimentation in bandit problems," *Journal of Economic Dynamics and Control*, vol. 27, no. 1, pp. 87–108, 2002.

[38] E. Hazan, "Introduction to online convex optimization," *Foundations and Trends in Optimization*, vol. 2, no. 3-4, pp. 157–325, 2016.

[39] L. Breiman, "First exit times for a square root boundary," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory, Part 2*. University of California Press, 1967, pp. 9–16.

[40] P. Greenwood and E. Perkins, "A conditioned limit theorem for random walk and brownian local time on square root boundaries," *Annals of Probability*, vol. 11, pp. 227–261, 1983.

[41] R. Durrett, *Probability: Theory and Examples*, 5th ed. Cambridge University Press, 2019.

[42] L. A. Shepp, "A first passage problem for the Wiener process," *The Annals of Mathematical Statistics*, vol. 38, no. 6, pp. 1912–1914, 1967.

[43] K. Chaudhuri, Y. Freund, and D. J. Hsu, "A parameter-free hedging algorithm," in *Advances in Neural Information Processing Systems 22*, 2009, pp. 297–305.

[44] A. Klenke, *Probability Theory: A Comprehensive Course*. Springer, 2008.

[45] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales. Volume 1: Foundations*, 2nd ed. Cambridge University Press, 2000.

[46] D. Williams, *Probability with Martingales*. Cambridge University Press, 1991.

[47] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*. Springer Science & Business Media, 2013.

[48] J. L. Doob, *Classical Potential Theory and Its Probabilistic Counterparts*. Springer-Verlag, 1984.