

Optimal anytime regret with two experts

Nicholas J. A. Harvey, Christopher Liaw, Edwin Perkins, and
Sikander Randhawa

Abstract. We consider the classical problem of prediction with expert advice. In the fixed-time setting, where the time horizon is known in advance, algorithms that achieve the optimal regret are known when there are two, three, or four experts or when the number of experts is large. Much less is known about the problem in the anytime setting, where the time horizon is *not* known in advance. No minimax optimal algorithm was previously known in the anytime setting, regardless of the number of experts. Even for the case of two experts, Luo and Schapire have left open the problem of determining the optimal algorithm.

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 analog of the regret problem, which is solved using ideas from stochastic calculus.

1. Introduction

We study the problem of prediction with expert advice, whose origin can be traced back to the 1950s [35]. (The problem of prediction with expert advice is sometimes worded slightly differently than we do but these formulations are essentially equivalent. For more details, see the discussion in Appendix D.) 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 loss (or cost) in $[0, 1]$ for each expert, with knowledge of the algorithm’s distribution but not its random choice. The loss of each expert is then revealed to the algorithm, and the algorithm incurs the loss 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 loss

and the total loss of the best expert. This problem and its variants have been a key component in numerous results; we refer the reader to [5].

The most well-known algorithm for the experts problem is the celebrated multiplicative weights update algorithm (MWU) [42, 54]. 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 [12, 13]. This bound on the regret of MWU is known to be tight for any even n [32]. It is also known that $\sqrt{(T/2) \ln n}$ is the regret of the best possible algorithm, asymptotically for large n and T . (A precise statement may be found in the references [13, Corollary 3.2.2] [14, Theorem 3.7].) Interestingly, MWU is *not* optimal for small values of n . For $n = 2$, Cover [18] observed decades earlier that a natural dynamic programming formulation of the problem leads to a simple analysis showing that the minimax optimal regret is $\sqrt{T/2\pi}$, asymptotically for large T (a proof of this can also be found in [13, §3], [37, Theorem 18.5.5]).

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*,¹ 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 [43]. For example, the *geometric horizon setting* of Gravin, Peres, and Sivan [31] assumes that T is a geometric random variable. In this setting, they gave the optimal algorithm for two and three experts. Moreover, they 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” [13, §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 $t \geq 1$. On the one hand, this is a conceptually simple and generic reduction. On the other hand, restarting the algorithm and discarding its state is clearly wasteful and probably not very practical.

Instead of using the doubling trick, one can use variants of MWU with a dynamic step size; see, e.g., [14, §2.3], [45, Theorem 1], [11, §2.5]. This is a much more elegant and practical approach and is even simpler to implement. However, the analysis is more subtle than for MWU with a fixed step size. It is known that, with an appro-

¹Other authors have referred to this setting as an “unknown time horizon” or “bounds that hold uniformly over time”.

priate choice of step sizes, MWU can guarantee² a regret of $\sqrt{t \ln n}$ for all $t \geq 1$ and all $n \geq 2$ (see [11, Theorem 2.4] or [29, Proposition 2.1]). However, it is unknown whether $\sqrt{t \ln n}$ is the minimax optimal anytime regret, for any value of n . Indeed, Luo and Schapire [43] have also stated that “finding the minimax solution to this setting seems to be quite challenging, even for the simplest case of $n = 2$ ”.

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 [46]. (Note that $\gamma/2 \approx 0.653 < 0.833 \approx \sqrt{\ln 2}$.) This also answers a question that has been left open by Luo and Schapire [43]. 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 [31]) and the optimal lower bound for $n \rightarrow \infty$ 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 analog 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 note that Freund [27] has used stochastic differential equations for a continuous-time formulation of the experts problem, although he did not discuss the discrete-time problem.

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 [30, §5.5].

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 $E[|X_\tau|] \leq \gamma E[\sqrt{\tau}]$ for every stopping time τ ; moreover, the constant γ cannot be improved.³ This result is originally due to Davis [19, 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 Section 2.4.

²It can be shown, by modifying arguments of [32], that this is the optimal anytime analysis for MWU with step sizes c/\sqrt{t} .

³At 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 $E[\sqrt{\tau}] = \infty$ (despite τ being a.s. finite) and the inequality is trivial.

Related work. The minimax regret for the experts problem has been well studied in the fixed-time horizon setting. As mentioned above, some tight asymptotics of the minimax regret were known decades ago: for $n = 2$, it is $\sqrt{T/2\pi}$ (see [18]), whereas asymptotically in n , it is $\sqrt{T \ln(n)/2}$ (see [12, 13]). Recent work, building on results of Gravin et al. [31], shows that the minimax regret is $\sqrt{8T/9\pi}$ for $n = 3$ (see [1]) and $\sqrt{\pi T/8}$ for $n = 4$ (see [6]). The anytime setting is not as well understood. In the two-experts setting, Luo and Schapire [43] 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 $\sqrt{T/2\pi}$ in the fixed-time setting, which establishes that the adversary has more power to cause regret in the anytime setting.

Recently, there has been a line of work that makes connections between the experts problem (in the finite-time horizon and geometric-time horizon setting) and PDEs [4, 6, 7, 23, 24, 39, 40]. There is also work connecting regret minimization to option pricing [20] and to the Black–Scholes formula [2], which is based on Brownian motion and stochastic calculus. Intuitively, stochastic calculus is a crucial tool to optimally hedge against future losses, which we exploit too.

Our algorithm chooses the distribution on the experts using the discrete derivative of a potential function. This idea has also been used in the AdaNormalHedge algorithm [44], although their potential function was not derived in continuous time.

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., [10]).

2. Discussion of results and techniques

2.1. Formal problem statement

We will formulate the problem in the style of online convex optimization [52], in which at each time step a deterministic algorithm picks a distribution on experts. An alternative formulation would be to have a randomized algorithm pick a single expert; see, e.g., [14, Chapter 4]. Using the randomized formulation in the anytime setting has certain subtleties which we discuss in Section 2.1.1.

Let n denote the number of experts. There is a deterministic algorithm \mathcal{A} which we think of as an infinite sequence of functions $\mathcal{A} = (A_1, A_2, \dots)$, where

$$A_t: ([0, 1]^n)^{t-1} \times \Delta_n^{t-1} \rightarrow \Delta_n.$$

Here, $\Delta_n = \{x \in \mathbb{R}_{\geq 0}^n : \sum_{i=1}^n x_i = 1\}$ is the set of probability distribution in \mathbb{R}^n . In addition, we have a deterministic adversary \mathcal{B} which we think of as an infinite

sequence of loss vectors in $[0, 1]^n$, i.e., $\mathcal{B} = (\ell_1, \ell_2, \dots) \in ([0, 1]^n)^{\mathbb{N}}$, where $\ell_t \in [0, 1]^n$ for each $t \geq 1$. We refer to the coordinate $\ell_{t,j}$ as the loss of the j th expert at time t . Next, given \mathcal{A} and \mathcal{B} , we define $x_1 = A_1$ and for $t \geq 1$, we recursively define

$$x_t = A_t(\ell_1, \dots, \ell_{t-1}, x_1, \dots, x_{t-1}).$$

Further, for $T \geq 1$, we write $\mathcal{A}_T = (A_1, \dots, A_T)$ and $\mathcal{B}_T = (\ell_1, \dots, \ell_T)$ for the algorithm and the adversary restricted to the first T steps.

The loss 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 loss” of the algorithm, although the algorithm is actually deterministic. The total expected loss of the algorithm up to time t is $\sum_{i=1}^t \langle x_i, \ell_i \rangle$. For $j \in [n]$, the total loss 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 loss and the total loss of the best expert, i.e.,

$$\text{Regret}(n, t, \mathcal{A}, \mathcal{B}) = \sum_{i=1}^t \langle x_i, \ell_i \rangle - \min_{j \in [n]} L_{t,j}. \quad (2.1)$$

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} . The formal definition is somewhat different: the algorithm knows that it will run for at most T steps, but then we take the limit $T \rightarrow \infty$. For a fixed time T , we define the anytime normalized regret up to time T as follows.

$$\text{AnytimeNormRegret}(n, T) := \inf_{\mathcal{A}_T} \sup_{\mathcal{B}_T} \max_{1 \leq t \leq T} \frac{\text{Regret}(n, t, \mathcal{A}_T, \mathcal{B}_T)}{\sqrt{t}}. \quad (2.2)$$

This definition uses a single infimum over a sequence of functions, corresponding to the algorithm (which knows T), and a single supremum over a sequence of loss vectors, corresponding to the adversary (which knows T and \mathcal{A}_T). It may be more intuitive to think of a sequence of successive minimizations and maximizations over the algorithm’s actions and adversary’s actions. This leads to the following definition which, as shown in Appendix C, is equivalent:

$$\begin{aligned} & \text{AnytimeNormRegret}(n, T) \\ & := \min_{x_1} \max_{\ell_1} \cdots \min_{x_T} \max_{\ell_T} \max_{1 \leq t \leq T} \frac{1}{\sqrt{t}} \left(\sum_{i=1}^t \langle x_i, \ell_i \rangle - \min_{j \in [n]} \sum_{i=1}^t \ell_{i,j} \right). \end{aligned} \quad (2.3)$$

In (2.3), the minimization is over all vectors from the probability simplex Δ_n and the maximization is over $[0, 1]^n$. The anytime normalized regret is defined to equal

$$\text{AnytimeNormRegret}(n) := \lim_{T \rightarrow \infty} \text{AnytimeNormRegret}(n, T).$$

This limit exists because (2.2) is clearly non-decreasing in T , and it is finite due to known algorithms such as MWU.

The value of $\text{AnytimeNormRegret}(n)$ was previously unknown even for $n = 2$. The best known bounds prior to our work were

$$0.564 \approx \sqrt{1/\pi} \leq \text{AnytimeNormRegret}(2) \leq \sqrt{\ln 2} \approx 0.833. \quad (2.4)$$

The lower bound, due to [43], demonstrates a gap between the anytime setting and the fixed-time setting, where the optimal normalized regret is $\sqrt{1/2\pi}$ [18]. Our main result is that $\text{AnytimeNormRegret}(2) = \gamma/2 \approx 0.653$ and consequently neither inequality in (2.4) is tight.

As mentioned above, MWU with a dynamic step size shows that

$$\text{AnytimeNormRegret}(n) \leq \sqrt{\ln n}$$

for all $n \geq 2$ (see [11, §2.5]). The lower bound

$$\liminf_{n \rightarrow \infty} \text{AnytimeNormRegret}(n)/\sqrt{\ln n} \geq \sqrt{1/2}$$

follows from the bound in the fixed-time setting [13]. Thus, the upper bound is loose by at most a factor $\sqrt{2}$.

2.1.1. Randomized formulations. Several alternative formulations of the problem arise if \mathcal{A} selects a single expert $I_t \in [n]$ randomly at each time t according to a probability vector $x_t \in \mathbb{R}^n$, and the adversary chooses an ending time τ . We mention three possibilities, differing in the power of the adversary \mathcal{B} . For this subsection alone, we define

$$\text{Regret}(n, t, \mathcal{A}, \mathcal{B}) = \sum_{s=1}^t \ell_{I_s} - \min_{j \in [n]} \sum_{s=1}^t \ell_{s,j},$$

which is the *realized* regret of the algorithm.

- An excessively powerful adversary would allow ℓ_t to depend on I_t . This definition is of little interest, and is mentioned primarily for contrast with the next two possibilities. The trouble is that one easily obtains an adversary \mathcal{B} with

$$\text{Regret}(n, t, \mathcal{A}, \mathcal{B}) = \Omega(t)$$

by setting $\ell_{t,I_t} = 1$ and $\ell_{t,i} = 0$ for $i \neq I_t$. In this case, the algorithm’s realized loss is t while there exists an expert with total loss at most t/n . Hence, the algorithm’s realized regret is at least $(1 - 1/n)t$.

- An adversary of intermediate power allows the loss vector ℓ_t and the event $\tau = t$ to be determined by I_1, \dots, I_{t-1} . This is analogous to the “non-oblivious opponent”

of [14, §4.1]. Interestingly, one can design an adversary \mathcal{B} for which

$$\mathbb{E} \left[\frac{\text{Regret}(n, \tau, \mathcal{A}, \mathcal{B})}{\sqrt{\tau \log \log \tau}} \right] = \Omega(1).$$

The surprising aspect is the $\sqrt{\log \log \tau}$ in the denominator, which arises due to the law of the iterated logarithm. A proof of this statement can be found in our technical report [36].

- The weakest adversary requires that ℓ_t and τ depend only on \mathcal{A} and not its random choices I_1, I_2, \dots . This is analogous to the “oblivious opponent” of [14, §4.1], which we discuss further in Appendix D. The expected regret in this model is identical to the regret in the deterministic model described at the start of Section 2.1. Indeed, in this model, the expectation of the realized regret is

$$\mathbb{E}[\text{Regret}(n, t, \mathcal{A}, \mathcal{B})] = \sum_{s=1}^t \mathbb{E}[\ell_{I_s}] - \min_{j \in [n]} \sum_{s=1}^t \ell_{s,j} = \sum_{s=1}^t \langle x_s, \ell_s \rangle - \min_{j \in [n]} \sum_{s=1}^t \ell_{s,j},$$

which is precisely (2.1).

In this paper, we study this third model. It is intriguing that in the anytime setting, the non-oblivious opponent has more power than the oblivious opponent. In contrast, the two adversaries have the same power in the fixed time setting [14, §4.1].

2.2. Statement of results

To state our results, we require two definitions:

$$\begin{aligned} \text{erfi}(x) &= \frac{2}{\sqrt{\pi}} \int_0^x e^{z^2} dz, \\ M_0(x) &= e^x - \sqrt{\pi x} \text{erfi}(\sqrt{x}). \end{aligned} \tag{2.5}$$

The first is the imaginary error function, a well-known special function that relates to the Gaussian error function. The second is an example of a confluent hypergeometric function, a very broad class of special functions that includes, e.g., Bessel functions and Laguerre polynomials. (See Section 2.6 for formal definitions.) Our analysis makes use of a few elementary properties of these functions. A key constant used in this paper is γ , which is defined to be the smallest⁴ positive root⁵ of $M_0(x^2/2)$, i.e.,

$$\gamma := \min\{x > 0 : M_0(x^2/2) = 0\} \approx 1.3069 \dots \tag{2.6}$$

⁴In fact, γ is the *unique* positive root; see Fact 2.7.

⁵The *roots* of certain confluent hypergeometric functions have appeared in studying some natural phenomena of Brownian motion; for some examples, see [8, 19, 33, 46].

Theorem 2.1 (Main result). *In the anytime setting with two experts, the minimax optimal normalized regret (over deterministic algorithms \mathcal{A} and adversaries \mathcal{B}) is*

$$\text{AnytimeNormRegret}(2) = \lim_{T \rightarrow \infty} \inf_{\mathcal{A}_T} \sup_{\mathcal{B}_T} \max_{1 \leq t \leq T} \frac{\text{Regret}(2, t, \mathcal{A}_T, \mathcal{B}_T)}{\sqrt{t}} = \frac{\gamma}{2}. \quad (2.7)$$

The proof of this theorem has two parts: an upper bound, in Section 3, which exhibits an optimal algorithm, and a lower bound, in Section 4, which exhibits an optimal randomized adversary. The algorithm is very short, and it appears below in Algorithm 1. We also note that the statement allows for the algorithm to depend on T . However, the algorithm we design to achieve the upper bound does not depend on T . On the other hand, our lower bound shows that even if the algorithm was provided with T beforehand, it is unable to achieve regret better than $\gamma/2$.

One might imagine that some form of duality theory is involved in our matching upper and lower bounds. Indeed, if the losses are in $\{0, 1\}$ one may write

$$\text{AnytimeNormRegret}(2)$$

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 (4.3), whereas γ arises in our upper bound as the minimizer in (5.13). We are not aware of any direct relationship between those two equations. 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. [31] in the fixed-time and geometric horizon settings using von Neumann’s minimax theorem (see also [37, §18.5]). 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 (2.7). Our results of Section 4 may be viewed as using random walks to construct feasible dual solutions of value $\gamma/2 - \varepsilon$ for all $\varepsilon > 0$, 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 3 and 5.

A conjecture for n experts. We suspect that hypergeometric functions may also have a key role to play in designing optimal algorithms when there are $n > 2$ experts. As we will see in Section 3, functions involving $M_0(x^2/2)$ have properties that are useful when employed as potential functions for online learning algorithms. It is known

that $M_0(x^2/2) \approx -e^{-x^2/2}/x^2$, and therefore the equation $M_0(x^2/2) = -n$ has the solution $x \approx \sqrt{2 \ln n}$. This motivates our conjecture that

$$\lim_{n \rightarrow \infty} \frac{\text{AnytimeNormRegret}(n)}{\sqrt{\ln(n)/2}} = 1.$$

In other words, we conjecture that the optimal regret for n experts in the anytime setting is asymptotically the same as the fixed time setting, in the limit as n increases.

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

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

The key aspect is here the lim sup rather than a sup. 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. This strengthened result is proved in Section 4.1.

The algorithm’s description and analysis relies heavily on a function $R: \mathbb{R}_{\geq 0} \times \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$R(t, g) = \begin{cases} 0 & (t = 0), \\ \frac{g}{2} + \kappa \sqrt{t} \cdot M_0\left(\frac{g^2}{2t}\right) & (t > 0 \text{ and } g \leq \gamma \sqrt{t}), \\ \frac{\gamma \sqrt{t}}{2} & (t > 0 \text{ and } g \geq \gamma \sqrt{t}), \end{cases} \quad (2.8)$$

where

$$\kappa = \frac{1}{\sqrt{2\pi} \operatorname{erfi}(\gamma/\sqrt{2})}$$

and M_0 as defined in (2.5). The function R may seem mysterious at first, but in fact arises naturally from the solution to a stochastic calculus problem⁶ in Section 5. 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 $\{(t, \gamma \sqrt{t}) : t > 0\}$ since γ satisfies $M_0(\gamma^2/2) = 0$. We next define a function p to be

$$p(t, g) = \frac{1}{2} (R(t, g+1) - R(t, g-1)). \quad (2.9)$$

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

⁶As we will see below, the regret against a random adversary is a stochastic integral. Viewing this problem in continuous time, then designing a function to minimize the integral leads to a PDE which R solves.

greatest accumulated loss so far at time t . 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 3.5) for all $t \geq 1$ so the algorithm places equal mass on both experts when their cumulative losses are equal.

Algorithm 1 The algorithm achieving the minimax anytime regret for two experts. At each time step, each expert incurs a loss in the interval $[0, 1]$, so the loss vector ℓ_t lies in $[0, 1]^2$.

- 1: Initialize $L_0 \leftarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.
 - 2: **for** $t = 1, 2, \dots$ **do**
 - 3: If necessary, swap indices so that $L_{t-1,1} \geq L_{t-1,2}$.
 - 4: The current gap is $g_{t-1} \leftarrow L_{t-1,1} - L_{t-1,2}$.
 - 5: Set $x_t \leftarrow [p(t, g_{t-1}), 1 - p(t, g_{t-1})]$, where p is the function defined by (2.9).
 - 6: ▷ Observe loss vector ℓ_t and incur loss $\langle x_t, \ell_t \rangle$.
 - 7: $L_t \leftarrow L_{t-1} + \ell_t$.
 - 8: **end for**
-

2.3. Techniques

Lower bound. A common approach to prove lower bounds in the experts problem is to consider a random adversary. With two experts, this adversary chooses one of the two experts uniformly at random and assigns a loss of 1 to that expert and a loss of 0 to the other expert. One can equivalently think of the adversary in the following way, which is more convenient for our purposes. If the gap between the experts at time $t - 1$ is positive then at time t , the adversary changes the gap by ± 1 uniformly at random. On the other hand, if the gap between the experts at time $t - 1$ is 0 then at time t , the adversary *deterministically* increases the gap by 1. This gap process is known as a reflected random walk. 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 perhaps unsurprising that an adversary using a fixed time horizon does not provide the optimal anytime lower bound.

To obtain the optimal lower bound, we allow the adversary to select a *random time* τ at which the game ends. In general, a random adversary in the anytime setting could generate an infinitely long sequence of random bits as its losses, then select the ending time τ as a function of the entire sequence. We will consider a weaker random

adversary in which τ is not a function of the entire sequence, but instead the event $\tau = t$ is known at time t ; that is, τ is a stopping time [38, §9.1]. There is an adversary of this weaker form that is nonetheless optimal, as we discuss next.

First, 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 $\text{Regret}(t) = Z_t + g_t/2$, where Z_t is a martingale with $Z_0 = 0$ and g_t is the current gap at time t . When τ is a stopping time satisfying certain hypotheses, the *optional stopping theorem* (OST) yields $E[Z_\tau] = Z_0 = 0$. We will restrict our attention to adversaries whose stopping times satisfy the hypotheses of the OST. (The stopping times in the fixed-time and geometric horizon settings trivially satisfy the hypotheses.) It is not a priori obvious that there is an optimal adversary in the anytime setting satisfying this restriction.

Concretely, we will consider adversaries that select τ to be the first time that the gap g_t exceeds⁷ some time dependent boundary $f(t)$. This approach follows an established doctrine that connects optimal stopping and stochastic control problems to free-boundary problems [16, 47]. The conclusion of the OST is then that

$$E[\text{Regret}(\tau)] = E[g_\tau]/2 \geq E[f(\tau)]/2.$$

However, the hypotheses of the OST must be respected, 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.

To understand what boundaries $f(t)$ to consider, let us discuss the OST hypotheses. First, 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 [25] to prove lower bounds that contradict the $O(\sqrt{t})$ upper bound of Cover or MWU.) At this point a lucky guess is required: we will consider boundaries of the form $f(t) = c\sqrt{t}$, since this is consistent with the known $\Theta(\sqrt{t})$ regret bounds. It is known [8, 53] that choosing $c < 1$ is necessary and sufficient to ensure that $E[\tau] < \infty$, which is a sufficient hypothesis for the OST. Unfortunately this only yields a regret lower bound of $\sqrt{t}/2$, which is trivial. (The algorithm can easily be forced to have regret $1/2$ at time $t = 1$.) Therefore, the condition $E[\tau] < \infty$ is too restrictive.

Fortunately, there is a less-widely known hypothesis for the OST that allowed us to determine the optimal boundary. It is stated as follows: 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 $E[\sqrt{\tau}] < \infty$, then $E[Z_\tau] = E[Z_0] = 0$. The crucial detail is to bound the *expected square root* of τ . This result is stated formally in Theorem 4.2.

⁷Note that $\tau = \min\{t \geq 0 : g_t \geq f(t)\}$ is a stopping time.

It remains to choose as large a boundary as possible such that the associated stopping time of hitting the boundary satisfies $E[\sqrt{\tau}] < \infty$. Using classical results of Breiman [8] and Greenwood and Perkins [33], we show that the optimal choice of c is γ . While this version of the OST provided intuition leading to the correct boundary, interestingly our proof in Section 4 does not explicitly use this version of the OST.

Upper bound. Our analysis of Algorithm 1, to prove the upper bound in Theorem 2.1, uses a deceptively simple argument where R defined in (2.8) 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 $R(t, g_t) - R(t - 1, g_{t-1})$. By telescoping, this immediately 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., [15, 44]). 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 analog 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 analog of a random walk). Using this adversary, the continuous-time regret becomes a stochastic integral.

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.⁸ However there is a stochastic analog of the FTC, namely the (continuous) Itô formula, which we state in Theorem 5.3. 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

⁸The integrator is reflected Brownian motion, which is not of bounded variation.

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.

2.4. Application

As mentioned in Section 1, 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 2.2 (Davis [19]). *Let $(X_t)_{t \geq 0}$ be a standard random walk. Then $E[|X_\tau|] \leq \gamma E[\sqrt{\tau}]$ for every stopping time τ ; moreover, the constant γ cannot be improved.*

Proof. We begin by proving the first assertion. Suppose that $\text{Regret}(T)$ is the regret process when Algorithm 1 is used against a random adversary. As discussed in Section 2.3 and (4.1), we can write the regret process as

$$\text{Regret}(T) = Z_T + g_T/2,$$

where Z_T is a martingale and g_T evolves as a reflected random walk.⁹ Moreover, if τ is a stopping time satisfying $E[\sqrt{\tau}] < \infty$, then $E[Z_\tau] = 0$ (see Theorem 4.2).

The upper bound in Theorem 2.1 asserts that

$$\gamma\sqrt{T}/2 \geq \text{Regret}(T) = Z_T + g_T/2$$

simultaneously for all $T \geq 0$. Hence, $\gamma E[\sqrt{\tau}]/2 \geq E[g_\tau]/2$. 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 [8] and Greenwood and Perkins [33] as mentioned in Section 2.3 (see also Section 4 or [19]). ■

Remark. Davis [19] proved Theorem 2.2 for both random walks and Brownian motion. We are also able to recover the result for Brownian motion as a corollary of our continuous-time result (Theorem 5.2). The proof is very similar to that above.

⁹Equality holds because our algorithm satisfies $p(t, 0) = 1/2$; this is discussed in the text preceding (4.1).

Remark. In retrospect, our arguments in Section 5 have some similarities with Davis’ proof of the Brownian motion version of Theorem 2.2 [19, Theorem 1.1]. For example, in Section 5, we see that the backwards heat equation appears naturally in the design of the continuous algorithm. Analogously, the backwards heat equation also appears in the proof of [19, Theorem 1.1] but as a result of searching for a supermartingale.

2.5. An expression for the regret involving the gap

In our two-expert prediction problem, the most important scenario restricts each loss vector ℓ_t to be either $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. That is, at each time step, some expert incurs loss 1 and the other expert incurs no loss. This restricted scenario is equivalent to the condition $g_t - g_{t-1} \in \{\pm 1\}$ for all $t \geq 1$, where $g_t := |L_{t,1} - L_{t,2}|$ is the gap at time t . To prove the optimal lower bound it suffices to consider this restricted scenario. The optimal upper bound is first proven in the restricted scenario, then extended to general loss vectors in Section 3.3. With the sole exception of Section 3.3, 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 $\text{Regret}(t) := \text{Regret}(2, t, \mathcal{A}, \mathcal{B})$ where \mathcal{A} and \mathcal{B} are usually implicit from the context.

Proposition 2.3. *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 $\ell_1, \dots, \ell_{t-1}$. Then*

$$\text{Regret}(T) = \sum_{t=1}^T p_t \cdot (g_t - g_{t-1}) \cdot \mathbf{1}[g_{t-1} \neq 0] + \sum_{t=1}^T \langle x_t, \ell_t \rangle \cdot \mathbf{1}[g_{t-1} = 0]. \quad (2.10)$$

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

$$\text{Regret}(T) = \sum_{t=1}^T p_t \cdot (g_t - g_{t-1}). \quad (2.11)$$

Remark. Observe that (2.11) 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 (2.11) 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 (2.10) is the Doob decomposition [38, Theorem 10.1] of the regret process $(\text{Regret}(t))_{t \geq 0}$, i.e., the first sum is a martingale and the second sum is an increasing predictable process.

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

$$\Delta_{\mathbf{R}}(t) = \langle x_t, \ell_t \rangle - (L_t^* - L_{t-1}^*).$$

Case 1: $g_{t-1} \neq 0$. In this case, the best expert at time $t - 1$ remains a best expert at time t . Note that this uses the assumption that $g_t - g_{t-1} \in \{\pm 1\}$, so $g_{t-1} \geq 1$. If the worst expert incurs loss 1, then the algorithm incurs loss p_t and the best expert incurs loss 0, so $\Delta_{\mathbf{R}}(t) = p_t$ and $g_t - g_{t-1} = 1$. Otherwise, the best expert incurs loss 1 and the algorithm incurs loss $1 - p_t$, so $\Delta_{\mathbf{R}}(t) = -p_t$ and $g_t - g_{t-1} = -1$. For either choice of loss, we have $\Delta_{\mathbf{R}}(t) = p_t \cdot (g_t - g_{t-1})$.

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

The above two cases prove (2.10). 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 (2.10) into one to get (2.11). ■

2.6. Basic facts about confluent hypergeometric functions

For any $a, b \in \mathbb{R}$ with $b \notin \mathbb{Z}_{\leq 0}$, the confluent hypergeometric function of the first kind is defined as

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

where $(x)_n := \prod_{i=0}^{n-1} (x + i)$ is the Pochhammer symbol. See, e.g., Abramowitz and Stegun [3, eq. (13.1.2)].

For notational convenience, for $i \in \{0, 1, 2, \dots\}$, we write

$$M_i(x) = M(i - 1/2, i + 1/2, x). \quad (2.12)$$

Fact 2.4. If $b \notin \mathbb{Z}_{\leq 0}$ then $\frac{d}{dx} M(a, b, x) = \frac{a}{b} \cdot M(a + 1, b + 1, x)$. Consequently,

- (1) $M'_0(x) = -M_1(x)$; and
- (2) $M'_1(x) = \frac{1}{3} \cdot M_2(x)$.

Proof. See [3, eq. (13.4.9)]. ■

Fact 2.5. The following identities hold:

- (1) $M_0(x) = -\sqrt{\pi x} \operatorname{erfi}(\sqrt{x}) + e^x$.

- (2) $M_1(x) = \frac{\sqrt{\pi} \operatorname{erfi}(\sqrt{x})}{2\sqrt{x}}$.
- (3) $M_2(x) = \frac{3(2e^x \sqrt{x} - \sqrt{\pi} \operatorname{erfi}(\sqrt{x}))}{4x^{3/2}}$.
- (4) $\frac{2}{3} \cdot M_2(x) \cdot x + M_1(x) = e^x$.

Proof. (2) See [3], equations (7.1.21) or (13.6.19), and use that $\operatorname{erfi}(x) = -i \operatorname{erf}(ix)$, where $i = \sqrt{-1}$.

(1) Differentiating the right-hand side (using the definition of erfi in (2.5)) yields

$$-\frac{\sqrt{\pi} \operatorname{erfi}(\sqrt{x})}{2\sqrt{x}}.$$

So the right-hand side is an anti-derivative of $-M_1(x)$, by part (2). Thus, the identity (1) follows from Fact 2.4 (1) and the initial condition $M_0(0) = 1$.

(3) This follows directly by differentiating (2) and Fact 2.4 (2).

(4) Immediate from (2) and (3). ■

Fact 2.6. The function $M_0(x)$ is decreasing and concave on $[0, \infty)$.

Remark. In fact, $M_0(x)$ is decreasing and concave on \mathbb{R} but we will not require this fact.

Remark. The function $M_0(x^2/2)$ is also decreasing and concave on $[0, \infty)$. Indeed, the concavity follows from the fact that if f is a non-increasing concave function on \mathbb{R} and g is a convex function on \mathbb{R} then $f(g(x))$ is concave. Although this fact is crucial for our algorithm, we do not explicitly make reference to this in the paper.

Proof. By Fact 2.4, we have $M_0'(x) = -M_1(x)$ and $M_0''(x) = -\frac{1}{3} \cdot M_2(x)$. Note that the coefficients of $M_1(x)$, $M_2(x)$ in their Taylor series are all non-negative. As $x \geq 0$, we have that $M_0'(x), M_0''(x) \leq 0$, as desired. ■

Fact 2.7. The function $x \mapsto M_0(x^2/2)$ has a unique positive root at $x = \gamma$. Moreover, $M_0(x^2/2) > 0$ for $x \in (0, \gamma)$ and $M_0(x^2/2) < 0$ for $x \in (\gamma, \infty)$.

Proof. The Maclaurin expansion of $M_0(x^2/2)$ is given by

$$M_0\left(\frac{x^2}{2}\right) = 1 - \sum_{k=1}^{\infty} \frac{1}{(2k-1)k!} \frac{x^{2k}}{2^k}.$$

Note that $M_0(0) = 1$. It is clear, from the series expansion above (and Fact 2.6), that $M_0(x^2/2)$ is strictly decreasing in x on $(0, \infty)$ and $\lim_{x \rightarrow \infty} M_0(x^2/2) = -\infty$. Hence, $M_0(x^2/2)$ contains a positive root γ and it is unique. Finally, it is clear that $M_0(x^2/2)$ is positive on $(0, \gamma)$ and negative on (γ, ∞) . ■

The proof of the lower bound requires the following fact on M which is a straightforward corollary of [46, Proposition 1]. The proof can be found in Appendix A.

Claim 2.8. *For any $\varepsilon > 0$, there exists $a_\varepsilon \in (-1, -1/2)$ such that the smallest¹⁰ positive root c_ε of $z \mapsto M(a_\varepsilon, 1/2, z^2/2)$ satisfies $c_\varepsilon \geq \gamma - \varepsilon$.*

3. Upper bound

In this section, we prove the upper bound in Theorem 2.1 via a sequence of simple steps. The main ideas of the proof are contained in the restricted setting where the gap changes by ± 1 each step. This corresponds to each loss vector ℓ_t being either $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. We first prove the upper bound in Theorem 2.1 in this restricted setting. In Section 3.3, we extend the analysis to general loss vectors in $[0, 1]^2$ through the use of concavity arguments.

The proof in this section uses the potential function R which, as explained in Section 2.3, is defined via continuous-time arguments in Section 5. Moreover, the structure of the proof is heavily inspired by the proof in the continuous setting.

Moving forward, we need a few observations about the functions R and p , which were defined in equations (2.8) and (2.9). First, we require two straightforward calculations. These are special cases of Lemma 5.9 (with $\tilde{R}_\gamma = \tilde{R}$ and $R_\gamma = R$). For convenience, we restate them here without the subscript but only prove Lemma 5.9 later in the paper.

Lemma 3.1. *Consider the function $\tilde{R}(t, g) = \frac{g}{2} + \kappa \sqrt{t} M_0(\frac{g^2}{2t})$. Then*

$$\frac{\partial}{\partial g} \tilde{R}(t, g) = \frac{1}{2} \left(1 - \frac{\operatorname{erfi}(g/\sqrt{2t})}{\operatorname{erfi}(\gamma/\sqrt{2})} \right).$$

Note that $R(t, g) = \tilde{R}(t, g)$ for $g \leq \gamma \sqrt{t}$ and $R(t, g) = \tilde{R}(t, \gamma \sqrt{t}) = \gamma \sqrt{t}/2$ for $g \geq \gamma \sqrt{t}$. So one should think of \tilde{R} as the “untruncated” version of R . Here, and for the remainder of the paper, we write $(x)_+ = \max\{x, 0\}$.

Lemma 3.2. *For all $t > 0$ and $g \geq 0$, it holds that*

$$\frac{\partial}{\partial g} R(t, g) = \frac{1}{2} \left(1 - \frac{\operatorname{erfi}(g/\sqrt{2t})}{\operatorname{erfi}(\gamma/\sqrt{2})} \right)_+.$$

Lemma 3.3. *For any $t > 0$, $R(t, g)$ is concave and non-decreasing in g .*

¹⁰In fact, there is a unique positive root.

Proof. The fact that $R(t, g)$ is non-decreasing in g follows from Lemma 3.2. The concavity of $R(t, g)$ (in g) follows from the fact that erfi is non-decreasing, so $\frac{\partial}{\partial g} R(t, g)$ is non-increasing in g . ■

As a consequence of Lemma 3.3, we can easily get the maximum value of $R(t, g)$ for any t .

Lemma 3.4. *For any $t > 0$, we have $R(t, g) \leq \gamma\sqrt{t}/2$.*

Proof. Lemma 3.3 shows that $R(t, g)$ is non-decreasing in g . By definition, $R(t, g)$ is constant for $g \geq \gamma\sqrt{t}$. It follows that $\max_g R(t, g) \leq R(t, \gamma\sqrt{t}) = \gamma\sqrt{t}/2$. ■

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 3.5 implies that $p(t, g) \in [0, 1/2]$ for all t, g .

Lemma 3.5. *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 was shown in Lemma 3.3, and an elementary property of concave functions (Fact A.1). The final assertion holds because R is non-decreasing in g , which is also shown in Lemma 3.3. ■

3.1. Analysis when gap increments are ± 1

In this subsection we prove the upper bound of Theorem 2.1 for a restricted class of adversaries (that nevertheless capture the core of the problem). The analysis is extended to all adversaries in Section 3.3.

Theorem 3.6. *Let \mathcal{A} be the algorithm described in Algorithm 1. For any adversary \mathcal{B} such that each loss vector ℓ_t is either $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, we have*

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

Our analysis relies on an identity known as the discrete Itô formula, which is the discrete analog of Itô's formula from stochastic analysis (see Theorem 5.3). To make this connection (in addition to future connections) more apparent, we define the discrete derivatives of a function f to be

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

It was remarked earlier that $p(t, g)$ (see (2.9)) is the discrete derivative of R , and this is because

$$p(t, g) = R_g(t, g). \quad (3.1)$$

Lemma 3.7 (Discrete Itô formula). *Let g_0, g_1, \dots be any sequence of real numbers (not necessarily random) satisfying $|g_t - g_{t-1}| = 1$. Then for any function f and any fixed time $T \geq 1$, we have*

$$\begin{aligned} f(T, g_T) - f(0, g_0) &= \sum_{t=1}^T f_g(t, g_{t-1}) \cdot (g_t - g_{t-1}) \\ &\quad + \sum_{t=1}^T \left(\frac{1}{2} f_{gg}(t, g_{t-1}) + f_t(t, g_{t-1}) \right). \end{aligned} \quad (3.2)$$

This lemma is a small generalization of [28, §2] and [41, Theorem 2] to accommodate a bivariate function f that depends on t . The proof is essentially identical.

Proof. By telescoping,

$$f(T, g_T) - f(0, g_0) = \sum_{t=1}^T (f(t, g_t) - f(t-1, g_{t-1})).$$

Consider a fixed $t \in [T]$. We can write

$$\begin{aligned} f(t, g_t) - f(t-1, g_{t-1}) &= \left(f(t, g_t) - \frac{f(t, g_{t-1}+1) + f(t, g_{t-1}-1)}{2} \right) \\ &\quad + \left(\frac{f(t, g_{t-1}+1) + f(t, g_{t-1}-1)}{2} - f(t-1, g_{t-1}) \right). \end{aligned}$$

For the first bracketed term, by considering the cases $g_t = g_{t-1} + 1$ and $g_t = g_{t-1} - 1$, we have

$$\begin{aligned} f(t, g_t) - \frac{f(t, g_{t-1}+1) + f(t, g_{t-1}-1)}{2} &= \frac{f(t, g_{t-1}+1) - f(t, g_{t-1}-1)}{2} \cdot (g_t - g_{t-1}) \\ &= f_g(t, g_{t-1}) \cdot (g_t - g_{t-1}). \end{aligned} \quad (3.3)$$

Note that the above step is the only place where the assumption that $|g_t - g_{t-1}| = 1$ is used. For the second bracketed term, we have

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

This gives the desired formula. ■

Now we show that the regret involves a discrete integral of the same form as (3.2). Let us recall that Lemma 3.5(1) guarantees $p(t, 0) = 1/2$, i.e., $x_t = [1/2, 1/2]$. Hence, (2.11) gives

$$\text{Regret}(T) = \sum_{t=1}^T p(t, g_{t-1}) \cdot (g_t - g_{t-1}), \tag{3.4}$$

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 apply the discrete Itô formula to (3.4), taking $f = R$. Since $p = R_g = f_g$, observe that the main difference between (3.2) and (3.4) is the absence of $\frac{1}{2} f_{gg}(t, g_{t-1}) + f_t(t, g_{t-1})$ in (3.4). 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 3.8 (Discrete backwards heat inequality). *For all $t \geq 1$ and $g \geq 0$, it holds that*

$$\frac{1}{2} R_{gg}(t, g) + R_t(t, g) \geq 0.$$

This lemma is the most technical part of the discrete analysis. Its proof appears in Section 3.2.

Proof of Theorem 3.6. First we apply Lemma 3.7 to the function R and the sequence g_0, g_1, \dots of (integer) gaps produced by the adversary \mathcal{B} . Then, for any time $T \geq 0$,

$$\begin{aligned} & R(T, g_T) - R(0, g_0) \\ &= \sum_{t=1}^T R_g(t, g_{t-1}) \cdot (g_t - g_{t-1}) + \sum_{t=1}^T \left(\frac{1}{2} R_{gg}(t, g_{t-1}) + R_t(t, g_{t-1}) \right) \\ & \hspace{15em} \text{(by Lemma 3.7)} \end{aligned}$$

$$\begin{aligned}
 &\geq \sum_{t=1}^T p(t, g_{t-1}) \cdot (g_t - g_{t-1}) && \text{(by (3.1) and Lemma 3.8)} \\
 &= \text{Regret}(T) && \text{(by (3.4)).}
 \end{aligned}$$

Since $g_0 = 0$ and $R(0, 0) = 0$, applying Lemma 3.4 shows that

$$\text{Regret}(T) \leq R(T, g_T) \leq \gamma \sqrt{T}/2. \quad \blacksquare$$

This completes the proof of Theorem 3.6. However, the proof does not reveal why Algorithm 1 is optimal. The constant γ in the regret bound appears due to properties of the function R , whose definition has yet to be explained. In Section 5, we will define the function R specifically to obtain γ in the preceding analysis. In Section 4, we prove a matching lower bound and show that γ is indeed the right constant.

We remark that the proof of Theorem 3.6 may also be viewed as an amortized analysis, in which the algorithm incurs amortized regret at most

$$\frac{\gamma}{2}(\sqrt{t} - \sqrt{t-1}) \approx \frac{\gamma}{4\sqrt{t}}$$

at each time step t . This viewpoint may be helpful to see how the potential function used in our setting relates to the potential functions in traditional algorithm design [17, §17.3].

3.2. Proof of Lemma 3.8

Lemma 3.9. *For all $u \in [0, 1/2]$, we have $M_0(u) \geq \sqrt{1-2u}$.*

Proof. The Maclaurin expansion of $M_0(u)$ is given by

$$M_0(u) = 1 - \sum_{k=1}^{\infty} \frac{1}{(2k-1)k!} u^k.$$

Note that

$$\frac{d^k}{dx^k} \sqrt{1-2x} = -\frac{(2k-3)!!}{(1-2x)^{(2k-1)/2}},$$

where $(n)!!$ denotes the double factorial (note that $(-1)!! = 1$).¹¹ Hence, the Maclaurin expansion of $\sqrt{1-2u}$ is

$$\sqrt{1-2u} = 1 - \sum_{k=1}^{\infty} \frac{(2k-3)!!}{k!} u^k.$$

¹¹If $n \in \mathbb{Z}_{\geq 0}$, we define $(n)!! = \prod_{k=0}^{\lceil n/2 \rceil - 1} (n-2k)$. If $n \in \mathbb{Z}_{< 0}$, we define $(n)!!$ via the recursive relation $(n)!! = \frac{(n+2)!!}{n+2}$, so that $(-1)!! = \frac{(1)!!}{1} = 1$.

It is not hard to verify that

$$(2k - 3)!! \geq \frac{1}{2k - 1}.$$

This implies that $M_0(u) \geq \sqrt{1 - 2u}$. ■

Lemma 3.10. *For all $z \in [0, 1)$ and $x \in \mathbb{R}$, we have*

$$M_0\left(\frac{(x + z)^2}{2}\right) + M_0\left(\frac{(x - z)^2}{2}\right) \geq 2\sqrt{1 - z^2}M_0\left(\frac{x^2}{2(1 - z^2)}\right).$$

Proof. Fix $z \in [0, 1)$ and consider the function

$$h_z(x) = M_0\left(\frac{(x + z)^2}{2}\right) + M_0\left(\frac{(x - z)^2}{2}\right) - 2\sqrt{1 - z^2}M_0\left(\frac{x^2}{2(1 - z^2)}\right).$$

Note that $h_z(0) \geq 0$ by applying Lemma 3.9 with $u = z^2/2$. We will show that $x = 0$ is the minimizer of h_z which implies the lemma.

Indeed, computing derivatives, we have

$$\begin{aligned} h'_z(x) &= -M_1\left(\frac{(x + z)^2}{2}\right) \cdot (x + z) - M_1\left(\frac{(x - z)^2}{2}\right) \cdot (x - z) \\ &\quad + 2M_1\left(\frac{x^2}{2(1 - z^2)}\right) \cdot \frac{x}{\sqrt{1 - z^2}}. \end{aligned}$$

As $h'_z(0) = 0$, $x = 0$ is a critical point of h_z . We will now show that h_z is convex which certifies that $x = 0$ is indeed a minimizer.

To obtain h''_z , we differentiate term-by-term. Let $u = (x + z)^2/2$. Then

$$\begin{aligned} \frac{d}{dx}M_1\left(\frac{(x + z)^2}{2}\right) \cdot (x + z) &= \frac{M_2\left(\frac{(x+z)^2}{2}\right) \cdot (x + z)^2}{3} + M_1\left(\frac{(x + z)^2}{2}\right) \\ &= \frac{2M_2(u) \cdot u}{3} + M_1(u) \\ &= \frac{2u(2e^u \sqrt{u} - \sqrt{\pi} \operatorname{erfi}(\sqrt{u}))}{4u^{3/2}} + \frac{\sqrt{\pi} \operatorname{erfi}(\sqrt{u})}{2\sqrt{u}} \\ &= e^u = \exp\left(\frac{(x + z)^2}{2}\right). \end{aligned}$$

The first equality is by Fact 2.4 and the third equality is by identities (2) and (3) in Fact 2.5. We can similarly show that

$$\frac{d}{dx}M_1\left(\frac{(x - z)^2}{2}\right) \cdot (x - z) = \exp\left(\frac{(x - z)^2}{2}\right).$$

Finally, for the last term, we have

$$\begin{aligned}
 & \frac{d}{dx} M_1 \left(\frac{x^2}{2(1-z^2)} \right) \cdot \frac{x}{\sqrt{1-z^2}} \\
 &= \frac{1}{3} M_2 \left(\frac{x^2}{2(1-z^2)} \right) \cdot \frac{x^2}{(1-z^2)^{3/2}} + M_1 \left(\frac{x^2}{2(1-z^2)} \right) \cdot \frac{1}{\sqrt{1-z^2}} \\
 &= \frac{1}{\sqrt{1-z^2}} \left(\frac{2}{3} M_2 \left(\frac{x^2}{2(1-z^2)} \right) \cdot \frac{x^2}{2(1-z^2)} + M_1 \left(\frac{x^2}{2(1-z^2)} \right) \right) \\
 &= \frac{\exp\left(\frac{x^2}{2(1-z^2)}\right)}{\sqrt{1-z^2}},
 \end{aligned}$$

where the first equality uses Fact 2.4 and the last equality is by identity (4) in Fact 2.5.

Hence, we have

$$h_z''(x) = \frac{2e^{x^2/2(1-z^2)} - (e^{(x+z)^2/2} + e^{(x-z)^2/2})\sqrt{1-z^2}}{\sqrt{1-z^2}}.$$

So to check that $h_z''(x) \geq 0$ for all $x \in \mathbb{R}$, it suffices to check that

$$\frac{(e^{(x+z)^2/2} + e^{(x-z)^2/2})\sqrt{1-z^2}}{2} \leq e^{x^2/2(1-z^2)}.$$

Indeed, we have

$$\begin{aligned}
 \frac{(e^{(x+z)^2/2} + e^{(x-z)^2/2})\sqrt{1-z^2}}{2} &\leq \frac{(e^{(x+z)^2/2} + e^{(x-z)^2/2})e^{-z^2/2}}{2} \\
 &= e^{x^2/2} \frac{(e^{xz} + e^{-xz})}{2} \leq e^{x^2/2} e^{x^2 z^2/2} \\
 &= e^{x^2(1+z^2)/2} \leq e^{x^2/2(1-z^2)},
 \end{aligned}$$

where the first inequality is because $1 - a \leq e^{-a}$ for all $a \in \mathbb{R}$, the second inequality is because $(e^a + e^{-a})/2 = \cosh(a) \leq e^{a^2/2}$ for all $a \in \mathbb{R}$, and the last inequality is because $1 + a \leq 1/(1 - a)$ for all $a < 1$. This proves that h_z is convex which concludes the proof that $x = 0$ is a minimizer for h_z and hence, completes the proof of the lemma. \blacksquare

Proof of Lemma 3.8. The inequality $R_t(t, g) + \frac{1}{2}R_{gg}(t, g) \geq 0$ is equivalent to

$$R(t, g + 1) + R(t, g - 1) \geq 2R(t - 1, g). \quad (3.5)$$

We first prove the claim for $t = 1$. In this case, the right-hand side of (3.5) is identically zero. On the other hand, the left-hand side of (3.5) is non-decreasing in g by

Lemma 3.3. Hence, it suffices to prove the inequality for $g = 0$. With $t = 1$ and $g = 0$, we have

$$R(1, 1) + R(1, -1) = 2\kappa M_0(1/2).$$

As M_0 is decreasing (Fact 2.6) and $1/2 \leq \gamma^2/2$, we have $M_0(1/2) \geq M_0(\gamma^2/2) = 0$. So (3.5) holds for $t = 1$ and $g \geq 0$.

For the remainder of the proof, we assume that $t > 1$. We consider a few cases depending on the value of g and t .

Case 1: $g \leq \min\{\gamma\sqrt{t} - 1, \gamma\sqrt{t-1}\}$. In this case, $g + 1 \leq \gamma\sqrt{t}$, $g \leq \gamma\sqrt{t-1}$, and $g - 1 \leq \gamma\sqrt{t}$. Hence,

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

So (3.5) is equivalent to

$$\sqrt{t} \cdot M_0\left(\frac{(g + 1)^2}{2t}\right) + \sqrt{t} \cdot M_0\left(\frac{(g - 1)^2}{2t}\right) \geq 2\sqrt{t-1} \cdot M_0\left(\frac{g^2}{2(t - 1)}\right), \quad (3.6)$$

or rearranging, is equivalent to

$$M_0\left(\frac{(g + 1)^2}{2t}\right) + M_0\left(\frac{(g - 1)^2}{2t}\right) \geq 2\sqrt{1 - 1/t} \cdot M_0\left(\frac{g^2}{2(t - 1)}\right).$$

The latter inequality is true by Lemma 3.10 using $x = g/\sqrt{t}$ and $z = 1/\sqrt{t} \in (0, 1)$.

Case 2: $\gamma\sqrt{t} - 1 \leq g \leq \gamma\sqrt{t-1}$. Let \tilde{R} be the function defined in Lemma 3.1. In this case, we have

$$R(t, g + 1) = \gamma\sqrt{t} = \tilde{R}(t, \gamma\sqrt{t}) \geq \tilde{R}(t, g + 1) = \frac{g + 1}{2} + \kappa\sqrt{t} \cdot M_0\left(\frac{(g + 1)^2}{2t}\right).$$

The inequality is by Lemma 3.1 which implies that $\tilde{R}(t, g + 1)$ is *non-increasing* for $g \in (\gamma\sqrt{t} - 1, \infty)$. Using the lower bound on $R(t, g + 1)$, (3.5) is again implied by (3.6) and we have already verified that (3.6) is true.

Case 3: $\gamma\sqrt{t-1} \leq g \leq \gamma\sqrt{t} - 1$. In this case

$$R(t, g + 1) = \frac{g + 1}{2} + \kappa\sqrt{t} \cdot M_0\left(\frac{(g + 1)^2}{2t}\right),$$

$$R(t, g - 1) = \frac{g - 1}{2} + \kappa \sqrt{t} \cdot M_0 \left(\frac{(g - 1)^2}{2t} \right),$$

$$R(t - 1, g) = \frac{\gamma}{2} \sqrt{t - 1}.$$

As $g \leq \gamma \sqrt{t} - 1$, we have $M_0(\frac{(g-1)^2}{2t}) \geq M_0(\frac{(g+1)^2}{2t}) \geq M_0(\frac{\gamma^2}{2}) = 0$. Here, the first two inequalities are because M_0 is decreasing (Fact 2.6). Hence,

$$R(t, g + 1) + R(t, g - 1) \geq g \geq \gamma \sqrt{t - 1} = 2R(t - 1, g),$$

which is precisely (3.5).

Case 4: $\max\{\gamma \sqrt{t - 1}, \gamma \sqrt{t} - 1\} \leq g$. In this case, $R(t - 1, g)$ and $R(t, g + 1)$ are constant in g but $R(t, g - 1)$ is non-decreasing in g . Hence, it suffices to check (3.5) for $g = \max\{\gamma \sqrt{t - 1}, \gamma \sqrt{t} - 1\}$, which holds by either case 2 (if $\gamma \sqrt{t} - 1 \leq \gamma \sqrt{t - 1}$) or case 3 (if $\gamma \sqrt{t - 1} \leq \gamma \sqrt{t} - 1$). ■

3.3. Analysis of Algorithm 1 for general loss vectors

Theorem 3.6 proves Theorem 2.1 in the special case that $\ell_t \in \{0, 1\}^2$. A standard approach to generalize to the setting where $\ell_t \in [0, 1]^2$ is to simulate ℓ_t as a distribution over loss vectors in $\{0, 1\}^2$ (see [32, Claim 4, Appendix A]). Here, we take a different approach and directly show that Algorithm 1 achieves the optimal regret without the need to do this simulation.

Theorem 3.11. *Let \mathcal{A} be the algorithm described in Algorithm 1. For any adversary \mathcal{B} (allowing any loss vectors $\ell_t \in [0, 1]^2$), we have*

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

In Section 3.1, since the gap was integer-valued, the identity of the best expert could only change when the gap is exactly 0 (at which time there are two best experts). In general, the gap can be real-valued, so the best expert can switch abruptly, which affects our formula for the regret. We will need to generalize Proposition 2.3 to deal with this possibility. Let $\Delta_R(t) = \text{Regret}(t) - \text{Regret}(t - 1)$.

Proposition 3.12. *Let g_{t-1} be the gap after time $t - 1$ but before playing an action at time t . Let g_t be the gap after time t . Let $p(t, g_{t-1})$ denote the probability mass assigned to the worst expert at time t . Suppose that $p(t, 0) = 1/2$ for all $t \geq 1$.*

(1) *If a best expert at time $t - 1$ remains a best expert at time t , then*

$$\Delta_R(t) = (g_t - g_{t-1})p(t, g_{t-1}).$$

(2) If a best expert at time $t - 1$ is no longer a best expert at time t then

$$\Delta_R(t) = g_t - (g_t + g_{t-1})p(t, g_{t-1}).$$

Moreover, $g_t + g_{t-1} \leq 1$.

The proof of this is very similar to that of Proposition 2.3 and appears in Section 3.3.1

Remark. Note that, at any specific time, the set of best experts may have size either one or two so the choice of the best expert in Proposition 3.12 may be ambiguous. However, note that if $g_{t-1} = 0$ (i.e., there are two best experts at time $t - 1$), then $p(t, g_{t-1}) = 1/2$, so both formulas give $\Delta_R(t) = \frac{1}{2}g_t$. On the other hand, if $g_t = 0$ (i.e., there are two best experts at time t), then both formulas give $\Delta_R(t) = -g_{t-1}p(t, g_{t-1})$. Hence, there is no issue with the ambiguity.

We will need the following identity which is essentially the same as Lemma 3.7 but without specializing to the case where $|g_t - g_{t-1}| = 1$.

Lemma 3.13. *Let g_0, g_1, \dots be a sequence of real numbers. Then for any function f and any fixed time $T \geq 1$, we have*

$$\begin{aligned} f(T, g_T) - f(0, g_0) &= \sum_{t=1}^T f(t, g_t) - \frac{f(t, g_{t-1} + 1) + f(t, g_{t-1} - 1)}{2} \\ &\quad + \sum_{t=1}^T \left(\frac{1}{2} f_{gg}(t, g_{t-1}) + f_t(t, g_{t-1}) \right). \end{aligned}$$

Proof. The proof is identical to the proof of Lemma 3.7 except that we do not perform the simplification in (3.3). ■

When we assumed the gaps were integer-valued, we had

$$\Delta_R(t) = R(t, g_t) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2}$$

because both sides were equal to $R_g(t, g_{t-1}) \cdot (g_t - g_{t-1})$; see (2.11) and (3.3). This does not hold in the general setting, but we will be able to prove the following inequality.

Lemma 3.14. *For all $t \geq 1$,*

$$\Delta_R(t) \leq R(t, g_t) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2}.$$

The proof of Lemma 3.14 appears in Section 3.3.2. Given Lemma 3.14, we can now prove our upper bound in general.

Proof of Theorem 3.11. Fix any $T \geq 1$. Then

$$\begin{aligned}
 R(T, g_T) - R(0, g_0) &= \sum_{t=1}^T R(t, g_t) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2} \\
 &\quad + \sum_{t=1}^T \left(\frac{1}{2} R_{gg}(t, g_{t-1}) + R_t(t, g_{t-1}) \right) \quad (\text{Lemma 3.13}) \\
 &\geq \sum_{t=1}^T \Delta_R(t) \quad (\text{Lemmas 3.14 and 3.8}) \\
 &= \text{Regret}(T).
 \end{aligned}$$

As $g_0 = 0$ and $R(0, 0) = 0$, we have $\text{Regret}(T) \leq R(T, g_T) \leq \gamma\sqrt{T}/2$, where the last inequality is by Lemma 3.4. ■

3.3.1. Proof of Proposition 3.12.

Proof of Proposition 3.12. Fix t and for notational convenience, let $p = p(t, g_{t-1})$ throughout the proof. In addition, throughout the proof, we use expert 1 to refer to the worst expert at time $t - 1$ (chosen arbitrarily if the choice of worst expert is not unique) and use expert 2 to refer to the other expert. Let $\ell_{t,1}, \ell_{t,2} \in [0, 1]$ be the respective losses at time t and $L_{t,1}, L_{t,2}$ be the respective *cumulative* losses up to time t . Note that $g_{t-1} = L_{t-1,1} - L_{t-1,2}$. Finally, we set $L_t^* = \min_{i \in \{1,2\}} L_{t,i}$. By assumption, $L_{t-1}^* = L_{t-1,2}$.

For the first assertion we have $L_t^* = L_{t,2}$ (because a best expert remains a best expert). Note that

$$\ell_{t,1} - \ell_{t,2} = (L_{t,1} - L_{t,2}) - (L_{t-1,1} - L_{t-1,2}) = g_t - g_{t-1}.$$

So the change in the loss of the algorithm can be written as

$$p\ell_{t,1} + (1-p)\ell_{t,2} = p(\ell_{t,1} - \ell_{t,2}) + \ell_{t,2} = p(g_t - g_{t-1}) + \ell_{t,2}.$$

On the other hand, the change in the loss of the best expert is

$$L_t^* - L_{t-1}^* = L_{t,2} - L_{t-1,2} = \ell_{t,2}.$$

Subtracting this from the above equation gives $\Delta_R(t) = (g_t - g_{t-1})p$.

In the second assertion, we have $L_t^* = L_{t,1}$, so $g_t = L_{t,2} - L_{t,1}$. Again, the algorithm incurs loss $p\ell_{t,1} + (1-p)\ell_{t,2}$. This time, note that

$$\ell_{t,1} - \ell_{t,2} = (L_{t,1} - L_{t,2}) - (L_{t-1,1} - L_{t-1,2}) = -g_t - g_{t-1}.$$

So the algorithm incurs loss $-p(g_t + g_{t-1}) + \ell_{t,2}$. On the other hand, the change in the loss of the best expert is

$$\begin{aligned} L_t^* - L_{t-1}^* &= L_{t,1} - L_{t-1,2} = L_{t,1} - L_{t-1,1} + L_{t-1,1} - L_{t-1,2} \\ &= \ell_{t,1} + g_{t-1} = \ell_{t,2} - g_t, \end{aligned}$$

where the last equality uses the identity $\ell_{t,1} - \ell_{t,2} = -g_t - g_{t-1}$. Subtracting this last quantity with the change in the algorithm's loss gives $\Delta_R(t) = g_t - p(g_t + g_{t-1})$.

To complete the proof for the second assertion, it remains to check that $g_t + g_{t-1} \leq 1$. From above, we have the identity $g_t + g_{t-1} = \ell_{t,2} - \ell_{t,1} \leq \ell_{t,2} \leq 1$, as desired. \blacksquare

3.3.2. Proof of Lemma 3.14.

Proof of Lemma 3.14. Fix $t \geq 1$. We will consider the two cases corresponding to the two cases in Proposition 3.12.

Case 1: A best expert at time $t - 1$ remains a best expert at time t . In this case, $\Delta_R(t) = (g_t - g_{t-1})p(t, g_{t-1})$, so it suffices to check that

$$p(t, g_{t-1}) \cdot (g_t - g_{t-1}) \leq R(t, g_t) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2}. \quad (3.7)$$

Rearranging, the above inequality is equivalent to

$$R(t, g_t) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2} - p(t, g_{t-1}) \cdot (g_t - g_{t-1}) \geq 0.$$

If g_{t-1} is fixed then notice that the left-hand side of the above expression is concave in g_t . To see this, Lemma 3.3 implies that $R(t, g_t)$ is concave in g_t , the second term is constant in g_t , and the last term is linear in g_t . Hence, it suffices to verify the inequality when $g_t = g_{t-1} \pm 1$ (Fact A.2). Indeed, if $|g_t - g_{t-1}| = 1$ then, as in (3.3)

$$\begin{aligned} R(t, g_t) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2} \\ &= \frac{R(t, g_{t-1} + 1) - R(t, g_{t-1} - 1)}{2} \cdot (g_t - g_{t-1}) \\ &= p(t, g_{t-1}) \cdot (g_t - g_{t-1}), \end{aligned}$$

where the second equality used the definition of p .

Case 2: A best expert at time $t - 1$ is no longer a best expert at time t . This case is nearly identical to the previous case but in this case

$$\Delta_R(t) = g_t - (g_t + g_{t-1})p(t, g_{t-1})$$

with the promise that $g_t + g_{t-1} \leq 1$. Hence, the inequality we need to verify is that

$$g_t - (g_t + g_{t-1})p(t, g_{t-1}) \leq R(t, g_t) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2}. \quad (3.8)$$

Once again, we do this via a concavity argument. Fix $g_{t-1} \in [0, 1]$. Since $g_t + g_{t-1} \leq 1$, we have $g_t \in [0, 1 - g_{t-1}]$. Notice that the left-hand side of (3.8) is linear in g_t and the right-hand side of (3.8) is concave in g_t (by Lemma 3.3). Hence, again by Fact A.2, it suffices to check the inequality assuming $g_t \in \{0, 1 - g_{t-1}\}$. Note that the case $g_t = 0$ is handled by case 1 since the left-hand side of (3.7) and (3.8) are identical (see also the remark after Proposition 3.12).

Now assume that $g_t = 1 - g_{t-1}$. Then (3.8) becomes

$$1 - g_{t-1} - p(t, g_{t-1}) \leq R(t, 1 - g_{t-1}) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2}.$$

Recall that $p(t, g) = (R(t, g + 1) - R(t, g - 1))/2$ so that the above inequality is equivalent to

$$\begin{aligned} 1 - g_{t-1} - \frac{R(t, g_{t-1} + 1) - R(t, g_{t-1} - 1)}{2} \\ \leq R(t, 1 - g_{t-1}) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2}. \end{aligned}$$

Rearranging the inequality becomes

$$1 \leq g_{t-1} + R(t, 1 - g_{t-1}) - R(t, g_{t-1} - 1).$$

Note that $g_{t-1} \leq 1 \leq \gamma\sqrt{t}$ (since $t \geq 1$ and $\gamma \geq 1$). Hence, by definition of R , the right-hand side of the above inequality is

$$\begin{aligned} g_{t-1} + R(t, 1 - g_{t-1}) - R(t, g_{t-1} - 1) \\ = g_{t-1} + \frac{1 - g_{t-1}}{2} + \kappa\sqrt{t}M_0\left(\frac{(1 - g_{t-1})^2}{2}\right) \\ - \frac{g_{t-1} - 1}{2} - \kappa\sqrt{t}M_0\left(\frac{(g_{t-1} - 1)^2}{2}\right) \\ = 1, \end{aligned}$$

and obviously, $1 \leq 1$. This proves that the desired inequality holds with equality. ■

4. Lower bound

The main result of this section is the following theorem, which implies the lower bound in Theorem 2.1.

Theorem 4.1. *For any $\varepsilon > 0$, there exists T such that for all algorithms \mathcal{A} , there exists an adversary \mathcal{B} such that*

$$\max_{1 \leq t \leq T} \frac{\text{Regret}(2, t, \mathcal{A}, \mathcal{B})}{\sqrt{t}} \geq \frac{\gamma - \varepsilon}{2}.$$

As remarked earlier, we can also show that large regret is obtained infinitely often; see Section 4.1.

It is common in the literature for regret lower bounds to be proven by random adversaries; see, e.g., [14, Theorem 3.7]. We will 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 loss vectors $\ell_1, \ell_2, \dots \in \{0, 1\}^2$ as follows. For all $t \geq 1$:

- If $g_{t-1} > 0$, then ℓ_t is randomly chosen to be one of the vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, uniformly and independent of $\ell_1, \dots, \ell_{t-1}$. Thus, $g_t - g_{t-1}$ is uniform in $\{\pm 1\}$.
- If $g_{t-1} = 0$, then $\ell_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ if $x_{t,1} \geq 1/2$, and $\ell_t = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 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 (2.10) 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

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

(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 [38, Ex. 10.8], the local time can be written as $\sum_{s=1}^t \mathbf{1}[g_{s-1} = 0] = g_t - Z'_t$, where Z'_t is a martingale with uniformly bounded increments and $Z'_0 = 0$. Thus, combining the two martingales, we have

$$\text{Regret}(t) \geq Z_t + \frac{g_t}{2} \quad \forall t \in \mathbb{Z}_{\geq 0}, \tag{4.1}$$

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 $E[\text{Regret}(\tau)] \geq E[g_\tau]/2$ (under some hypotheses on τ and Z). Thus it is natural to design a stopping time τ that maximizes $E[g_\tau]$ and satisfies the hypotheses. We know from (2.4) that the optimal anytime regret at time t is $\Theta(\sqrt{t})$, so one reasonable stopping time would be

$$\tau(c) := \min\{t > 0 : g_t \geq c\sqrt{t}\} \quad (4.2)$$

for some constant c yet to be determined. If $\tau(c)$ and Z satisfy the hypotheses of the optional stopping theorem, then it holds that $E[\text{Regret}(\tau(c))] \geq \frac{c}{2} E[\sqrt{\tau(c)}]$. 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 OST as formulated in standard references requires one of the following hypotheses: (i) τ is almost surely bounded, or (ii) $E[\tau]$ 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., [9, Theorem 5.33], [25, Theorem 4.8.5], [38, Theorem 10.11], [34, Theorem 12.5.1], [50, Theorem II.57.4], or [55, 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 [8, 53] that $E[\tau(c)] < \infty$, with $\tau(c)$ as above, if and only if $c < 1$; this yields a weak lower bound on the regret. Instead, we will 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 of this theorem can be found in our technical report [36].

Theorem 4.2. *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 $E[\sqrt{\tau}] < \infty$, then $E[Z_\tau] = E[Z_0]$.*

Optimizing the stopping time. Since the martingale Z_t defined above has bounded increments, Theorem 4.2 may be applied so long as $E[\sqrt{\tau(c)}] < \infty$, in which case the preceding discussion yields $\text{AnytimeNormRegret}(2) \geq c/2$. We reiterate that the condition $E[\sqrt{\tau(c)}] < \infty$ is a stronger assumption than $\tau(c)$ being almost surely finite. So it remains to determine

$$\sup\{c \geq 0 : E[\sqrt{\tau(c)}] < \infty\}, \quad (4.3)$$

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 theorem, in which M is the confluent hypergeometric function defined in Section 2.6, and the meaning of a *slowly-varying function* π is that $\lim_{x \rightarrow \infty} \pi(ax)\pi(x)^{-1} = 1$ for all $a > 0$; see [26, Chapter VIII, eq. (8.6)].

Theorem 4.3 ([8, Theorem 2], [33, Theorem 5]). *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*

$$\Pr[\tau(c) > u] = u^a \pi(u) = 1,$$

where π is a slowly-varying function.

Fact 4.4 ([26, Lemma VIII.8.2]). Let π be a slowly-varying function. Then for all $\varepsilon > 0$, there exists M_ε such that $\pi(x) \leq x^\varepsilon$ for all $x \geq M_\varepsilon$.

By combining Theorem 4.3 and Fact 4.4, we see that if c is the smallest positive root of the function $x \mapsto M(a, 1/2, x^2/2)$ then for any $\delta > 0$, there exists a constant C_δ such that $\Pr[\tau(c) > u] \leq C_\delta u^{a+\delta}$.

Recall the definition of γ in (2.6). For intuition, let us apply Theorem 4.3 with $c = \gamma$, which is defined so that it is the root for $a = -1/2$ (see (2.12) and Fact 2.5). It then follows that (ignoring the slowly varying function for now),

$$\mathbb{E}[\sqrt{\tau(\gamma)}] = \int_0^\infty \Pr[\sqrt{\tau(\gamma)} > s] ds = \int_0^\infty \Pr[\tau(\gamma) > s^2] ds \sim K \int_0^\infty s^{-1} ds,$$

by Theorem 4.3. This integral is infinite, so the OST (Theorem 4.2) 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 $\mathbb{E}[\sqrt{\tau(c)}]$ is finite.

We now prove Theorem 4.1. Note that while it does not explicitly make use of Theorem 4.2, the above discussion remained fruitful as it led us to the correct constant γ .

Proof of Theorem 4.1. Fix any $\varepsilon > 0$ that is sufficiently small and any algorithm \mathcal{A} . Consider the random adversary, \mathcal{B} , and the stopping times $\tau(c)$ described above. Let $\text{Regret}(t) = \text{Regret}(2, t, \mathcal{A}, \mathcal{B})$ denote the (random) regret at time t for algorithm \mathcal{A} against the random adversary \mathcal{B} . By Claim 2.8, there exists $a_\varepsilon \in (-1, -1/2)$ and $c_\varepsilon \geq \gamma - \varepsilon$ such that c_ε is the unique positive root of $z \mapsto M(a_\varepsilon, 1/2, z^2/2)$. Let $\delta > 0$ be a constant such that $a_\varepsilon + \delta < -1/2$. By Theorem 4.3 and Fact 4.4, there is a constant C_δ such that

$$\Pr[\tau(c_\varepsilon) \geq s] \leq C_\delta s^{a_\varepsilon + \delta}. \tag{4.4}$$

Fix $T \geq 1$. For notation, we write $a \wedge b = \min\{a, b\}$. We have that

$$\begin{aligned} \mathbb{E}[\text{Regret}(\tau(c_\varepsilon) \wedge T)] &\geq \mathbb{E}[Z_{\tau(c_\varepsilon) \wedge T}] + \frac{1}{2} \mathbb{E}[g_{\tau(c_\varepsilon) \wedge T}] && \text{(by (4.1))} \\ &= \mathbb{E}[Z_0] + \frac{1}{2} \mathbb{E}[g_{\tau(c_\varepsilon) \wedge T}] && \text{(by OST)} \\ &\geq \frac{1}{2} \mathbb{E}[g_{\tau(c_\varepsilon) \wedge T} \cdot \mathbf{1}[\tau(c_\varepsilon) < T]] \end{aligned}$$

$$\begin{aligned}
 &\geq \frac{c_\varepsilon}{2} \mathbb{E}[\sqrt{\tau(c_\varepsilon) \wedge T} \cdot \mathbf{1}[\tau(c_\varepsilon) < T]] && \text{(by (4.2))} \\
 &= \frac{c_\varepsilon}{2} \mathbb{E}[\sqrt{\tau(c_\varepsilon) \wedge T}] - \frac{c_\varepsilon}{2} \mathbb{E}[\sqrt{T} \cdot \mathbf{1}[\tau(c_\varepsilon) \geq T]]. && (4.5)
 \end{aligned}$$

Note that in the first equality the optional stopping theorem asserts that $\mathbb{E}[Z_{\tau(c_\varepsilon) \wedge T}] = \mathbb{E}[Z_0]$, since $\tau(c_\varepsilon) \wedge T$ is an almost surely bounded stopping time. Our goal now is to show that the second term in (4.5) tends to 0 as T tends to infinity. Indeed, by (4.4), we have

$$\mathbb{E}[\sqrt{T} \cdot \mathbf{1}[\tau(c_\varepsilon) \geq T]] = \sqrt{T} \cdot \Pr[\tau(c_\varepsilon) \geq T] \leq C_\delta \cdot T^{1/2+a_\varepsilon+\delta}.$$

Since $a_\varepsilon + \delta < -1/2$, we thus conclude that the last term in (4.5) tends to 0 as $T \rightarrow \infty$. To conclude, for any $\varepsilon > 0$, there exists a fixed time T_ε (depending on ε but not on \mathcal{A}) such that

$$\begin{aligned}
 \mathbb{E}[\text{Regret}(\tau(c_\varepsilon) \wedge T_\varepsilon)] &\geq (1 - \varepsilon) \cdot \frac{c_\varepsilon}{2} \mathbb{E}[\sqrt{\tau(c_\varepsilon) \wedge T_\varepsilon}] \\
 &\geq \frac{(\gamma - \varepsilon)(1 - \varepsilon)}{2} \cdot \mathbb{E}[\sqrt{\tau(c_\varepsilon) \wedge T_\varepsilon}].
 \end{aligned}$$

By the probabilistic method, we conclude that for any $\varepsilon > 0$, there exists $T_\varepsilon > 0$ such that for all algorithms \mathcal{A} , there exists losses $\ell_1, \dots, \ell_{T_\varepsilon}$ such that for some $1 \leq t \leq T_\varepsilon$, the regret of \mathcal{A} is at least $\frac{(\gamma - \varepsilon)(1 - \varepsilon)}{2} \cdot \sqrt{t}$. A suitable adjustment to ε completes the proof. \blacksquare

4.1. Large regret infinitely often

In this subsection, we prove the following extension of Theorem 4.1, which shows that one can achieve regret $\gamma\sqrt{t}/2$ infinitely often.

Theorem 4.5. *For any $\varepsilon > 0$, there exists an infinite sequence of times $T_1 < T_2 < \dots$ such that for all algorithms \mathcal{A} , there is an adversary \mathcal{B} such that for every $i \geq 1$,*

$$\max_{T_{i-1} < t \leq T_i} \frac{\text{Regret}(2, t, \mathcal{A}, \mathcal{B})}{\sqrt{t}} \geq \frac{\gamma - \varepsilon}{2}. \quad (4.6)$$

The proof of Theorem 4.5 is very similar to that of Theorem 4.1 except that the proof ignores the first T_0 time steps. Specifically, fix a $T_0 > 0$. We show that there is an adversary that ensures that the regret at time $t + T_0$, for some $t > 0$, is roughly $c\sqrt{t + T_0}/2$, for some c close to (but smaller than) γ .

To accomplish this, first observe that for any algorithm, the regret at time T_0 is at least $-T_0$, since any algorithm incurs at least regret -1 at each time step. Thus, to prove a regret bound of $c\sqrt{t + T_0}/2$ at time $t + T_0$ for any algorithm, it suffices to prove a regret bound of $c\sqrt{t + T_0}/2 + T_0$ at time t for any algorithm. Similar to

Theorem 4.1, the proof makes use of the stopping time of a reflected random walk to hit the $c\sqrt{t + T_0} + 2T_0$ boundary. The difference is that this boundary is no longer a square root boundary; however, it is still an asymptotically square root boundary. Fortunately, there is an extension of Theorem 4.3 to functions which are asymptotically square root.

Theorem 4.6 ([33, Theorem 5]). *Let $c > 1$ and $a < 0$ be such that c is the smallest positive root of the function $x \mapsto M(a, 1/2, x^2/2)$. Let $f(t)$ be a function such that $\lim_{t \rightarrow \infty} f(t)t^{-1/2} = c$. Let $\tau = \inf\{t > 0 : g_t \geq f(t)\}$. Then $\Pr[\tau > u] = u^a \pi(u)$, where π is a slowly-varying function.*

Proof of Theorem 4.5. We show how, given any $T_0 \geq 0$, one can find an integer $T > 0$ such that for some $1 \leq t \leq T$, the regret at time $t + T_0$ is at least $(\gamma - \varepsilon)\sqrt{t + T_0}/2$ regardless of the value of $\text{Regret}(2, T_0, \mathcal{A}, \mathcal{B})$. The claim follows from this by first taking $T_0 = 0$ to find a T_1 such that

$$\max_{0=T_0 < t \leq T_1} \text{Regret}(2, t, \mathcal{A}, \mathcal{B})/\sqrt{t} \geq (\gamma - \varepsilon)/2.$$

Having found T_{i-1} , we set $T_0 = T_{i-1}$ to find T_i such that (4.6) holds.

Fix any $\varepsilon > 0$ that is sufficiently small and any algorithm \mathcal{A} . Let \mathcal{B} be the random adversary from the proof of Theorem 4.1. Let

$$\text{Regret}(t) = \text{Regret}(2, T_0 + t, \mathcal{A}, \mathcal{B}) - \text{Regret}(2, T_0, \mathcal{A}, \mathcal{B})$$

denote the regret incurred by \mathcal{A} from time T_0 to time $T_0 + t$ against the adversary \mathcal{B} .

By Claim 2.8, there exists $a_\varepsilon \in (-1, -1/2)$ and $c_\varepsilon \geq \gamma - \varepsilon$ such that c_ε is the unique positive root of $z \mapsto M(a_\varepsilon, 1/2, z^2/2)$. Let

$$f(t) = c_\varepsilon \sqrt{t + T_0} + 2T_0/(1 - \varepsilon).$$

As before, let g_t be the absolute value of a standard random walk. Note that

$$\lim_{t \rightarrow \infty} f(t)t^{-1/2} = c_\varepsilon.$$

Define the stopping time $\tau = \inf\{t > 0 : g_t \geq f(t)\}$. Finally, let $\delta > 0$ be a constant such that $a_\varepsilon + \delta < -1/2$. By Theorem 4.6 and Fact 4.4, there is a constant C_δ such that $\Pr[\tau > s] \leq C_\delta s^{a_\varepsilon + \delta}$. Fix $T \geq 1$. For notation, we write $a \wedge b = \min\{a, b\}$. Following the proof of Theorem 4.1, we have that there is a martingale Z_t such that

$$\begin{aligned} \mathbb{E}[\text{Regret}(\tau \wedge T)] &\geq \mathbb{E}[Z_{\tau \wedge T}] + \frac{1}{2} \mathbb{E}[g_{\tau \wedge T}] \\ &= \mathbb{E}[Z_0] + \frac{1}{2} \mathbb{E}[g_{\tau \wedge T}] \end{aligned}$$

$$\begin{aligned}
 &\geq \frac{1}{2} \mathbb{E}[g_{\tau \wedge T} \cdot \mathbf{1}[\tau < T]] \\
 &\geq \frac{1}{2} \mathbb{E}[f(\tau \wedge T) \cdot \mathbf{1}[\tau < T]] \\
 &= \frac{1}{2} \mathbb{E}[f(\tau \wedge T)] - \frac{1}{2} \mathbb{E}[f(T) \cdot \mathbf{1}[\tau \geq T]]. \tag{4.7}
 \end{aligned}$$

Our goal now is to show that the second term in (4.7) tends to 0 as T tends to infinity. Indeed, we have

$$\mathbb{E}[f(T) \cdot \mathbf{1}[\tau \geq T]] = f(T) \cdot \Pr[\tau \geq T] \leq C_\delta \cdot f(T) \cdot T^{a_\varepsilon + \delta}.$$

Note that

$$f(T) \cdot T^{a_\varepsilon + \delta} = (f(T) \cdot T^{-1/2}) \cdot (T^{1/2 + a_\varepsilon + \delta}).$$

The first factor tends to c_ε as $T \rightarrow \infty$ and the second factor tends to 0 as $T \rightarrow \infty$ since $a_\varepsilon + \delta < -1/2$. We thus conclude that the last term in (4.7) tends to 0 as $T \rightarrow \infty$. To conclude, for any $\varepsilon > 0$, there exists a fixed time T_ε (depending on ε but not on \mathcal{A}) such that

$$\begin{aligned}
 \mathbb{E}[\text{Regret}(\tau \wedge T_\varepsilon)] &\geq \frac{1 - \varepsilon}{2} \mathbb{E}[f(\tau \wedge T_\varepsilon)] \\
 &\geq \frac{(1 - \varepsilon)c_\varepsilon}{2} \mathbb{E}[\sqrt{T_0 + (\tau \wedge T_\varepsilon)}] + T_0.
 \end{aligned}$$

Now recall that we had defined

$$\text{Regret}(t) = \text{Regret}(2, T_0 + t, \mathcal{A}, \mathcal{B}) - \text{Regret}(2, T_0, \mathcal{A}, \mathcal{B})$$

and that $\text{Regret}(2, T_0, \mathcal{A}, \mathcal{B}) \geq -T_0$. By the probabilistic method, we conclude that for any $\varepsilon > 0$, there exists $T_\varepsilon > 0$ such that for all algorithms \mathcal{A} , there exists losses $\ell_{T_0+1}, \dots, \ell_{T_0+T_\varepsilon}$ (independent of $\ell_1, \dots, \ell_{T_0}$) such that, for some $1 \leq t \leq T_\varepsilon$, the regret of \mathcal{A} at time $T_0 + t$ is at least

$$\frac{(1 - \varepsilon)}{2} \cdot f(t) - T_0 \geq \frac{(\gamma - \varepsilon)(1 - \varepsilon)}{2} \sqrt{T_0 + t}. \quad \blacksquare$$

5. Derivation of a continuous-time analog of Algorithm 1

The purpose of this section is to show how the potential function R defined in (2.8) arises naturally as the solution of a stochastic calculus problem. The derivation of that function is accomplished by defining, then solving, an analog of the regret minimization problem in continuous time. The main advantage of considering this continuous setting is the wealth of analytic methods available, such as stochastic calculus.

5.1. Defining the continuous regret problem

Continuous time regret problem. The continuous regret problem is inspired by equation (2.11). Notice that, when the adversary chooses loss vectors in $\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\}$, the sequence of gaps g_0, g_1, g_2, \dots 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 5.1 (Continuous regret). Let $p: \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ 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

$$\text{ContRegret}(T, p, B) = \int_0^T p(t, |B_t|) d|B_t|. \tag{5.1}$$

Remark. The condition $p(t, 0) = 1/2$ is due to (5.1) being inspired by (2.11), which requires this condition.

In this definition we may think of p as a continuous-time algorithm and B as a continuous-time adversary. The goal for the remainder of this section is to prove the following result.

Theorem 5.2. *There exists a continuous-time algorithm p^* such that*

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

Remark. A natural question arises upon reviewing the definition of continuous regret: What role does Brownian motion play in Definition 5.1 and is it the “correct” stochastic process to consider in order to uncover the optimal algorithm? In the analysis that follows, the only properties of reflected Brownian motion that we use are its non-negativity and that its *quadratic variation* is t . It turns out that one can generalize Theorem 5.2 by allowing any non-negative, continuous semi-martingale X to control the gap process, and by letting time grow at the rate of the *quadratic variation* of X . See Appendix B.2 for more details.

5.2. Connections to stochastic calculus and the backward heat equation

Since $\text{ContRegret}(T)$ evolves as a stochastic integral with respect to a semi-martingale¹² (namely reflected Brownian motion), Itô's lemma provides an insightful decomposition. The following statement of Itô's lemma is a specialization of [49, Theorem IV.3.3] for the special case of reflected Brownian motion.¹³

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

Theorem 5.3 (Itô's formula). *Let $f: \mathbb{R}_{\geq 0} \times \mathbb{R} \rightarrow \mathbb{R}$ be $C^{1,2}$. Then, almost surely,*

$$\begin{aligned} f(T, |B_T|) - f(0, |B_0|) &= \int_0^T \partial_x f(t, |B_t|) d|B_t| \\ &\quad + \int_0^T \underbrace{\left[\partial_t f(t, |B_t|) + \frac{1}{2} \partial_{xx} f(t, |B_t|) \right]}_{=: \overset{*}{\Delta} f(t, |B_t|)} dt. \end{aligned}$$

The integrand of the second integral is an important quantity arising in PDEs and stochastic processes (see, e.g., [22, p. 263]). We will denote it by

$$\overset{*}{\Delta} f(t, x) := \partial_t f(t, x) + \frac{1}{2} \partial_{xx} f(t, x).$$

Some discussion about the statement of Theorem 5.3 appears in Appendix B.1.

Applying Itô's formula to the continuous regret. Comparing these equations, 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 $\partial_{xx} f(t, 0) = 1/2$; the latter two conditions are needed for Definition 5.1 to be applicable. Itô's formula then yields

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

¹²A semi-martingale is a stochastic process that can be written as the sum of a local martingale and a process of finite variation.

¹³Specifically, we are using the statement of Itô's formula that appears in [49, Remark 1 after Theorem IV.3.3] with $X_t = |B_t|$ and $A_t = t$. Note that y in their notation is t in ours and $\langle |B|, |B| \rangle_t = t$.

Path independence and the backward heat equation. At this point a useful idea arises: as a thought experiment, suppose that $\overset{*}{\Delta}f = 0$. Then the second integral would vanish, and we would have the appealing expression

$$\text{ContRegret}(T, p, B) = f(T, |B_T|).$$

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: at time T , the continuous regret of the algorithm p depends only on T and $|B_T|$ (the gap). We say that such an 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 5.4. Let $f: \mathbb{R}_{>0} \times \mathbb{R} \rightarrow \mathbb{R}$ 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*.

We may summarize the preceding discussion with the following proposition.

Proposition 5.5. Let $f: \mathbb{R}_{>0} \times \mathbb{R} \rightarrow \mathbb{R}$ be a $C^{1,2}$ function that satisfies $\overset{*}{\Delta}f = 0$ everywhere with $f(0, 0) = 0$. Let $p = \partial_x f$. Then,

$$\int_0^T p(t, |B_t|) d|B_t| = f(T, |B_T|).$$

Suppose that a function f satisfies the hypothesis of Proposition 5.5 and in addition $p = \partial_x f \in [0, 1]$ with $p(t, 0) = 1/2$. Then, we would have

$$\text{ContRegret}(T, p, B) = f(T, |B_T|). \tag{5.2}$$

We are unable to derive a function that satisfies the properties required for (5.2) to hold along with $\max_{x \geq 0} f(T, x) \leq \gamma \sqrt{T}/2$. Instead, we will begin by relaxing the constraint that $p(t, x) \in [0, 1]$ and allow $p(t, x)$ to be negative. We will overload the notation $\text{ContRegret}(\cdot)$ to include such functions. In the next section, we will derive a family of such functions that all achieve

$$\text{ContRegret}(T, p, |B_T|) = f(T, |B_T|) = O(\sqrt{T}).$$

This is done by setting up and solving the backwards heat equation. Next, we use a “smoothing” argument to obtain a family of functions that all achieve

$$\text{ContRegret}(T, p, |B_T|) = O(\sqrt{T}),$$

and that do satisfy $p(t, x) \in [0, 1]$. Finally, we will optimize $\text{ContRegret}(T, \cdot, |B_T|)$ over this family of functions to prove Theorem 5.2. The constant γ will appear as a consequence of this optimization problem.

5.2.1. Satisfying the backward heat equation. The main result of this section is the derivation of a family of functions $\tilde{p}: \mathbb{R}_{>0} \times \mathbb{R} \rightarrow \mathbb{R}$ that satisfy $\tilde{p}(t, x) \leq 1$, $\tilde{p}(t, 0) = 1/2$ and

$$\text{ContRegret}(T, \tilde{p}, B) = f(T, |B_T|) = O(\sqrt{T}), \tag{5.3}$$

but do not necessarily satisfy $\tilde{p}(t, x) \geq 0$.

The first step is to find a function f which satisfies the partial differential equation $\overset{*}{\Delta} f = 0$. Since the boundary condition $\tilde{p}(t, 0) = 1/2$ is a condition on $\tilde{p} = \partial_x f$, not on f itself, it will be convenient to solve a PDE for \tilde{p} instead, and then to derive f by integrating. However, some care is needed since not all antiderivatives of \tilde{p} (in x) will satisfy the backwards heat equation. Fortunately, we have a useful lemma showing that if \tilde{p} satisfies the backward heat equation, then we can construct an f that also does.

Lemma 5.6. *Suppose that $h: \mathbb{R}_{>0} \times \mathbb{R} \rightarrow \mathbb{R}$ is a $C^{1,2}$ function. Define*

$$f(t, x) := \int_0^x h(t, y) \, dy - \frac{1}{2} \int_0^t \partial_x h(s, 0) \, ds.$$

Then,

- (1) $f \in C^{1,2}$,
- (2) If $\overset{*}{\Delta} h = 0$ over $\mathbb{R}_{>0} \times \mathbb{R}$, then $\overset{*}{\Delta} f = 0$ over $\mathbb{R}_{>0} \times \mathbb{R}$,
- (3) $h = \partial_x f$.

Proof. First, we check that $f \in C^{1,2}$. Let $(t, x) \in \mathbb{R}_{>0} \times \mathbb{R}$. It is easy to check via standard applications of the dominated convergence theorem (DCT) and the fundamental theorem of calculus (FTC) that

- (1) $\partial_t f(t, x) = \int_0^x \partial_t h(t, y) \, dy - \frac{1}{2} \partial_x h(t, 0)$,
- (2) $\partial_x f(t, x) = h(t, x)$, and
- (3) $\partial_{xx} f(t, x) = \partial_x h(t, x)$.

All of the above partial derivatives are clearly continuous since h is $C^{1,2}$.

Next, we show that if $\overset{*}{\Delta} h(t, x) = 0$ for all $(t, x) \in \mathbb{R}_{>0} \times \mathbb{R}$, then $\overset{*}{\Delta} f(t, x) = 0$ for all $\mathbb{R}_{>0} \times \mathbb{R}$. Indeed,

$$\begin{aligned} \overset{*}{\Delta} f(t, x) &= \left(\partial_t + \frac{1}{2} \partial_{xx} \right) f(t, x) \\ &= \int_0^x \partial_t h(t, y) \, dy - \frac{1}{2} \partial_x h(t, 0) + \frac{1}{2} \partial_x h(t, x) \quad (\text{by (1) and (3)}) \end{aligned}$$

$$\begin{aligned}
 &= \int_0^x \underbrace{\left(\partial_t h(t, y) + \frac{1}{2} \partial_{xx} h(t, y) \right)}_{=0} dy && \text{(by FTC)} \\
 &= 0,
 \end{aligned}$$

as claimed. ■

Defining boundary conditions for p . Obtaining a particular solution to the backward heat equation requires sufficient boundary conditions in order to uniquely identify \tilde{p} . The boundary condition mentioned above is that $\tilde{p}(t, 0) = 1/2$ for all t . This condition together with the backward heat equation clearly do not suffice to uniquely determine \tilde{p} . Therefore, we impose some reasonable boundary conditions on \tilde{p} .

What should the value be at the boundary? Intuitively, $x \mapsto \tilde{p}(t, x)$ should be a decreasing function because \tilde{p} represents the weight placed on the worst expert as a function of the gap. Therefore, it is natural to consider an “upper boundary” which specifies the point at which the difference in experts’ total losses is so great that the algorithm places zero weight on the worst expert. The upper boundary can be specified by a curve, $\{(t, \phi(t)) : t > 0\}$ for some continuous function $\phi: \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$. We will incorporate this idea by requiring $\tilde{p}(t, \phi(t)) = 0$ for all $t > 0$.

Where should the boundary be? One reasonable choice for the boundary is to use $\phi_\alpha(t) = \alpha\sqrt{t}$ for some constant $\alpha > 0$, as this is similar to the boundary used by the random adversary in the lower bound of Section 4. For now, we leave α as an unknown parameter whose value can be optimized later. These conditions are combined into the following partial differential equation:

$$\text{Backward heat equation: } \partial_t u(t, x) + \frac{1}{2} \partial_{xx} u(t, x) = 0 \quad \text{for all } (t, x) \in \mathbb{R}_{>0} \times \mathbb{R}, \tag{5.4}$$

$$\text{Upper boundary: } u(t, \alpha\sqrt{t}) = 0 \quad \text{for all } t > 0, \tag{5.5}$$

$$\text{Lower boundary: } u(t, 0) = \frac{1}{2} \quad \text{for all } t > 0. \tag{5.6}$$

Next we show that the following function solves this PDE. Define $\tilde{p}_\alpha: \mathbb{R}_{>0} \times \mathbb{R} \rightarrow \mathbb{R}$ by

$$\tilde{p}_\alpha(t, x) := \frac{1}{2} \left(1 - \frac{\operatorname{erfi}(x/\sqrt{2t})}{\operatorname{erfi}(\alpha/\sqrt{2})} \right). \tag{5.7}$$

Lemma 5.7. *The function \tilde{p}_α satisfies the following properties:*

- (1) \tilde{p}_α is $C^{1,2}$ over $\mathbb{R}_{>0} \times \mathbb{R}$,
- (2) \tilde{p}_α satisfies the constraints in (5.4), (5.5) and (5.6), and
- (3) For all $t > 0$ and all $x \geq 0$, $\tilde{p}_\alpha(t, x) \leq 1/2$.

Proof. Let us assume that we can write $u(t, x) = v(x/\sqrt{t})$. Then, we have

$$\partial_t u(t, x) = -\frac{x}{2t^{3/2}}v'(x/\sqrt{t}) \quad \text{and} \quad \frac{1}{2}\partial_{xx}u(t, x) = \frac{1}{2t}v''(x/\sqrt{t}).$$

The backward heat equation enforces that

$$v''(x/\sqrt{t}) = \frac{x}{\sqrt{t}}v'(x/\sqrt{t}).$$

By a change of variables ($z = x/\sqrt{t}$), we obtain the following ordinary differential equation

$$v''(z) = z \cdot v'(z).$$

Hence, $v'(z) = C \cdot e^{z^2/2}$ for some constant C . We can then integrate to obtain

$$v(z) = \int_0^z C e^{y^2/2} dy + D = \int_0^{z/\sqrt{2}} \sqrt{2}C e^{r^2} dr + D$$

for some constant D . For the last equality, we made the change of variables $r = y/\sqrt{2}$ in the integral. Therefore, by the definition of erfi (and a different constant C), we have $v(z) = C \operatorname{erfi}(z/\sqrt{2}) + D$. Hence, for some constants $C, D \in \mathbb{R}$, we have

$$u(t, x) = C \operatorname{erfi}(x/\sqrt{2t}) + D.$$

Plugging in the boundary condition at $x = 0$ and recalling that $\operatorname{erfi}(0) = 0$, we see that $D = 1/2$. Plugging in the boundary condition that $u(t, \alpha\sqrt{t}) = 0$ and using that $D = 1/2$, we see that

$$C = -\frac{1}{2 \operatorname{erfi}(\alpha/\sqrt{2})}.$$

Therefore, we have that the following function

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

satisfies the backwards heat equation and the boundary conditions. Moreover, $u \in C^{1,2}$ on $\mathbb{R}_{>0} \times \mathbb{R}$. ■

Lemma 5.7 shows that $\tilde{p}_\alpha(t, x)$ nearly defines a valid continuous time algorithm, in that it satisfies the conditions of Definition 5.1 except for non-negativity. Next, we will integrate \tilde{p}_α as described in Lemma 5.6. Define the function $\tilde{R}_\alpha: \mathbb{R}_{>0} \times \mathbb{R} \rightarrow \mathbb{R}$ as

$$\tilde{R}_\alpha(t, x) = \frac{x}{2} + \kappa_\alpha \sqrt{t} \cdot M_0\left(\frac{x^2}{2t}\right), \quad \text{where } \kappa_\alpha = \frac{1}{\sqrt{2\pi} \operatorname{erfi}(\alpha/\sqrt{2})}. \quad (5.8)$$

Lemma 5.8. *The following equation holds:*

$$\tilde{R}_\alpha(t, x) = \int_0^x \tilde{p}_\alpha(t, y) \, dy - \frac{1}{2} \int_0^t \partial_x \tilde{p}_\alpha(s, 0) \, ds.$$

First we need to compute some derivatives.

Lemma 5.9. *The following identities hold for every $\alpha > 0$:*

- (1) $\partial_x \tilde{R}_\alpha(t, x) = \tilde{p}_\alpha(t, x) = \frac{1}{2} \left(1 - \frac{\operatorname{erfi}(x/\sqrt{2t})}{\operatorname{erfi}(\alpha/\sqrt{2t})} \right),$
- (2) $\partial_{xx} \tilde{R}_\alpha(t, x) = \partial_x \tilde{p}_\alpha(t, x) = -\kappa_\alpha \cdot \frac{\exp(x^2/2t)}{\sqrt{t}}.$

Proof. The proof is a straightforward calculation. We have

$$\begin{aligned} \partial_x \tilde{R}_\alpha(t, x) &= \frac{1}{2} - \kappa_\alpha \frac{x}{\sqrt{t}} \cdot M_1\left(\frac{x^2}{2t}\right) \\ &= \frac{1}{2} - \frac{1}{\sqrt{2\pi} \operatorname{erfi}(\alpha/\sqrt{2})} \cdot \frac{x}{\sqrt{t}} \cdot \frac{\sqrt{\pi} \operatorname{erfi}(x/\sqrt{2t})}{2 \cdot x/\sqrt{2t}} \\ &= \frac{1}{2} \left(1 - \frac{\operatorname{erfi}(x/\sqrt{2t})}{\operatorname{erfi}(\alpha/\sqrt{2})} \right), \end{aligned}$$

where the first equality uses Fact 2.4 and the second equality uses identity (2) in Fact 2.5. This proves the first identity.

For the second identity, using the definition of $\operatorname{erfi}(\cdot)$, we have

$$\partial_{xx} \tilde{R}_\alpha = \partial_x \tilde{p}_\alpha(t, x) = -\frac{\exp(x^2/2t)}{\sqrt{2\pi} \operatorname{erfi}(\alpha/\sqrt{2})\sqrt{t}} = -\kappa_\alpha \cdot \frac{\exp(x^2/2t)}{\sqrt{t}}. \quad \blacksquare$$

Proof of Lemma 5.8. By the first identity in Lemma 5.9, we have

$$\int_0^x \tilde{p}_\alpha(t, y) \, dy = \tilde{R}_\alpha(t, x) - \tilde{R}_\alpha(t, 0). \tag{5.9}$$

Note that $\tilde{R}_\alpha(t, 0) = \kappa_\alpha \sqrt{t}$. Next, the second identity of Lemma 5.9 implies that

$$-\partial_x \tilde{p}_\alpha(s, 0) = \frac{\kappa_\alpha}{\sqrt{s}}.$$

Hence,

$$-\frac{1}{2} \int_0^t \partial_x \tilde{p}_\alpha(s, 0) \, ds = \kappa_\alpha \sqrt{t} = \tilde{R}_\alpha(t, 0). \tag{5.10}$$

Summing (5.9) and (5.10) gives

$$\begin{aligned} \int_0^x \tilde{p}_\alpha(t, y) \, dy - \frac{1}{2} \int_0^t \partial_x \tilde{p}_\alpha(s, 0) \, ds &= \tilde{R}_\alpha(t, x) - \tilde{R}_\alpha(t, 0) + \tilde{R}_\alpha(t, 0) \\ &= \tilde{R}_\alpha(t, x). \quad \blacksquare \end{aligned}$$

By Lemma 5.7, the function \tilde{p}_α satisfies the hypothesis of the function h in Lemma 5.6. Applying Lemma 5.6 with $h = \tilde{p}_\alpha$ and $f = \tilde{R}_\alpha$, we have the following observation on \tilde{R}_α .

Observation 5.10. The function \tilde{R}_α satisfies the following properties:

- (1) \tilde{R}_α is $C^{1,2}$,
- (2) \tilde{R}_α satisfies $\Delta^* \tilde{R}_\alpha = 0$ over $\mathbb{R}_{>0} \times \mathbb{R}$,
- (3) $\partial_x \tilde{R}_\alpha(t, x) = \tilde{p}_\alpha(t, x)$.

Observation 5.10 shows that \tilde{R}_α satisfies the hypotheses of Proposition 5.5. Hence, we have

$$\text{ContRegret}(T, \tilde{p}_\alpha, B) = \tilde{R}_\alpha(T, |B_T|).$$

Since $\text{erfi}(\cdot)$ is a strictly increasing function with $\text{erfi}(0) = 0$, observe that $\partial_x \tilde{R}_\alpha = \tilde{p}_\alpha$ has exactly one root at $\alpha\sqrt{t}$. In particular, for any fixed $T > 0$, the function $\tilde{R}_\alpha(T, x)$ is maximized at $x = \alpha\sqrt{T}$. Therefore, for every T , we have

$$\tilde{R}_\alpha(T, |B_T|) \leq \max_{x \geq 0} \tilde{R}_\alpha(T, x) \leq \tilde{R}_\alpha(T, \alpha\sqrt{T}) = \left(\frac{\alpha}{2} + \kappa_\alpha M_0 \left(\frac{\alpha^2}{2}\right)\right) \sqrt{T},$$

where the equality is by definition of \tilde{R}_α in equation (5.8). To summarize, we have shown that

$$\text{ContRegret}(T, \tilde{p}_\alpha, B) \leq \left(\frac{\alpha}{2} + \kappa_\alpha M_0 \left(\frac{\alpha^2}{2}\right)\right) \sqrt{T}. \quad (5.11)$$

This establishes (5.3), as desired.

5.2.2. Resolving the non-negativity issue. The only remaining step is to modify \tilde{p}_α so that it lies in the interval $[0, 1/2]$. We modify \tilde{p}_α in the most natural way: by modifying all negative values to be zero. Specifically, we set

$$p_\alpha(t, x) := \begin{cases} 0 & (t = 0) \\ (\tilde{p}_\alpha(t, x))_+ & (t > 0) \end{cases} = \begin{cases} 0 & (t = 0) \\ \frac{1}{2} \left(1 - \frac{\text{erfi}(x/\sqrt{2t})}{\text{erfi}(\alpha/\sqrt{2})}\right)_+ & (t > 0) \end{cases}. \quad (5.12)$$

Here, we use the notation $(x)_+ = \max\{0, x\}$. Note that $p_\alpha(t, 0) = 1/2$ for all $t > 0$ and $p_\alpha(t, x) \in [0, 1/2]$ for all $t, x \geq 0$. So p_α defines a valid continuous-time algorithm. From (5.12), we obtain a truncated version of \tilde{R}_α as

$$R_\alpha(t, x) := \begin{cases} 0 & (t = 0), \\ \tilde{R}_\alpha(t, x) & (t > 0 \wedge x \leq \alpha\sqrt{t}), \\ \tilde{R}_\alpha(t, \alpha\sqrt{t}) & (t > 0 \wedge x \geq \alpha\sqrt{t}). \end{cases}$$

It is straightforward to verify that $\partial_x R_\alpha = p_\alpha$. This is because for $x \leq \alpha\sqrt{t}$, we have

$$p_\alpha(t, x) = \tilde{p}_\alpha(t, x) \quad \text{and} \quad R_\alpha(t, x) = \tilde{R}_\alpha(t, x)$$

(in Observation 5.10, we noted that $\partial_x \tilde{R}_\alpha = p_\alpha$). In addition, $R_\alpha(t, x)$ is constant (in x) for $x \geq \alpha\sqrt{t}$ so its derivative (in x) is 0.

If R_α were sufficiently smooth then we could immediately apply Itô's formula (Theorem 5.3) to obtain a formula for the regret of p_α . For $0 < x < \alpha\sqrt{t}$, we have

$$\overset{*}{\Delta} R_\alpha(t, x) = 0$$

by Observation 5.10 (because $R_\alpha(t, x) = \tilde{R}_\alpha(t, x)$, whenever $0 < x < \alpha\sqrt{t}$) and for $x > \alpha\sqrt{t}$, it is not difficult to verify that

$$\overset{*}{\Delta} R_\alpha(t, x) > 0.$$

Itô's formula would then suggest that

$$\text{ContRegret}(T, p_\alpha, B) \leq \text{Regret}(T, |B_T|).$$

The only flaw is that $\partial_{xx} R_\alpha$ is not well defined on the curve $\{(t, \alpha\sqrt{t}) : t > 0\}$ so R_α is not in $C^{1,2}$ and Theorem 5.3 cannot be applied directly. Fortunately, it is possible to resolve this issue and to formally prove the following lemma which bounds the continuous regret for the algorithm $p_\alpha = \partial_x R_\alpha$ in terms of the potential function R_α . As the details are quite technical and distract from our main goal of providing intuition on how the discrete algorithm is obtained, the proof of Lemma 5.11 is omitted from the present paper but may be found in our technical report [36].

Lemma 5.11. *Fix $\alpha > 0$. Then, almost surely, for all $T \geq 0$,*

$$\text{ContRegret}(T, p_\alpha, B) \leq R_\alpha(T, |B_T|).$$

The remainder of this section proves Theorem 5.2 by setting $p^* = p_\alpha$ for the optimal α .

Remark. The definition of p_α has an interesting interpretation. Let B be a Brownian motion. Fix a time t and a position $x > 0$. Now let $\tau = \inf\{s > t : |B_s| \geq \alpha\sqrt{s}\}$. It is known [21] that

$$p_\alpha(t, x) = \Pr[B_\tau < 0 \mid B_t = x].$$

In words, $p_\alpha(t, x)$ is the probability that a Brownian motion started at time t and position x crosses the bottom $-\alpha\sqrt{t}$ boundary before crossing the top $+\alpha\sqrt{t}$ boundary. As a sanity check, one may observe that $p_\alpha(t, \alpha\sqrt{t}) = 0$ and $p_\alpha(t, 0) = 0.5$. Interestingly, the optimal algorithms for two experts in both the finite-time horizon setting [18] and the geometric time horizon setting have a similar interpretation [32]. In both cases, the optimal algorithm is to assign the probability that a random walk started at position $x > 0$ at time t remains positive at the stopping time. In the finite-time case, the stopping time is a deterministic quantity T whereas in the geometric-time case, the stopping time is a geometric random variable. A similar connection also exists for three and four experts [6, 32].

5.3. Optimizing the boundary to minimize continuous regret

By Lemma 5.11,

$$\text{ContRegret}(T, \partial_x R_\alpha, B) \leq R_\alpha(T, |B_T|) \leq R_\alpha(T, \alpha\sqrt{T}),$$

where the last inequality is because $\partial_x R_\alpha(t, x) = p_\alpha(t, x)$ is positive for $x \in [0, \alpha\sqrt{t}]$ and 0 for $x \geq \alpha\sqrt{t}$. As observed in (5.11), we have the formula

$$R_\alpha(T, \alpha\sqrt{T}) = (\alpha/2 + \kappa_\alpha M_0(\alpha^2/2))\sqrt{T}.$$

Thus, to minimize $R_\alpha(T, \alpha\sqrt{T})$, it is convenient to define

$$h(\alpha) := R_\alpha(1, \alpha) = \frac{\alpha}{2} + \kappa_\alpha M_0(\alpha^2/2).$$

The only remaining task is now to solve the following optimization problem.

$$\min_{\alpha>0} h(\alpha) = \min_{\alpha>0} \left\{ \frac{\alpha}{2} + \kappa_\alpha \cdot M_0\left(\frac{\alpha^2}{2}\right) \right\} \quad (5.13)$$

The following lemma verifies that there exists some α for which

$$\text{ContRegret}(T, \partial_x R_\alpha, B) \leq \frac{\gamma\sqrt{T}}{2},$$

completing the proof of Theorem 5.2.

Lemma 5.12. *The function $h(\alpha)$ is minimized at $\alpha = \gamma$ and $h(\gamma) = \gamma/2$. Consequently, for any fixed $T > 0$,*

$$\min_{\alpha} R_\alpha(T, \alpha\sqrt{T}) = R_\gamma(T, \gamma\sqrt{T}) = \frac{\gamma\sqrt{T}}{2}.$$

Lemma 5.12 follows easily from the following claim.

Claim 5.13. *For all $\alpha > 0$, it holds that*

$$h'(\alpha) = -\frac{\exp(\alpha^2/2)}{\pi \operatorname{erfi}(\alpha/\sqrt{2})} \cdot M_0(\alpha^2/2).$$

In particular, $h'(\alpha) < 0$ for $\alpha \in (0, \gamma)$, $h'(\gamma) = 0$, and $h'(\alpha) > 0$ for $\alpha \in (\gamma, \infty)$.

Proof. Recall that $h(\alpha) = \frac{\alpha}{2} + \frac{M_0(\alpha^2/2)}{\sqrt{2\pi} \operatorname{erfi}(\alpha/\sqrt{2})}$ and that $\frac{d}{dx} \operatorname{erfi}(x/\sqrt{2}) = \sqrt{\frac{2}{\pi}} e^{x^2/2}$. Hence,

$$\begin{aligned} h'(\alpha) &= \frac{1}{2} - \frac{\alpha \cdot M_1(\alpha^2/2)}{\sqrt{2\pi} \operatorname{erfi}(\alpha/\sqrt{2})} - \frac{\exp(\alpha^2/2) \cdot M_0(\alpha^2/2)}{\pi \operatorname{erfi}(\alpha/\sqrt{2})^2} \quad (\text{by Fact 2.4}) \\ &= -\frac{\exp(\alpha^2/2) \cdot M_0(\alpha^2/2)}{\pi \operatorname{erfi}(\alpha/\sqrt{2})^2} \quad (\text{by Fact 2.5 (2)}). \end{aligned}$$

This proves the first assertion.

Next, observe that $\exp(\alpha^2/2)/\operatorname{erfi}(\alpha/\sqrt{2})^2$ is positive for all $\alpha > 0$. Hence, by Fact 2.7, we have that $h'(\alpha) < 0$ for $\alpha \in (0, \gamma)$, $h'(\gamma) = 0$, and $h'(\alpha) > 0$ for $\alpha \in (\gamma, \infty)$. ■

Proof of Lemma 5.12. Claim 5.13 implies that γ is the global minimizer for $h(\alpha)$. Since γ is a root of $M_0(\alpha^2/2)$, it follows that $h(\gamma) = \gamma/2$. This proves the first assertion. Next, for every $\alpha > 0$, we have

$$R_\alpha(T, \alpha\sqrt{T}) = \sqrt{T} \cdot h(\alpha) \geq \sqrt{T} \cdot h(\gamma) = \gamma\sqrt{T}/2,$$

which proves the second assertion. ■

A. Standard concavity facts

Fact A.1. Suppose $f: \mathbb{R} \rightarrow \mathbb{R}$ is concave. Then for any $\alpha < \beta$, the function $g(t) = f(t + \beta) - f(t + \alpha)$ is non-increasing.

Fact A.2. Suppose $f: \mathbb{R} \rightarrow \mathbb{R}$ is concave. Let $\alpha < \beta$. Then $f(x) \geq \min\{f(\alpha), f(\beta)\}$ for all $x \in [\alpha, \beta]$.

Claim 2.8. For any $\varepsilon > 0$, there exists $a_\varepsilon \in (-1, -1/2)$ such that the smallest¹⁴ positive root c_ε of $z \mapsto M(a_\varepsilon, 1/2, z^2/2)$ satisfies $c_\varepsilon \geq \gamma - \varepsilon$.

Proof. Following Perkins' notation [46], let $\lambda_0(-c, c)$ be such that c is the smallest positive root of

$$x \mapsto M(-\lambda_0(-c, c), 1/2, x^2/2).$$

By [46, Proposition 1], the map $c \mapsto \lambda_0(-c, c)$ is strictly decreasing and continuous on $\mathbb{R}_{>0}$, so it has a continuous inverse α . From (2.6) and Fact 2.5 (1), we see that

$$\lambda_0(-\gamma, \gamma) = 1/2,$$

hence $\alpha(1/2) = \gamma$. By continuity, for all $\varepsilon > 0$, there exists $\delta \in (0, 1/2)$ such that $\alpha(1/2 + \delta) > \gamma - \varepsilon$. Then we may take $a_\varepsilon = -(1/2 + \delta)$ and $c_\varepsilon = \alpha(1/2 + \delta)$. ■

The following fact is well known; see, e.g., [48, Lemma 2].

Fact A.3. Let \mathcal{X} and \mathcal{Y} be topological spaces, and let $f: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ be a (jointly) continuous function. If \mathcal{Y} is compact then

$$g(x) = \inf_{y \in \mathcal{Y}} f(x, y) \quad \text{and} \quad h(x) = \sup_{y \in \mathcal{Y}} f(x, y)$$

are continuous functions on \mathcal{X} .

¹⁴In fact, there is a unique positive root.

B. Additional discussions for Section 5

B.1. Discussion on the statement of Theorem 5.3

In this paper, we use the version of Itô’s formula that appears in [49, Remark 1 after Theorem IV.3.3]. It states that if $f \in C^{1,2}$, X is a continuous semi-martingale¹⁵ and A is a process with bounded variation, then

$$\begin{aligned}
 f(A_T, X_T) - f(A_0, X_0) &= \int_0^T \partial_x f(A_t, X_t) dX_t + \int_0^T \partial_t f(A_t, X_t) dA_t \\
 &\quad + \frac{1}{2} \int_0^T \partial_{xx} f(A_t, X_t) d\langle X, X \rangle_t. \tag{B.1}
 \end{aligned}$$

In our setting, we take $X_t = |B_t|$ and $A_t = t$. We now explain the notation $\langle X, X \rangle$.

- (1) For a continuous local martingale M , $\langle M, M \rangle$ is the unique increasing continuous process vanishing at 0 such that $M^2 - \langle M, M \rangle$ is a martingale [49, Theorem IV.1.8].
- (2) If X is a continuous semi-martingale with M being the (continuous) local martingale part then $\langle X, X \rangle = \langle M, M \rangle$ [49, Definition IV.1.20].

Tanaka’s formula [51, Theorem IV.43.3] asserts that $|B_t| = W_t + L_t$, where W_t is a Brownian motion and L_t is the local time of B_t at 0, which is an increasing, continuous, adapted process. Hence, $|B_t|$ is a semi-martingale with $\langle |B|, |B| \rangle_t = \langle W, W \rangle_t = t$. Plugging these into (B.1) gives

$$\begin{aligned}
 f(T, |B_T|) - f(0, |B_0|) &= \int_0^T \partial_x f(t, |B_t|) d|B_t| \\
 &\quad + \int_0^T \left[\partial_t f(t, |B_t|) + \frac{1}{2} \partial_{xx} f(t, |B_t|) \right] dt,
 \end{aligned}$$

which is what appears in Theorem 5.3.

B.2. Continuous regret against any continuous semi-martingale

Recall that the continuous regret upper bound (Theorem 5.2) involved the adversary evolving the gap process as a reflected Brownian motion, which is a continuous semi-martingale. In this section, we generalize the definition of continuous regret to allow arbitrary, non-negative, continuous semi-martingales to control the gap process, and derive an analog of Theorem 5.2 in this generalized setting. We use the notation $[X]_t$

¹⁵A continuous semi-martingale X is a process that can be written as $X = M + N$, where M is a continuous local martingale and N is a continuous adapted process of finite variation.

to refer to $\langle X, X \rangle_t$, the quadratic variation process of X , which was introduced in Appendix B.1.

We begin with a generalized definition of continuous regret.

Definition B.1 (Continuous regret). Let $p: \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ be a continuous function that satisfies $p(t, 0) = 1/2$ for every $t > 0$. Let X_t be a continuous, non-negative, semi-martingale. Then, the *continuous regret* of p with respect to X is the stochastic integral

$$\text{ContRegret}(T, p, X) = \int_0^T p(t, X_t) dX_t.$$

The main result for this generalized setting is as follows.

Theorem B.2. *There exists a continuous-time algorithm p^* such that for any continuous, non-negative, semi-martingale X ,*

$$\text{ContRegret}(T, p^*, X) \leq \frac{\gamma}{2} \sqrt{[X]_T} \quad \forall T \in \mathbb{R}_{\geq 0}, \text{ almost surely.}$$

We provide an overview of the proof of this result below. For the sake of exposition, we sketch the proof of Theorem B.2 in the setting where we allow p^* to take values in $(-\infty, 1]$. Truncating p^* as was done in Section 5.2.2 yields Theorem B.2 as stated.

Proof sketch. Let $p^*(t, x) := \tilde{p}_\gamma([X]_t, x)$ and $R(t, x) := \tilde{R}_\gamma(t, x)$. (See equations (5.7) and (5.8) for definitions of \tilde{p}_γ and \tilde{R}_γ .) Recall the following three important properties of R from Observation 5.10:

- (1) R is $C^{1,2}$,
- (2) R satisfies $\overset{*}{\Delta} R = 0$ over $\mathbb{R}_{>0} \times \mathbb{R}$,
- (3) $\partial_x R(t, x) = \tilde{p}_\gamma(t, x)$.

Since R is $C^{1,2}$, we may apply Itô's formula (specifically equation (B.1) with $A_t = [X]_t$, which is a bounded variation process since it is increasing) to obtain

$$\begin{aligned} & R([X]_T, X_T) \\ &= \int_0^T \partial_x R([X]_t, X_t) dX_t + \int_0^T \partial_t R([X]_t, X_t) + \frac{1}{2} \partial_{xx} R([X]_t, X_t) d[X]_t \\ &= \int_0^T p^*(t, X_t) dX_t \\ &\quad + \underbrace{\int_0^T \left(\partial_t R([X]_t, X_t) + \frac{1}{2} \partial_{xx} R([X]_t, X_t) \right) d[X]_t}_{=\overset{*}{\Delta} R([X]_t, X_t)} \quad (\partial_x R = \tilde{p}_\gamma) \end{aligned}$$

$$\begin{aligned}
 &= \int_0^T p^*(t, X_t) dX_t && (\overset{*}{\Delta}R = 0) \\
 &= \text{ContRegret}(T, p^*, X).
 \end{aligned}$$

Next, recall the upper bound on R from equation (5.11):

$$R(t, x) = R_\gamma(t, x) \leq \left(\frac{\gamma}{2} + \kappa_\gamma M_0 \left(\frac{\gamma^2}{2} \right) \right) \sqrt{t} = \frac{\gamma}{2} \sqrt{t},$$

where the final equality is because γ is a root of $M_0(x^2/2)$. Putting everything together, we have

$$\text{ContRegret}(T, p^*, X) = R([X]_T, X_T) \leq \frac{\gamma}{2} \sqrt{[X]_T},$$

as desired. ■

C. Equivalent definitions of anytime regret

Let \mathcal{X} and \mathcal{L} be subsets of \mathbb{R}^n . For all $t \geq 1$, let

$$f_t: \mathcal{X}^t \times \mathcal{L}^t \rightarrow \mathbb{R}$$

be a function. Using this function we define

$$P(T) = \inf_{x_1 \in \mathcal{X}} \sup_{\ell_1 \in \mathcal{L}} \dots \inf_{x_T \in \mathcal{X}} \sup_{\ell_T \in \mathcal{L}} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T).$$

For all $t \geq 1$, let

$$A_t: \mathcal{X}^{t-1} \times \mathcal{L}^{t-1} \rightarrow \mathcal{X}$$

be a function. Using those functions, we can iteratively define the points

$$x_t = A_t(x_1, \dots, x_{t-1}, \ell_1, \dots, \ell_{t-1})$$

for all $t \geq 1$. For convenience, we have the following sequences:

$$\mathcal{A}_T = (A_1, \dots, A_T),$$

$$\mathcal{B}_T = (\ell_1, \dots, \ell_T).$$

Lastly, define

$$Q(T) = \inf_{\mathcal{A}_T} \sup_{\mathcal{B}_T} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T).$$

Proposition C.1. *The following holds: $P(T) = Q(T)$.*

Proof. First, we show $P(T) \leq Q(T)$. For fixed $x_1, \dots, x_{T-1}, \ell_1, \dots, \ell_{T-1}$, we have

$$\inf_{x_T \in \mathcal{X}} \sup_{\ell_T \in \mathcal{L}} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T) = \inf_{A_T} \sup_{\ell_T \in \mathcal{L}} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T),$$

since we may define $A_T(x_1, \dots, x_{T-1}, \ell_1, \dots, \ell_{T-1}) = x_T$. This equality holds because the function A_T is only ever evaluated at the single point

$$(x_1, \dots, x_{T-1}, \ell_1, \dots, \ell_{T-1}).$$

Next, we may take the supremum over ℓ_{T-1} to obtain

$$\begin{aligned} & \sup_{\ell_{T-1} \in \mathcal{L}} \inf_{A_T} \sup_{\ell_T \in \mathcal{L}} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T) \\ & \leq \inf_{A_T} \sup_{\ell_{T-1}, \ell_T \in \mathcal{L}} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T). \end{aligned}$$

Next, we may take the infimum over x_{T-1} . Arguing as above, we have

$$\begin{aligned} & \inf_{x_{T-1} \in \mathcal{X}} \inf_{A_T} \sup_{\ell_{T-1}, \ell_T \in \mathcal{L}} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T) \\ & = \inf_{A_{T-1}, A_T} \sup_{\ell_{T-1}, \ell_T \in \mathcal{L}} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T) \end{aligned}$$

since A_{T-1} is only ever evaluated at the single point $(x_1, \dots, x_{T-2}, \ell_1, \dots, \ell_{T-2})$. Repeating this argument proves $P(T) \leq Q(T)$.

Next we show $Q(T) \leq P(T)$. Fix $\varepsilon > 0$. There exist functions A_1, \dots, A_T such that, for all $\ell_1, \dots, \ell_T \in \mathcal{L}$, and setting $x_t = A_t(x_1, \dots, x_{t-1}, \ell_1, \dots, \ell_{t-1})$, we have

$$f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T) \leq Q(T) + \varepsilon.$$

This implies, by reordering the quantifiers,

$$\exists A_1, \forall \ell_1, \exists A_2, \forall \ell_2, \dots, \exists A_T, \forall \ell_T, \quad f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T) \leq Q(T) + \varepsilon.$$

Now defining $x_1 = A_1$, and in general $x_t = A_t(x_1, \dots, x_{t-1}, \ell_1, \dots, \ell_{t-1})$, we obtain

$$\exists x_1, \forall \ell_1, \exists x_2, \forall \ell_2, \dots, \exists x_T, \forall \ell_T, \quad f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T) \leq Q(T) + \varepsilon.$$

This implies that $P(T) \leq Q(T) + \varepsilon$. Since ε is arbitrary, the proof is complete. ■

Proposition C.2. *If \mathcal{X} and \mathcal{L} are compact and f_T is jointly continuous, then*

$$P(T) = \min_{x_1 \in \mathcal{X}} \max_{\ell_1 \in \mathcal{L}} \dots \min_{x_T \in \mathcal{X}} \max_{\ell_T \in \mathcal{L}} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T).$$

Proof sketch. For any fixed values x_1, \dots, x_T and $\ell_1, \dots, \ell_{T-1}$, the function f_T is a continuous function of ℓ_T . Compactness of \mathcal{Y} implies that the supremum is achieved, so we may write

$$\sup_{\ell_T \in \mathcal{L}} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T) = \max_{\ell_T \in \mathcal{L}} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T).$$

By Fact A.3 this is a jointly continuous function of $x_1, \dots, x_T, \ell_1, \dots, \ell_{T-1}$. Hence, for any fixed values of $x_1, \dots, x_{T-1}, \ell_1, \dots, \ell_{T-1}$ it is a continuous function of x_T . By compactness, the infimum is achieved and we may write

$$\inf_{x_T \in \mathcal{X}} \sup_{\ell_T \in \mathcal{L}} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T) = \min_{x_T \in \mathcal{X}} \max_{\ell_T \in \mathcal{L}} f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T).$$

By Fact A.3 this is a jointly continuous function of $x_1, \dots, x_{T-1}, \ell_1, \dots, \ell_{T-1}$. An inductive argument completes the proof. ■

To summarize, we can now show the equivalence of the two definitions of anytime regret. The definitions (2.2) and (2.3) are equal by Propositions C.1 and C.2 using the compact sets $\mathcal{X} = \Delta_n$, $\mathcal{L} = [0, 1]^n$ and the function

$$f_T(x_1, \dots, x_T, \ell_1, \dots, \ell_T) = \max_{1 \leq t \leq T} \frac{1}{\sqrt{t}} \left(\sum_{i=1}^t \langle x_i, \ell_i \rangle - \min_{j \in [n]} \sum_{i=1}^t \ell_{t,j} \right),$$

which is easily seen to be jointly continuous. A related discussion appears in [14, Section 2.10].

D. Comparison with some other models

In this section, we provide a brief discussion on the similarities between our model and the “prediction with expert advice” (PWEA) model and “oblivious opponent” (OO) model discussed in [14].

Comparison with our model and PWEA. We note that our model and the PWEA model defined in [14, Chapter 2] are essentially equivalent without any further structure on the models. In the PWEA model, one assumes a convex decision space \mathcal{D} , an outcome space \mathcal{Y} , a finite set \mathcal{E} of experts, and a loss function $\ell: \mathcal{D} \times \mathcal{Y} \rightarrow [0, 1]$ which is convex in the first argument. This model proceeds as follows: For each round $t = 1, 2, \dots$

- (1) The environment chooses an outcome $y_t \in \mathcal{Y}$ and expert advice $\{f_{t,E} \in \mathcal{D} : E \in \mathcal{E}\}$. The expert advice is given to the forecaster.
- (2) The forecaster chooses prediction $\hat{p}_t \in \mathcal{D}$.

- (3) The environment reveals $y_t \in \mathcal{Y}$.
- (4) The forecaster incurs loss $\ell(\hat{p}_t, y_t)$ and each expert $E \in \mathcal{E}$ incurs loss $\ell(f_{t,E}, y_t)$.

The regret incurred by the forecaster is

$$\sum_{t=1}^T \ell(\hat{p}_t, y_t) - \min_{E \in \mathcal{E}} \sum_{t=1}^T \ell(f_{t,E}, y_t).$$

The environment (resp. forecaster) in the PWEA setting is analogous to the adversary (resp. algorithm) in our setting. PWEA captures our model because one can take $\mathcal{E} = [n]$, $\mathcal{Y} = [0, 1]^n$, $\mathcal{D} = \Delta_n$ (the probability simplex on \mathbb{R}^n), and the expert advice vectors are $f_{t,i} = e_i$, where $e_i \in \mathbb{R}^n$ is i th standard basis vector. Finally, the loss function is $\ell(p, y) = \langle p, y \rangle$ which is clearly convex in p .

On the other hand, it is not hard to see that any regret bound obtained in our model translates to a bound in the PWEA model that is no larger. Indeed, reindexing so that the set of experts $\mathcal{E} = [n]$, one can take $\ell_{t,i} = \ell(f_{t,i}, y_t)$ (the left-hand side is the loss defined in our model). Next, if the algorithm (in our model) chooses x_t then the forecaster (in the PWEA model) chooses $\sum_{i=1}^n x_{t,i} f_{t,i}$. The loss incurred in our model is $\langle x_t, \ell_t \rangle$ whereas the loss incurred in the PWEA model is

$$\ell\left(\sum_{i=1}^n x_{t,i} f_{t,i}, y_t\right) \leq \sum_{i=1}^n x_{t,i} \ell(f_{t,i}, y_t) = \langle x_t, \ell_t \rangle.$$

This latter explanation is also briefly discussed in [14, Chapter 4].

Comparison with OO. Our OO discussion in Section 2.1.1 is directly inspired by the OO discussion in [14] (see the discussion after Remark 4.1 in [14]). If one views our algorithm as a randomized algorithm which chooses an expert at time t according to a probability distribution x_t then one has the deal with some subtlety in what the adversary can do. The OO model in [14] assumes that the adversary’s sequence of outcomes (which are analogous to our loss vectors) may depend on the algorithm but are chosen before the algorithm’s random choices are revealed. This is exactly analogous to our setting where the adversary is allowed to know x_t but not the realization of the algorithm’s coin flips.

References

[1] Y. Abbasi-Yadkori, P. L. Bartlett, and V. Gabillon, Near minimax optimal players for the finite-time 3-expert prediction problem. In *NIPS 2017 – Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 3037–3045, Curran Associates, 2017

- [2] J. Abernethy, R. M. Frongillo, and A. Wibisono, [Minimax option pricing meets Black–Scholes in the limit](#). In *STOC 2012 – Proceedings of the 44th ACM Symposium on Theory of Computing*, pp. 1029–1039, ACM, New York, 2012 Zbl [1286.90165](#) MR [2961563](#)
- [3] M. Abramowitz and I. A. Stegun (eds.), *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*. Dover Publications, New York, 1966 Zbl [0171.38503](#) MR [0208797](#)
- [4] A. Andoni and R. Panigrahy, A differential equations approach to optimizing regret trade-offs. 2013, arXiv:[1305.1359](#)
- [5] S. Arora, E. Hazan, and S. Kale, [The multiplicative weights update method: a meta-algorithm and applications](#). *Theory Comput.* **8** (2012), 121–164 Zbl [1283.68414](#) MR [2948502](#)
- [6] E. Bayraktar, I. Ekren, and X. Zhang, [Finite-time 4-expert prediction problem](#). *Comm. Partial Differential Equations* **45** (2020), no. 7, 714–757 Zbl [07271723](#) MR [4120922](#)
- [7] E. Bayraktar, I. Ekren, and Y. Zhang, [On the asymptotic optimality of the comb strategy for prediction with expert advice](#). *Ann. Appl. Probab.* **30** (2020), no. 6, 2517–2546 Zbl [07420064](#) MR [4187120](#)
- [8] L. Breiman, First exit times from a square root boundary. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 2*, pp. 9–16, University of California Press, Berkeley, CA, 1967 Zbl [0241.60035](#) MR [0212865](#)
- [9] L. Breiman, *Probability*. Classics Appl. Math. 7, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992 Zbl [0753.60001](#) MR [1163370](#)
- [10] M. Brezzi and T. L. Lai, [Optimal learning and experimentation in bandit problems](#). *J. Econom. Dynam. Control* **27** (2002), no. 1, 87–108 Zbl [1168.91324](#) MR [1925627](#)
- [11] S. Bubeck, Introduction to online optimization. Lecture notes, Princeton University, 2011, <http://sbubeck.com/BubeckLectureNotes.pdf>
- [12] N. Cesa-Bianchi, [Analysis of two gradient-based algorithms for on-line regression](#). *J. Comput. System Sci.* **59** (1999), no. 3, 392–411 Zbl [0961.68148](#) MR [1726769](#)
- [13] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire, and M. K. Warmuth, [How to use expert advice](#). *J. ACM* **44** (1997), no. 3, 427–485 Zbl [0890.68066](#) MR [1470152](#)
- [14] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge University Press, Cambridge, 2006 Zbl [1114.91001](#) MR [2409394](#)
- [15] K. Chaudhuri, Y. Freund, and D. J. Hsu, A parameter-free hedging algorithm. In *NIPS 2009 – Proceedings of the 22nd International Conference on Neural Information Processing Systems*, pp. 297–305, Curran Associates, 2009
- [16] H. Chernoff, Optimal stochastic control. *Sankhyā Ser. A* **30** (1968), 221–252 Zbl [0206.15901](#) MR [241149](#)
- [17] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. Third edn., MIT Press, Cambridge, MA, 2009 Zbl [1187.68679](#) MR [2572804](#)
- [18] T. M. Cover, Behavior of sequential predictors of binary sequences. In *Trans. Fourth Prague Conf. on Information Theory, Statistical Decision Functions, Random Processes (Prague, 1965)*, pp. 263–272, Academia, Prague, 1967 MR [0217944](#)

- [19] B. Davis, [On the \$L^p\$ norms of stochastic integrals and other martingales](#). *Duke Math. J.* **43** (1976), no. 4, 697–704 Zbl [0349.60061](#) MR [418219](#)
- [20] P. DeMarzo, I. Kremer, and Y. Mansour, [Online trading algorithms and robust option pricing](#). In *STOC 2006 – Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pp. 477–486, ACM, New York, 2006 Zbl [1300.91047](#) MR [2277173](#)
- [21] D. S. Donchev, [Brownian motion hitting probabilities for general two-sided square-root boundaries](#). *Methodol. Comput. Appl. Probab.* **12** (2010), no. 2, 237–245 Zbl [1195.60108](#) MR [2609650](#)
- [22] J. L. Doob, *Classical potential theory and its probabilistic counterpart*. Grundlehren Math. Wiss. 262, Springer, New York, 1984 Zbl [0549.31001](#) MR [731258](#)
- [23] N. Drenska, A PDE approach to a prediction problem involving randomized strategies. PhD thesis, New York University, 2017
- [24] N. Drenska and R. V. Kohn, [Prediction with expert advice: A PDE perspective](#). *J. Nonlinear Sci.* **30** (2020), no. 1, 137–173 Zbl [1439.35484](#) MR [4053484](#)
- [25] R. Durrett, *Probability—theory and examples*. Camb. Ser. Stat. Probab. Math. 49, Cambridge University Press, Cambridge, 2019 Zbl [1440.60001](#) MR [3930614](#)
- [26] W. Feller, *An introduction to probability theory and its applications. Vol. II*. Second edn., John Wiley, New York-London-Sydney, 1971 Zbl [0219.60003](#) MR [0270403](#)
- [27] Y. Freund, A method for Hedging in continuous time. 2009, arXiv:[0904.3356](#)
- [28] T. Fujita, [A random walk analogue of Lévy’s theorem](#). *Studia Sci. Math. Hungar.* **45** (2008), no. 2, 223–233 Zbl [1164.60365](#) MR [2417970](#)
- [29] S. Gerchinovitz, Prediction of individual sequences and prediction in the statistical framework: Some links around sparse regression and aggregation techniques. PhD thesis, Université Paris-Sud, 2011
- [30] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete mathematics*. Second edn., Addison-Wesley, Reading, MA, 1994 Zbl [0836.00001](#) MR [1397498](#)
- [31] N. Gravin, Y. Peres, and B. Sivan, [Towards optimal algorithms for prediction with expert advice](#). In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 528–547, ACM, New York, 2016 Zbl [1411.68103](#) MR [3478415](#)
- [32] N. Gravin, Y. Peres, and B. Sivan, Tight lower bounds for multiplicative weights algorithmic families. In *44th International Colloquium on Automata, Languages, and Programming*, article no. 14, Leibniz Int. Proc. Inform. 80, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017 Zbl [1441.68204](#) MR [3685788](#)
- [33] P. Greenwood and E. Perkins, A conditioned limit theorem for random walk and Brownian local time on square root boundaries. *Ann. Probab.* **11** (1983), no. 2, 227–261 Zbl [0522.60030](#) MR [690126](#)
- [34] G. Grimmett and D. Stirzaker, *Probability and random processes*. Third edn., Oxford University Press, Oxford, 2001 Zbl [1015.60002](#) MR [667520](#)
- [35] J. Hannan, Approximation to Bayes risk in repeated play. In *Contributions to the theory of games. Vol. 3*, pp. 97–139, Ann. of Math. Stud. 39, Princeton University Press, Princeton, NJ, 1957 Zbl [0078.32804](#) MR [0100249](#)
- [36] N. J. A. Harvey, C. Liaw, E. Perkins, and S. Randhawa, Optimal anytime regret with two experts. 2020, arXiv:[2002.08994](#)

- [37] A. R. Karlin and Y. Peres, *Game theory, alive*. American Mathematical Society, Providence, RI, 2017 Zbl [1416.91001](#) MR [3642417](#)
- [38] A. Klenke, *Probability theory: A comprehensive course*. Universitext, Springer, London, 2008 Zbl [1141.60001](#) MR [2372119](#)
- [39] V. A. Kobzar, R. V. Kohn, and Z. Wang, New potential-based bounds for prediction with expert advice. 2019, arXiv:[1911.01641](#)
- [40] V. A. Kobzar, R. V. Kohn, and Z. Wang, New potential-based bounds for the geometric-stopping version of prediction with expert advice. 2019, arXiv:[1912.03132](#)
- [41] R. Kudźma, *Itô's formula for a random walk*. *Lith. Math. J.*, **22** (1982), 302–306 Zbl [0522.60074](#) MR [684465](#)
- [42] N. Littlestone and M. K. Warmuth, *The weighted majority algorithm*. *Inform. and Comput.* **108** (1994), no. 2, 212–261 Zbl [0804.68121](#) MR [1265851](#)
- [43] H. Luo and R. E. Schapire, Towards minimax online learning with unknown time horizon. In *Proceedings of the 31st International Conference on Machine Learning*, pp. 226–234, Proceedings of Machine Learning Research 32, PMLR, 2014
- [44] H. Luo and R. E. Schapire, Achieving all with no parameters: AdaNormalHedge. In *Proceedings of the 28th Conference on Learning Theory*, pp. 1286–1304, Proceedings of Machine Learning Research 40, PMLR, 2015
- [45] Y. Nesterov, *Primal-dual subgradient methods for convex problems*. *Math. Program.* **120** (2009), no. 1, Ser. B, 221–259 Zbl [1191.90038](#) MR [2496434](#)
- [46] E. Perkins, *On the Hausdorff dimension of the Brownian slow points*. *Z. Wahrsch. Verw. Gebiete* **64** (1983), no. 3, 369–399 Zbl [0506.60079](#) MR [716493](#)
- [47] G. Peskir and A. Shiryaev, *Optimal stopping and free-boundary problems*. Lectures in Mathematics ETH Zürich, Birkhäuser, Basel, 2006 Zbl [1115.60001](#) MR [2256030](#)
- [48] Z. B. Reichstein, *On a property of real plane curves of even degree*. *Canad. Math. Bull.* **62** (2019), no. 1, 179–182 Zbl [1409.14094](#) MR [3943776](#)
- [49] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*. Second edn., Grundlehren Math. Wiss. 293, Springer, Berlin, 1994 Zbl [0917.60006](#) MR [1303781](#)
- [50] L. C. G. Rogers and D. Williams, *Diffusions, Markov processes, and martingales. Vol. 1*. Camb. Math. Libr., Cambridge University Press, Cambridge, 2000 Zbl [0949.60003](#) MR [1796539](#)
- [51] L. C. G. Rogers and D. Williams, *Diffusions, Markov processes, and martingales. Vol. 2*. Camb. Math. Libr., Cambridge University Press, Cambridge, 2000 Zbl [0977.60005](#) MR [1780932](#)
- [52] S. Shalev-Shwartz, *Online learning and online convex optimization*. *Found. Trends Mach. Learn.* **4** (2011), no. 2, 107–194 Zbl [1253.68190](#)
- [53] L. A. Shepp, *A first passage problem for the Wiener process*. *Ann. Math. Statist.* **38** (1967), 1912–1914 Zbl [0178.19402](#) MR [217879](#)
- [54] Vladimir G. Vovk, *Aggregating strategies*. *Colt Proceedings 1990*, pp. 371–383, Morgan Kaufmann, San Francisco, CA, 1990
- [55] D. Williams, *Probability with martingales*. Cambridge Math. Textbooks, Cambridge University Press, Cambridge, 1991 Zbl [0722.60001](#) MR [1155402](#)

Received 26 August 2021; revised 1 August 2022.

Nicholas J. A. Harvey

Department of Computer Science, University of British Columbia, 201-2366 Main Mall,
Vancouver, BC V6T 1Z4, Canada; nickhar@cs.ubc.ca

Christopher Liaw

Department of Computer Science, University of British Columbia, 201-2366 Main Mall,
Vancouver, BC V6T 1Z4, Canada; cqliaw@cs.ubc.ca

Edwin Perkins

Department of Mathematics, University of British Columbia, 1986 Mathematics Road,
Vancouver, BC V6T 1Z2, Canada; perkins@math.ubc.ca

Sikander Randhawa

Department of Computer Science, University of British Columbia, 201-2366 Main Mall,
Vancouver, BC V6T 1Z4, Canada; srand@cs.ubc.ca