Efficient and Optimal Fixed-Time Regret with Two Experts

Laura Greenstreet   LAURA.GREENSTREET@GMAIL.COM
Nicholas J. A. Harvey   NICKHAR@CS.UBC.CA
Victor Sanches Portella   VICTORSP@CS.UBC.CA
University of British Columbia, Department of Computer Science

Abstract

Prediction with expert advice is a foundational problem in online learning. In instances with T rounds and n experts, the classical Multiplicative Weights Update method suffers at most $\sqrt{(T/2)\ln n}$ regret when T is known beforehand. Moreover, this is asymptotically optimal when both T and n grow to infinity. However, when the number of experts n is small/fixed, algorithms with better regret guarantees exist. Cover showed in 1967 a dynamic programming algorithm for the two-experts problem restricted to $\{0, 1\}$ costs that suffers at most $\sqrt{T/2\pi} + O(1)$ regret with $O(T^2)$ pre-processing time. In this work, we propose an optimal algorithm for prediction with two experts' advice that works even for costs in $[0, 1]$ and with $O(1)$ processing time per turn. Our algorithm builds up on recent work on the experts problem based on techniques and tools from stochastic calculus.

Keywords: experts, Cover, online learning, optimal, fixed-time

1. Introduction

The foundational problem in online learning of *prediction with expert advice* (or simply *experts' problem*) consists of a sequential game between a player and an adversary. In each turn, the player chooses (possibly randomly) one of n experts to follow. Concurrently, the adversary chooses for each expert a cost in $[0, 1]$. At the end of a turn, the player sees the costs of all experts and suffers the cost of the expert they followed. The performance of the player is usually measured by the *regret*: the difference between their cumulative loss and the cumulative loss of the best expert in hindsight. In this case, we are interested in strategies for the player whose (expected) regret against any adversary is sublinear in the total number of rounds of the game.

A well-known strategy for the player is the *Multiplicative Weights Update* (MWU) method (Arora et al., 2012). In the *fixed-time setting* — that is, when the player knows beforehand the total number of rounds T — MWU with a carefully-chosen fixed step-size suffers at most $\sqrt{(T/2)\ln n}$ regret. Additionally, this regret bound is asymptotically optimal when both n and T grow to infinity (Cesa-Bianchi et al., 1997). Yet, if the number of experts n is fixed/small, better regret guarantees may be possible. From a theoretical standpoint, there is a clear motivation for the case where n is fixed: MWU can suffer regret arbitrarily close to $\sqrt{(T/2)\ln n}$ for *any* n as the number of rounds T grows[1] (Gravin et al., 2017). This means that different ideas are necessary for player strategies to guarantee smaller regret.

---

1. This is known to hold when MWU is used with a fixed or decreasing step-size, which are the usual cases when MWU is applied to the experts' problem. When the step-size of MWU is allowed to be arbitrary, Gravin et al. (2017) show that MWU can suffer regret arbitrarily close to $(2/3)\sqrt{(T/2)\ln n}$ as the number of rounds T grows.

1.1. The Case of Two Experts

Of course, a natural question is whether regret smaller than $\sqrt{(T/2)\ln n}$ is even possible as T grows even if n is fixed. Cover (1967) showed that for two experts (that is, for $n = 2$) a regret of $\sqrt{T/2\pi} + O(1)$ is the best possible in the worst-case by showing an algorithm for the case with costs in $\{0, 1\}$. In related work, the attainable worst-case regret bounds for three (Abbasi-Yadkori et al., 2017; Kobzar et al., 2020) and four (Bayraktar et al., 2020) experts were recently improved, respectively, to $\sqrt{8T/(9\pi)} + O(\ln T)$ and $\sqrt{T\pi/8}$ up to lower-order terms. Although optimal, Cover's algorithm is based in a dynamic programming approach that takes $O(T^2)$ time in total. In comparison, MWU takes $O(1)$ time per round to compute the probabilities to assign to the two experts at each round. Finally, one can adapt Cover's algorithm for costs in $[0, 1]$, but the standard approach is to randomly round to costs to either 0 or 1. In this case, the regret guarantees only hold in expectation.

In a related line of work, Harvey et al. (2020b) study the 2-experts problem in the *anytime setting*, that is, in the case where the player/algorithm does not know the total number of rounds T ahead of time. They showed an *optimal* strategy for the player whose regret on *any* round $t \in \mathbb{N}$ is at most $(\gamma/2)\sqrt{t}$, where the constant $\gamma \approx 1.30693$ arises naturally in the study of Brownian motion (see Mörters and Peres, 2010 for an introduction to the field and historical references). Moreover, their algorithm can be computed (up to machine precision) in $O(1)$ time since it boils down to the evaluation of well-known mathematical functions such as the exponential function and the imaginary error function. In this work, we combine similar ideas based on stochastic calculus together with Cover's algorithm to propose an efficient and optimal fixed-time algorithm for 2-experts.

Known result (Cover, 1967): There is a dynamic programming algorithm for the 2-experts problem with costs in $\{0, 1\}$ that suffers at most $\sqrt{T/2\pi} + 0.5$ regret in games with T rounds and requires $O(T^2)$ pre-processing time.

Our contribution: An algorithm for the two experts' problem with costs in $[0, 1]$ that suffers at most $\sqrt{T/2\pi} + 1.3$ regret in games with T rounds. This new algorithm has running time $O(T)$ and is based on discretizing a continuous-time solution obtained using ideas from stochastic calculus.

More precisely, one of the key steps is deriving a player in the continuous-time setting from Harvey et al. (2020b) that exploits the knowledge of the time-horizon to obtain regret bounds better than in the anytime setting. However, unlike the anytime setting, discretizing this algorithm leads to non-negative discretization error. Another key contribution of our paper is showing that this discretization error is small. Finally, the connections to Cover's classical algorithm sheds new intuition into the classical optimal solution. Interestingly, our results could be formally presented without resorting to stochastic calculus. Yet, it is the stochastic calculus point of view that guides us through the design and analysis of the algorithm.

Text organization: We first formally define the experts problem and discuss some assumptions and simplifications in Section 2. In Section 3 we present a brief summary of Cover's optimal algorithm for two experts. In Section 4 we define an analogous continuous-time problem and describe a solution inspired by Cover's algorithm. Finally, in Section 5 we present and analyze a discretized version of the continuous-time algorithm, showing it enjoys optimal worst-case regret bounds.

2. Prediction with Expert Advice

In this section we shall more precisely define the problem of prediction with expert advice. The problem is parameterized by a fixed number $n \in \mathbb{N}$ of experts. A (strategy for the) **player** is a function $\mathcal{A}$ that, given cost vectors $\ell_1, \ldots, \ell_t \in [0,1]^n$ chosen by the adversary in previous rounds, outputs a probability distribution over the n experts represented by a vector $x_{t+1} \in \Delta_n :=$ $\left\{ x \in [0,1]^n : \sum_{i=1}^n x(i) = 1 \right\}$. Similarly, a (strategy for the) **adversary** is a function $\mathcal{B}$ that, given previous player's choices $x_1, \ldots, x_t \in \Delta_n$ of distributions over the experts, outputs a vector $c_{t+1} \in [0,1]^n$ of expert costs for round $t+1$, where $\ell_{t+1}(i)$ is the cost of expert $i \in [n] := \{1, \ldots, n\}$. The performance of a player strategy $\mathcal{A}$ in a game with $T \in \mathbb{N}$ rounds against an adversary $\mathcal{B}$ is measured by the **regret**, defined as

$$\mathrm{Regret}(T, \mathcal{A}, \mathcal{B}) := \sum_{t=1}^T \ell_t^\mathsf{T} x_t - \min_{i \in [n]} \sum_{t=1}^T \ell_t(i),$$

where above, and for the remainder of this section[2] we have $x_t := \mathcal{A}(\ell_1, \ldots, \ell_{t-1})$ and $\ell_t := \mathcal{B}(x_1, \ldots, x_{t-1})$ for all $t \in [T]$. Moreover, whenever the loss vectors $\ell_1, \ldots, \ell_T$ are clear from context, we define the **cumulative loss** of expert $i \in [n]$ at round $t \in [T]$ by $L_t(i) := \sum_{j=1}^t \ell_t(i)$. In this text, for each $T \in \mathbb{N}$ we want to devise a strategy $\mathcal{A}_T$ for the player that suffers regret at most sublinear in T against *any* adversary in a game with T rounds. That is, we want a family of strategies $\{\mathcal{A}_T\}_{T \in \mathbb{N}}$ such that

$$\lim_{T \to \infty} \frac{1}{T} \sup_{\mathcal{B}} \mathrm{Regret}(T, \mathcal{A}_T, \mathcal{B}) = 0, \tag{1}$$

where the supremum ranges over all possible adversaries, even those that have full knowledge of (and may even be adversarial to) the player's strategy.

2.1. Restricted Adversaries

In (1), the supremum ranges over all the possible adversaries for a game with T rounds. However, we need only consider in the supremum **oblivious adversaries** (Karlin and Peres, 2017, Section 18.5.4), that is, adversaries $\mathcal{B}$ whose choice on each round depends *only* on the round number and not on the choices of the player. For any $\ell = (\ell_1, \ldots, \ell_T)^\mathsf{T} \in (\mathbb{R}^n)^T$, we denote by $\mathcal{B}_\ell$ the oblivious adversary that plays ℓ_t on round $t \in [T]$.

In fact, we may restrict our attention to even smaller sets of adversaries (for details on these reductions, see Gravin et al., 2016 and Karlin and Peres, 2017, Section 18.5.3). First, in (1) we need only to consider **binary adversaries**, that is, adversaries which can assign only costs in $\{0,1\}$ to the experts. Furthermore, to obtain the value of the optimal regret *for two experts* we only need to consider adversaries that pick vector costs in $\mathcal{L} := \{(1,0)^\mathsf{T}, (0,1)^\mathsf{T}\}$, which we call **restricted binary adversaries**. Intuitively, the adversary can do no better by placing equal costs on both experts at any given round. The optimal algorithm for two experts proposed by Cover (1967) heavily relies on the assumption that the adversary is a restricted binary one and does not extend to general costs in $[0,1]$ without resorting to randomly rounding the costs — which makes the regret guarantees hold only in expectation.

---

2. If no specific strategies $\mathcal{A}$ or $\mathcal{B}$ are clear from the context, one may take $\mathcal{A}$ and $\mathcal{B}$ to be arbitrary strategies and we shall omit $\mathcal{A}$ and $\mathcal{B}$ when they are clear from context.

In this work we design an algorithm the suffers at most $\sqrt{T/(2\pi)} + O(1)$ regret for arbitrary $[0, 1]$ costs. Our initial analysis handles only restricted binary adversaries, but simple concavity arguments extend the upper bound to general adversaries. Throughout this text we fix a time horizon $T \in \mathbb{N}$.

2.2. The Gap Between Experts

The case where we have only 2 experts admits a simplification that aids us greatly in the design of upper- and lower-bounds on the optimal regret. Namely, the **gap** (between experts) at round $t \in [T]$ is given by $|L_t(1) - L_t(2)|$, where L_t is the cumulative loss vector at round t as defined in Section 2. Furthermore, we denote by **lagging expert** (on round $t \in [T]$) an expert with maximum cumulative loss on round t among both experts. Similarly, we denote by **leading expert** (on round $t \in [T]$) an expert with minimum cumulative loss on round t. The following proposition from Harvey et al. (2020b) shows that, for the restricted binary adversaries described earlier, the regret can be almost fully characterized by the expert gaps and the player's choices of distributions on the experts. In the next proposition (and throughout the remainder of the text), for any predicate P we define $[P]$ to be 1 if P is true and 0 otherwise.

Proposition 1 (Harvey et al., 2020a, Proposition 2.3) *Fix $T \in \mathbb{N}$, let $\mathcal{A}$ be a player strategy, and let $\ell_1, \ldots, \ell_T \in \{(1,0)^\mathsf{T}, (0,1)^\mathsf{T}\}$ be the expert costs chosen by the adversary. For each $t \in [T]$, set $x_t := \mathcal{A}(\ell_1, \ldots, \ell_{t-1})$, let $p_t \in \{x_t(1), x_t(2)\}$ be the probability mass placed on the lagging expert on round t, and let g_t be the gap between experts on round t. Then,*

$$\mathrm{Regret}(T) = \sum_{t=1}^{T} [g_{t-1} > 0] p_t \cdot (g_t - g_{t-1}) + \sum_{t=1}^{T} [g_{t-1} = 0] \ell_t^\mathsf{T} x_t,$$

where $g_0 := 0$. In particular, if for every $t \in [T]$ with $g_{t-1} = 0$ we have $x_t(1) = x_t(2) = 1/2$, then

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

3. An Overview of Cover's Algorithm

Although in this section we give only a brief overview of Cover's algorithm, for the sake of completeness we provide a full description and analysis of the algorithm in Appendix A. The key idea in Cover's algorithm is to compute optimal decisions for all possible scenarios beforehand. This is a feasible approach when we know the total number of rounds and the adversary is a (restricted) binary adversary. More precisely, we will focus our attention to player strategies $\mathcal{A}_p$ parameterized by functions $p : [T] \times \{0, \ldots, T-1\} \to [0, 1]$ which place $p(t, g)$ probability mass on the lagging expert on round t if the gap between experts is g, and $1 - p(t, g)$ mass on the leading expert. Then the "regret-to-be-suffered" by $\mathcal{A}_p$ at any round t with a given gap between experts g is

$$V_p[t, g] := \sup\{\mathrm{Regret}(T, \mathcal{A}_p, \mathcal{B}_\ell) - \mathrm{Regret}(t, \mathcal{A}_p, \mathcal{B}_\ell) : \ell \in \mathcal{L}^T \text{ s.t. } |L_t(1) - L_t(2)| = g\}. \quad (2)$$

We can compute all entries of V_p as defined above via a dynamic programming approach, starting with $V_p[T, g]$ for all $g \in \{0, \ldots, T-1\}$ and then computing these values for earlier rounds. Moreover,

there is a simple strategy p^* that minimizes the worst-case regret $V_p[0,0]$. Interestingly, the worst-case regret of $\mathcal{A}_{p^*}$ given by $V^*[0,0]$ is tightly connected with symmetric random walks, where a **symmetric random walk** (of length t starting at g) is a sequence of random variables $(S_i)_{i=0}^t$ with $S_i := g + X_1 + \cdots + X_i$ for each $i \in \{0, \ldots, t\}$ and $\{X_j\}_{j\in[t]}$ are i.i.d. uniform random variables on $\{\pm 1\}$. The next theorem summarizes the guarantees on the regret of $\mathcal{A}_{p^*}$, showing that it suffers no more than $\sqrt{T/2\pi} + O(1)$ regret. Moreover, it is worth noting that no player strategy can do any better asymptotically in T (for a complete proof of the lower bound, see Appendix A.5).

Theorem 2 (Cover, 1967, and Karlin and Peres, 2017, Section 18.5.3) *For every $r, g \in \mathbb{N}$, let the random variable $Z_r(g)$ be the number of passages through 0 of a symmetric random walk of length r starting at position g. Then $V^*[t,g] = \frac{1}{2}\mathbf{E}[Z_{T-t}(g)]$ for every $t, g \in \mathbb{N}$. In particular,*

$$V^*[0,0] = \frac{1}{2}\mathbf{E}[Z_T(0)] \leq \sqrt{\frac{T}{2\pi}} + \frac{1}{2}.$$

Finally, although not more efficiently computable than the dynamic programming approach, p^* has a closed form solution (see Karlin and Peres, 2017, Section 18.5.3) given, for $t \in [T]$ and $g \in \{0, \ldots, T-1\}$, by

$$p^*(t,g) = \frac{1}{2}\mathbf{P}(S_{T-t} = g) + \mathbf{P}(S_{T-t} > g), \tag{3}$$

where S_{T-t} is a symmetric random walk of length $T - t$. This closed-form solution will serve as inspiration for our continuous-time algorithm.

4. A Continuous-Time Problem

Cover's player strategy is optimal, but it is defined only for restricted binary adversaries. It is likely that it can be extended to binary adversaries, but it is definitely less clear how to extend such an algorithm for general adversaries picking costs in $[0, 1]$. Moreover, even when Cover's algorithm can be used, it is quite inefficient: we either need to compute V^* which has $O(T^2)$ entries, or at each round we need to compute the probabilities in (3). In the latter case, in the first round we already need $O(T^2)$ time to exactly compute the probabilities related to a length $T - 1$ random walk.

To devise a new algorithm for the two experts problem, we first look at an analogous continuous-time problem, first proposed by Harvey et al. (2020b) and with a similar setting previously studied by Freund (2009). The main idea is to translate the random walk connection from the discrete case into a stochastic problem in continuous time, and then exploit the heavy machinery of stochastic calculus to derive a continuous time solution.

4.1. Regret as a Discrete Stochastic Integral

Let us begin by further connecting Cover's algorithm to random walks. Let $\mathcal{A}_p$ be a player strategy induced by some function $p \colon [T] \times \{0, \ldots, T-1\} \to \mathbb{R}$. If $p(t,0) = 1/2$ for all $t \in \{0, \ldots, T-1\}$, then Proposition 1 tells us that, for any restricted binary adversary sequence of gaps $g_1, \ldots, g_T \in \mathbb{R}_{\geq 0}$ and for $g_0 := 0$ we have

$$\mathrm{Regret}(T) = \sum_{t=1}^T p(t, g_{t-1})(g_t - g_{t-1}). \tag{4}$$

The right-hand side of the above equation is a discrete analog of the Riemman-Stieltjes integral of p with respect to g. In fact, if $(g_t)_{t=0}^{T}$ is a random sequence[3], the above is also known as a discrete stochastic integral. In particular, consider the case where $(g_t)_{t=0}^{T}$ is a length T reflected (i.e., absolute value of a) symmetric random walk. Then, any possible sequence of deterministic gaps has a positive probability of being realized by $(g_t)_{t=0}^{T}$. In other words, any sequence of gaps is in the support of $(g_t)_{t=0}^{T}$. Thus, bounding the worst-case regret of $\mathcal{A}_p$ is equivalent to bounding almost surely the value of (4) when $(g_t)_{t=0}^{T}$ is a reflected symmetric random walk. This idea will prove itself powerful in continuous-time even though it is not very insightful for the discrete time problem.

4.2. A Continuous-Time Problem

A stochastic process that can be seen as the continuous-time analogue of symmetric random walks is *Brownian motion* (Revuz and Yor, 1999; Mörters and Peres, 2010). We fix a Brownian motion $(B_t)_{t \geq 0}$ throughout the remainder of this text. Inspired by the observation that the discrete regret boils down to a discrete stochastic integral, Harvey et al. (2020b) define a continuous analogue of regret as a continuous stochastic integral. More specifically, given a function $p \colon [0, T] \times \mathbb{R} \to [0, 1]$ such that $p(t, 0) = 1/2$ for all $t \in \mathbb{R}_{\geq 0}$, define the **continuous regret** at time T by

$$\mathrm{ContRegret}(T, p) \coloneqq \lim_{\varepsilon \downarrow 0} \int_{0}^{T-\varepsilon} p(t, |B_t|) \, \mathrm{d}|B_t|,$$

where the term in the limit of the right-hand above is the stochastic integral (from 0 to $T - \varepsilon$) of p with respect to the process $(|B_t|)_{t \geq 0}$. We take the limit as a mere technicality: p need not be defined at time T and we want to ensure left-continuity of the continuous regret (the limit is well-defined since a stochastic integral with respect to a reflected Brownian motion is guaranteed to have limits from the left and to be continuous from the right). It is worth noting that usually stochastic integrals are defined with respect to martingales or local martingales, but $(|B_t|)_{t \geq 0}$ is neither. Still, $(|B_t|)_{t \geq 0}$ happens to be a semi-martingale, which roughly means that it can be written as a sum of two processes: a local-martingale and a process of bounded variation. In this case one can still define stochastic integrals in a way that foundational results such as Itô's formula still hold and details can be found in Revuz and Yor (1999). We do not give the precise definition of a stochastic integral since we shall not deal with its definition directly. Still, one may think intuitively of such integrals as random Riemann-Stieltjes integrals, although the precise definition of stochastic integral is more delicate.

Let us now look for a continuous function $p \colon [0, T] \times \mathbb{R} \to [0, 1]$ with $p(t, 0) = 1/2$ for all $t \geq 0$ with small continuous regret. Note that without the conditions of continuity or the requirement of $p(t, 0) = 1/2$ for $t \geq 0$, the problem would be trivial. If we did not require $p(t, 0) = 1/2$ for all $t \in [0, T]$, then taking $p(t, g) \coloneqq 0$ everywhere would yield 0 continuous regret. Moreover, dropping this requirement would go against the analogous conditions needed in the discrete case, where regret could be written as a "discrete stochastic integral" on Proposition 1 only when the player chooses $(1/2, 1/2)^{\mathsf{T}}$ in rounds with 0 gap. Finally, requiring continuity of the p is a way to avoid technicalities and "unfair" player strategies.

When working with Riemann integrals, instead of manipulating the definitions directly we use powerful and general results such as the Fundamental Theorem of Calculus (FTC). Analogously, the following result, known as *Itô's formula*, is one of the main tools we use when manipulating stochastic integrals and which can be seen as an analogue of the FTC (and shows how stochastic

---

3. We usually also require some kind of restriction on $(g_t)_{t=0}^{T}$, such as requiring it to be a martingale or a local martingale.

integrals do not always follow the classical rules of calculus). We denote by $C^{1,2}$ the class of bivariate functions that are continuously differentiable with respect to their first argument and twice continuously differentiable with respect to their second argument. Moreover, for any function $f \in C^{1,2}$ we denote by $\partial_t f$ the partial derivative of f with respect to its first argument, and we denote by $\partial_g f$ and $\partial_{gg} f$, respectively, the first and second derivatives of f with respect to its second argument.

Theorem 3 (Itô's Formula, see [Revuz and Yor, 1999](#), Theorem IV.3.3) *Let* $f \colon [0,T] \times \mathbb{R} \to \mathbb{R}$ *be in* $C^{1,2}$ *and let* $T' \in [0,T)$. *Then, almost surely,*

$$f(T', |B_{T'}|) - f(0, |B_0|) = \int_0^{T'} \partial_g f(t, |B_t|)\, \mathrm{d}|B_t| + \int_0^{T'} \left[\partial_t f(t, |B_t|) + \tfrac{1}{2}\partial_{gg} f(t, |B_t|) \right]\, \mathrm{d}t. \quad (5)$$

Note that the first integral in the equation of the above theorem resembles the definition of the continuous regret. In fact, the above result shows an alternative way to write the continuous regret at time T of a function $p \colon [0,T] \times \mathbb{R} \to [0,1]$ such that there is $R \in C^{1,2}$ with $\partial_g R = p$. However, it might be hard to compute (or even to bound) the second integral on (5). A straightforward way to circumvent this problem is to look for functions such that the second integral in (5) is 0. For that, it suffices to consider functions $R \in C^{1,2}$ that satisfy the **backwards heat equation** on $[0,T] \times \mathbb{R}$, that is,

$$\overset{*}{\Delta} R(t,g) \coloneqq \partial_t R(t,g) + \frac{1}{2}\partial_{gg} R(t,g) = 0, \qquad \forall (t,g) \in [0,T) \times \mathbb{R}. \qquad \text{(BHE)}$$

We summarize the above discussion and its implications in the following lemma.

Lemma 4 *Let* $R \colon [0,T] \times \mathbb{R} \to [0,1]$ *be in* $C^{1,2}$ *and such that* $\partial_g R(t,g) = p(t,g)$ *for all* $(t,g) \in [0,T) \times \mathbb{R}_{\geq 0}$, *such that* (BHE) *holds and such that* $R(0,0) = 0$. *Then* $\lim_{t \uparrow T} R(t, |B_t|) = \mathrm{ContRegret}(T,p)$ *almost surely.*

4.3. A Solution Inspired by Cover's Algorithm

In the remainder of this text we will make extensive use of a well-known function related to the Gaussian distribution known as **complementary error function**, defined by

$$\mathrm{erfc}(z) \coloneqq 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2}\, \mathrm{d}x = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-x^2}\, \mathrm{d}x, \qquad \forall z \in \mathbb{R}.$$

In Section 4.4 we will show that the function $Q \colon (-\infty, T) \times \mathbb{R} \to [0,1]$ in $C^{1,2}$ given by

$$Q(t,g) \coloneqq \frac{1}{2} \mathrm{erfc}\left(\frac{g}{\sqrt{2(T-t)}} \right), \qquad \forall (t,g) \in (-\infty, T) \times \mathbb{R}$$

satisfies $\mathrm{ContRegret}(T,Q) = \sqrt{T/(2\pi)}$ almost surely. Before bounding the continuous regret, it is enlightening to see how Q is related to Cover's algorithm.

Specifically, let p^* be as in (3). Due to the Central Limit Theorem, Q can be seen as an approximation of p^*. To see why, let $(S_t)_{t=0}^\infty$ be a symmetric random walk, and define $X_t \coloneqq S_t - S_{t-1}$ and $Y_t \coloneqq (X_t + 1)/2$ for each $t \geq 1$. Note that Y_t follows a Bernoulli distribution with parameter $1/2$

for any $t \geq 1$. Moreover, let Z be a Gaussian random variable with mean 0 and variance 1. Then, by setting $\mu := \mathbf{E}[2Y_1] = 1$ and $\sigma^2 := \mathbf{E}[(2Y_1 - \mu)^2] = 1$, the Central Limit Theorem guarantees

$$\frac{1}{\sqrt{L}} S_L = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} X_i = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} (2Y_i - 1) = \frac{\sqrt{L}}{\sigma} \left(\frac{1}{L} \sum_{i=1}^{L} 2Y_i - \mu \right) \xrightarrow{L \to \infty} Z,$$

where the limit holds in distribution. Thus, we roughly have that S_L and $\sqrt{L} Z$ have similar distributions. Then,

$$p^*(t, g) = \frac{1}{2} \mathbf{P}(S_{T-t} = g) + \mathbf{P}(S_{T-t} > g) \approx \frac{1}{2} \mathbf{P}((\sqrt{T-t})Z = g) + \mathbf{P}((\sqrt{T-t})Z > g)$$

$$= \mathbf{P}\left(Z > \frac{g}{\sqrt{T-t}} \right) = \frac{1}{2} \operatorname{erfc}\left(\frac{g}{\sqrt{2(T-t)}} \right) = Q(t, g).$$

One may already presume that using Q in place of p^* in the discrete experts' problem should yield a regret bound close enough to the guarantees on the regret of Cover's algorithm. Indeed, using Berry-Esseen's Theorem (Durrett, 2019, Section 3.4.4) to more precisely bound the difference between p^* and Q yields a $O(\sqrt{T})$ regret bound with suboptimal constants against binary adversaries. However, it is not clear if the approximation error would yield the optimal constant in the regret bound. Additionally, these guarantees do not naturally extend to arbitrary experts' costs in $[0, 1]$. In Section 5 we will show how to use an algorithm closely related to Q that enjoys a clean bound on the discrete-time regret.

Deriving Q directly from a PDE. We have derived Q by a heuristic argument to approximate p^*. Yet, one can derive the same solution without ever making use of p^* by approaching the problem directly from the stochastic calculus point of view. Namely, consider player strategies that satisfy the BHE, are non-negative, and that place $1/2$ mass on each expert when the gap is 0. With only these conditions we would end up with anytime solutions similar to the ones considered by Harvey et al. (2020b). In the fixed-time case we can "invert time" by a change of variables $t \leftarrow T - t$. Then the BHE becomes the traditional heat equation, which Q satisfies together with the boundary conditions.

4.4. Bounding the Continuous Regret

Interestingly, not only is Q in $C^{1,2}$, but it also satisfies the backwards heat equation, even though we have never explicitly required such a condition to hold. Since the proof of this fact boils down to technical but otherwise straightforward computations, we defer it to Appendix C.

Lemma 5 *For all $t \in [0, T)$ and $g \in \mathbb{R}_{\geq 0}$ we have $\overset{*}{\Delta} Q(t, g) = 0$.*

However, recall that to use Lemma 4 we need a function $R \in C^{1,2}$ with $\partial_g R = Q$ that satisfies the backwards heat equation, not necessarily Q itself needs to satisfy the backwards heat equation. Luckily enough, the following lemma shows how to obtain such a function R based on Q.

Proposition 6 (Harvey et al., 2020a, Lemma 5.6) *Let $h \colon [0, T) \times \mathbb{R} \to \mathbb{R}$ be in $C^{1,2}$ and define*

$$f(t, g) := \int_0^g h(t, y) \, \mathrm{d}y - \frac{1}{2} \int_0^t \partial_g h(s, 0) \, \mathrm{d}s, \qquad \forall (t, g) \in [0, T) \times \mathbb{R}.$$

Then,

(i) $f \in C^{1,2}$,

(ii) If h satisfies (BHE), then so does f,

(iii) $h = \partial_g f$.

In light of the above proposition, for all $(t, g) \in (-\infty, T) \times \mathbb{R}$ define

$$R(t, g) := \int_0^g Q(t, x) \, \mathrm{d}x - \frac{1}{2} \int_0^t \partial_g Q(s, 0) \, \mathrm{d}s.$$

In the case above, we can evaluate these integrals and obtain a formula for R that is easier to analyze. Although we defer a complete proof of the next equation to Appendix C, using that $\int_0^y \mathrm{erfc}(x) \, \mathrm{d}x = y \, \mathrm{erfc}(y) - \frac{1}{\sqrt{\pi}} e^{-y^2} + \frac{1}{\sqrt{\pi}}$ (Olver et al., 2010, Section 7.7(i)) and that $\frac{\mathrm{d}}{\mathrm{d}x} \mathrm{erfc}(x) = -\frac{2}{\sqrt{\pi}} e^{-x^2}$ we can show for every $t \in (-\infty, T)$ and $g \in \mathbb{R}$ that

$$R(t, g) = \frac{g}{2} \mathrm{erfc}\left(\frac{g}{\sqrt{2(T-t)}}\right) - \sqrt{\frac{T-t}{2\pi}} \exp\left(-\frac{g^2}{2(T-t)}\right) + \sqrt{\frac{T}{2\pi}}. \qquad (6)$$

Since R satisfies (BHE), Lemma 4 shows the continuous regret of Q is given exactly by R. The following lemma shows a bound on R and, thus, a bound on the continuous regret of Q.

Lemma 7 *We have $R(0, 0) = 0$ and*

$$R(t, g) \le \sqrt{\frac{T}{2\pi}}, \qquad \forall (t, g) \in [0, T) \times \mathbb{R}.$$

Proof The facts that $R(0, 0) = 0$ and that $R(t, g) \le 0$ for $g \le 0$ and $t \in [0, T)$ are easily verifiable. For the bound on R for $g > 0$, note first that for any $z > 0$ we have

$$\mathrm{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{x^2} \, \mathrm{d}x = \frac{2}{\sqrt{\pi}} \int_z^\infty \frac{2x}{2x} e^{x^2} \, \mathrm{d}x \le \frac{1}{z\sqrt{\pi}} \int_z^\infty 2x e^{x^2} \, \mathrm{d}x = \frac{e^{-z^2}}{z\sqrt{\pi}}.$$

Therefore, for all $(t, g) \in [0, T) \times \mathbb{R}_{>0}$ we have

$$\frac{g}{2} \mathrm{erfc}\left(\frac{g}{\sqrt{2(T-t)}}\right) \le \frac{g}{2} \cdot \frac{\sqrt{2(T-t)} \exp\left(\frac{-g^2}{2(T-t)}\right)}{g\sqrt{\pi}} = \sqrt{\frac{T-t}{2\pi}} \exp\left(\frac{-g^2}{2(T-t)}\right).$$

Applying the above to (6) yields the desired bound. ■

Combining these results we get the desired bound on the continuous regret of Q, which we summarize in the following theorem.

Theorem 8 *We have* $\mathrm{ContRegret}(T, Q) \le \sqrt{T/(2\pi)}$ *almost surely.*

5. From Continuous to Discrete Time

In the continuous time algorithm we have that $R(t, g)$ is the continuous regret at time t with gap g of the strategy that places probability mass on the lagging expert[4] according to $Q(t, g) = \partial_g R(t, g)$. At the same time, for Cover's algorithm we have $V^*[t, g]$ as an upper-bound on the regret when the mass on the lagging expert is given by $p^*(t, g)$. Furthermore, similar to the relation between Q and R, we can write p^* as a function of V^* (details can be found on Appendix A): at round t with gap g at round $t - 1$, the probability mass placed on the lagging expert in Cover's algorithm is[5]

$$p^*(t, g) = \frac{V^*[t, g - 1] - V^*[t, g + 1]}{2} \approx \partial_g V^*[t, g].$$

That is, p^* is a sort of discrete derivative of V^* with respect to its second argument. From this analogy, one might expect that a discrete derivative of R with respect to its second argument yields a good strategy for the player in the original experts' problem. As we shall see, this is exactly the case. Additionally, computing the discrete derivative of R amounts to a couple of evaluations of the complementary error function, which we can assume to be computable (up to machine precision) in constant time.

In this section we shall describe the discretized algorithm and give an upper-bound on its regret against restricted binary adversaries, that is, adversaries that choose costs in $\{(0, 1)^\mathsf{T}, (1, 0)^\mathsf{T}\}$. Luckily, unlike Cover's algorithm, the strategy we shall see in this section smoothly extends to general costs in $[0, 1]$ while preserving its performance guarantees. Since the details of this extension amounts to concavity arguments, we defer the details of this extension to Appendix E.

5.1. Discrete Itô's Formula

In Section 4, the main tool to relate the continuous regret to the function R was Itô's formula. Similarly, one of the main tools for the analysis of the discretized continuous-time algorithm will be a discrete version of Itô's formula. In order to state such a formula and to describe the algorithm, some standard notation to denote discrete derivative will be useful. Namely, for any function $f\colon \mathbb{R}^2 \to \mathbb{R}$ and any $t, g \in \mathbb{R}$, define

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

We are now in place to state a discrete analogue of Itô's formula. One important assumption of the next theorem is that $g_0, \ldots, g_T \in \mathbb{R}$ are such that successive values have absolute difference equal to 1. In the case where $g_0, \ldots, g_T$ are gaps in a 2-experts problem, this means that the adversary needs to be a restricted binary adversary. The version of the next theorem as stated — including the dependence on t — can be found in Harvey et al. (2020b, Lemma 3.7). Yet, this theorem is a slight generalization of earlier results such as the ones due to Fujita (2008, Section 2) and Kudzhma (1982, Theorem 2)

---

4. We have never formally defined lagging and leading experts in continuous time, and we do not intend to do so. Here we are extrapolating the view given by Proposition 1 of regret as a stochastic integral of the probability put on the lagging expert with respect to the gaps for the sake of intuition.

5. For $g = 0$ this does not follow directly, but our goal at the moment is only to build intuition.

Theorem 9 (Discrete Itô's Formula) *Let $g_0, g_1, \ldots, g_T \in \mathbb{R}$ be such that $|g_t - g_{t-1}| = 1$ for every $t \in [T]$ and let $f \colon \mathbb{R}^2 \to \mathbb{R}$. Then,*

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

The first summation in the right-hand side of discrete Itô's formula can be seen as a discrete stochastic integral when $(g_t)_{t=0}^{T}$ is a random sequence. Remarkably, this term is extremely similar to the regret formula from Proposition 1. Thus, if we were to use discrete Itô's formula to bound the regret, it would be desirable for the second term to (approximately) satisfy an analogue of (BHE). In fact, the potential V^* from Cover's algorithm satisfies the discrete BHE (with some care needed when the gap is zero, see Appendix A.3). Furthermore, the connection between BHE seems to extend to other problems in online learning: in recent work, Zhang et al. (2022) showed how coin-betting with potentials that satisfy the BHE yield optimal algorithms for unconstrained online learning.

Since R satisfies (BHE), one might hope that R would also satisfy such a discrete backwards-heat inequality, yielding an upper-bound on the regret of the strategy given by R_g. In the work of Harvey et al. (2020b) in the anytime setting, it was the case that the terms in the second sum were non-negative, which in a sense means that the discretized algorithm suffers *negative* discretization error. In the fixed-time setting we are not as lucky.

5.2. Discretizing the Algorithm

Based on the discussion at the beginning of this section, a natural way to discretize the algorithm from Section 4 is to define the function $q \colon [T] \times \{0, \ldots, T-1\} \to \mathbb{R}$ by

$$q(t, g) := \begin{cases} R_g(t, g) & \text{if } t < T, \\ [g = 0]\frac{1}{2} & \text{if } t = T, \end{cases} \qquad \forall t \in [T], \forall g \in \{0, \ldots, T-1\},$$

where we need to treat the case at the very last step differently since R is not defined on $\{T\} \times \mathbb{R}$. It is not clear from its definition, but we indeed have $q(t, 0) = 1/2$ for all $t \in [T]$. We defer the (relatively technical) proof of the next result to Appendix D.

Lemma 10 *We have $q(t, 0) = 1/2$ for all $t \in [T]$.*

Our goal now is to combine Proposition 1 and the discrete Itô's formula to bound the regret of $\mathcal{A}_q$. Since R satisfies the (BHE), one might hope that R is close to satisfying the discrete version of this equation. To formalize this idea, for all $t \in (-\infty, T)$ and $g \in \mathbb{R}$ define

$$r_{gg}(t, g) := \partial_{gg} R(t, g) - R_{gg}(t, g) \qquad \text{and} \qquad r_t(t, g) := \partial_t R(t, g) - R_t(t, g).$$

The above terms measure how well the first derivative with respect to the first variable and the second derivative with respect to the second variable are each approximated by their discrete analogues. That is, these are basically the discretization errors on the derivatives of R. Then, combining the fact that R satisfies (BHE) together with Proposition 1 yields the following theorem.

Theorem 11 *Consider a game of $\mathcal{A}_q$ with a restricted binary adversary with gap sequence given by $g_0, g_1, g_2, \ldots, g_T \in \{0, \ldots, T\}$ such that $g_0 = 0$ and $|g_t - g_{t-1}| = 1$ for all $t \in [T]$. Then,*

$$\text{Regret}(T) \leq \sqrt{\frac{T}{2\pi}} + \frac{1}{2} + \frac{1}{2} \sum_{t=1}^{T-1} r_{gg}(t, g_{t-1}) + \sum_{t=1}^{T-1} r_t(t, g_{t-1}). \tag{7}$$

Proof Lemma 10 and Proposition 1 yield

$$\text{Regret}(T) = \sum_{t=1}^{T} q(t, g_{t-1})(g_t - g_{t-1}) \leq \sum_{t=1}^{T-1} q(t, g_{t-1})(g_t - g_{t-1}) + \frac{1}{2}, \tag{8}$$

where in the last inequality we used that $q(T, g_{T-1}) \leq 1/2$. Furthermore, by the discrete Itô's formula (Theorem 9), we have

$$R(T-1, g_{T-1}) - R(0, g_0) = \sum_{t=1}^{T-1} R_g(t, g_{t-1})(g_t - g_{t-1}) + \sum_{t=1}^{T-1} \left(\tfrac{1}{2} R_{gg}(t, g_{t-1}) + R_t(t, g_{t-1}) \right)$$

$$\overset{\text{(BHE)}}{=} \sum_{t=1}^{T-1} q(t, g_{t-1})(g_t - g_{t-1}) - \sum_{t=1}^{T-1} \left(\tfrac{1}{2} r_{gg}(t, g_{t-1}) + r_t(t, g_{t-1}) \right)$$

$$\overset{\text{(8)}}{\geq} \text{Regret}(T) - \frac{1}{2} - \sum_{t=1}^{T-1} \left(\tfrac{1}{2} r_{gg}(t, g_{t-1}) + r_t(t, g_{t-1}) \right).$$

Rearranging and using the facts given by Lemma 7 that $R(0,0) = 0$ and that $R(T-1, g_{T-1}) \leq \sqrt{T/(2\pi)}$ yield the desired bound on the regret. ∎

5.3. Bounding the Discretization Error

In light of Theorem 11, it suffices to bound the accumulated discretization error of the derivatives to obtain potentially good bounds on the regret of $\mathcal{A}_q$. The next two lemmas show that both $r_t(t, g)$ and $r_{gg}(t, g)$ are in $O((T-t)^{-3/2})$. Since

$$\sum_{t=1}^{T-1} (T-t)^{-3/2} \leq \int_{t=0}^{T-1} (T-t)^{-3/2} \, \mathrm{d}t = 2\left(1 - \frac{1}{\sqrt{T}}\right) \leq 2, \tag{9}$$

this will show that $\mathcal{A}_q$ suffers at most $\sqrt{T/2\pi} + O(1)$ regret.[6] Since the proof of these bounds are relatively technical but otherwise not considerably insightful, we defer them to Appendix D.

Lemma 12 *For any $t \in (-\infty, T)$ and $g \in \mathbb{R}$ we have*

$$r_t(t, g) \leq \frac{\sqrt{2}}{8\sqrt{\pi}} \cdot \frac{1}{(T-t)^{3/2}} \qquad \text{and} \qquad r_{gg}(t, g) \leq \frac{2\sqrt{2}}{3\sqrt{\pi}} \cdot \frac{1}{(T-t)^{3/2}}.$$

Combining the above lemmas together with Theorem 11 yields the following regret bound.

---

6. This together with Prop. 1 also shows that the difference between in the regret of $\mathcal{A}_q$ and $\mathcal{A}_Q$ is in $O(1)$.

Theorem 13 *Define* $q(t,g) := R_g(t,g)$ *for all* $(t,g) \in \{0,\ldots,T-1\}^2$, *consider a game of* $\mathcal{A}_q$ *against a restricted binary adversary. Then,*

$$\text{Regret}(T) \leq \sqrt{\frac{T}{2\pi}} + 1.24.$$

Proof Let $g_1, g_2, \ldots, g_T \in \{0,\ldots,T\}$ be the gap sequence and set $g_0 := 0$. We have

$$\text{Regret}(T) \leq \sqrt{\frac{T}{2\pi}} + \frac{1}{2} + \frac{1}{2}\sum_{t=1}^{T-1} r_{gg}(t, g_{t-1}) + \sum_{t=1}^{T-1} r_t(t, g_{t-1}) \qquad \text{(by Theorem 11)},$$

$$\leq \sqrt{\frac{T}{2\pi}} + \frac{1}{2} + \left(\frac{\sqrt{2}}{8\sqrt{\pi}} + \frac{\sqrt{2}}{3\sqrt{\pi}}\right) \sum_{t=1}^{T-1} \frac{1}{(T-t)^{3/2}} \qquad \text{(by Lemma 12)},$$

$$\leq \sqrt{\frac{T}{2\pi}} + \frac{1}{2} + \left(\frac{\sqrt{2}}{4\sqrt{\pi}} + \frac{2\sqrt{2}}{3\sqrt{\pi}}\right) \leq \sqrt{\frac{T}{2\pi}} + 1.24, \qquad \text{(by (9))}.$$

■

6. On Optimal Regret for More than Two Experts

In this paper we presented an efficient and optimal algorithm for two experts in the fixed time-setting. A natural question is whether similar techniques can be used to find the minimax regret when we have more than two experts. Encouragingly, techniques from stochastic calculus were also used to find the optimal regret for 4 experts (Bayraktar et al., 2020). Yet, it is not clear how to use similar techniques for cases with arbitrary number of experts. The approach used in this paper and by Harvey et al. (2020b) heavily relies on the gap parameterization for the problem. Although there is an analogous parameterization of the n experts' problem into $n-1$ gaps that yields a claim similar to Proposition 1, it is not clear what would be an analogous continuous-time problem to guide us in the algorithm design process since the gap process *are not independent*—even with independent costs on the experts. Moreover, many of the approaches in related work (Abbasi-Yadkori et al., 2017; Harvey et al., 2020b; Bayraktar et al., 2020) focus on specific adversaries such as the *comb adversary*. However, the latter does not seem to be a worst-case adversary for cases such as for five experts (Chase, 2019). We are not aware of adversaries that could yield worst-case regret for arbitrary *fixed* number of experts, although asymptotically in n and T it is well-known that assigning $\{0,1\}$ costs at random is minimax optimal (Cesa-Bianchi and Lugosi, 2006)

Acknowledgments

References

Yasin Abbasi-Yadkori, Peter L. Bartlett, and Victor Gabillon. Near minimax optimal players for the finite-time 3-expert prediction problem. In *Annual Conference on Neural Information Processing Systems (NIPS)*, pages 3033–3042, 2017.

Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. 8:121–164, 2012.

Erhan Bayraktar, Ibrahim Ekren, and Xin Zhang. Finite-time 4-expert prediction problem. *Communications in Partial Differential Equations*, 45(7):714–757, 2020.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006. ISBN 978-0-521-84108-5.

Nicolò Cesa-Bianchi, Yoav Freund, David Haussler, David P. Helmbold, Robert E. Schapire, and Manfred K. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.

Zachary Chase. Experimental evidence for asymptotic non-optimality of comb adversary strategy. 12 2019. URL http://arxiv.org/abs/1912.01548.

Thomas M. Cover. Behavior of sequential predictors of binary sequences. In *Trans. Fourth Prague Conf. on Information Theory, Statistical Decision Functions, Random Processes (Prague, 1965)*, pages 263–272. Academia, Prague, 1967.

Rick Durrett. *Probability—theory and examples*, volume 49 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2019. Fifth edition.

Yoav Freund. A method for hedging in continuous time. October 2009. URL http://arxiv.org/abs/0904.3356.

Takahiko Fujita. A random walk analogue of Lévy's theorem. *Studia Sci. Math. Hungar.*, 45(2):223–233, 2008.

Nick Gravin, Yuval Peres, and Balasubramanian Sivan. Towards optimal algorithms for prediction with expert advice. In *Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 528–547. ACM, New York, 2016.

Nick Gravin, Yuval Peres, and Balasubramanian Sivan. Tight lower bounds for multiplicative weights algorithmic families. In *44th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 80, pages Art. No. 48, 14. 2017.

Nicholas J. A. Harvey, Christopher Liaw, Edwin Perkins, and Sikander Randhawa. Optimal anytime regret with two experts. February 2020a. URL https://arxiv.org/abs/2002.08994v2.

Nicholas J. A. Harvey, Christopher Liaw, Edwin A. Perkins, and Sikander Randhawa. Optimal anytime regret for two experts. In *61st IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1404–1415. IEEE, 2020b.

Anna R. Karlin and Yuval Peres. *Game theory, alive*. American Mathematical Society, Providence, RI, 2017.

Vladimir A. Kobzar, Robert V. Kohn, and Zhilei Wang. New potential-based bounds for prediction with expert advice. In *Conference on Learning Theory, (COLT)*, volume 125 of *Proceedings of Machine Learning Research*, pages 2370–2405. PMLR, 2020.

R. Kudzhma. Itô's formula for a random walk. *Litovsk. Mat. Sb.*, 22(3):122–127, 1982.

Peter Mörters and Yuval Peres. *Brownian motion*, volume 30 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2010.

Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark, editors. *NIST handbook of mathematical functions*. U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC; Cambridge University Press, Cambridge, 2010.

Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

Herbert Robbins. A remark on Stirling's formula. *Amer. Math. Monthly*, 62:26–29, 1955.

Zhiyu Zhang, Ashok Cutkosky, and Ioannis Paschalidis. PDE-based optimal strategy for unconstrained online learning. January 2022. URL https://arxiv.org/abs/2201.07877.

Appendix A. Cover's Algorithm for Two Experts

In this section, we shall review the optimal algorithm for the 2-experts problem originally proposed by Cover (1967) and the matching lower-bound.

A.1. A Dynamic Programming View

In the fixed-time setting we know the total number of rounds before the start of the game. Thus, we may compute ahead of time all the possible states the game can be on each round and decide the probabilities on the experts the player should choose in each case to minimize the worst-possible regret. More specifically, we start with a function $p\colon [T] \times \{0, \ldots, T-1\} \to [0, 1]$ that represents our player strategy: for any $t \in [T]$, if at the end of a round $t-1$ the experts' gap is $g \in \{0, \ldots, T-1\}$, on round t the player places $p(t, g)$ probability mass on the lagging[7] expert and $1 - p(t, g)$ probability on the leading expert, and we denote by $\mathcal{A}_p$ such a player strategy defined by p. Now, for all $t \in \{0, \ldots, T\}$ and $g \in \{0, \ldots, T\}$, denote by $V_p[t, g]$ the maximum regret-to-be-suffered by the player strategy defined by p on rounds $t + 1, \ldots, T$ given that at the end of round t the gap between experts is g. Slightly more formally, we have

$$V_p[t, g] := \sup\{\, \mathrm{Regret}(T, \mathcal{A}_p, \mathcal{B}_\ell) - \mathrm{Regret}(t, \mathcal{A}_p, \mathcal{B}_\ell) : \ell \in \mathcal{L}^T \text{ such that } |L_t(1) - L_t(2)| = g\}, \tag{10}$$

---

7. When the gap is 0, which means that both experts have the same cumulative loss, we break ties arbitrarily. For the optimal algorithm we shall ultimately derive this will not matter since we will have $p(t, 0) = 1/2$ for all $t \in [T]$.

where $\mathcal{L} := \{(1,0)^\mathsf{T}, (0,1)^\mathsf{T}\}$. Above we take the supremum instead of the maximum only to account for cases where the set we are considering is empty (and, thus, the supremum evaluates to $-\infty$), such as when t and g have distinct parities or when $g > t$. Note that by definition of V_p we have

$$V_p[0,0] = \max_{\ell \in \mathcal{L}^T} \mathrm{Regret}(T, \mathcal{A}_p, \mathcal{B}_\ell).$$

Thus, if we compute $V_p[0,0]$, then we have a bound on the worst-case regret of $\mathcal{A}_p$ against restricted binary adversaries. The following proposition shows how we can compute this value in a dynamic programming style.

Theorem 14 *For any $p \colon [T] \times \{0, \ldots, T-1\} \to [0,1]$, and for all $t, g \in \{0, \ldots, T\}$ such that $V_p[t,g] \neq -\infty$ we have*

$$V_p[t,g] = 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{if } t = T, \qquad (11)$$

$$V_p[t,g] = \max \begin{cases} V_p[t+1, g+1] + p(t+1, g) \\ V_p[t+1, g-1] - p(t+1, g) \end{cases} \qquad \text{if } t < T \text{ and } g > 0, \qquad (12)$$

$$V_p[t,g] = V_p[t+1, 1] + \max\{p(t+1,0), 1 - p(t+1,0)\} \qquad \text{if } t < T \text{ and } g = 0, . \qquad (13)$$

Proof First, note that (11) clearly holds by the definition of V_p. To show that equations (12) and (13) hold, let $t, g \in \{0, \ldots, T\}$ be such that $V_p[t,g] \neq -\infty$. Let $\ell \in \mathcal{L}^T$ be a sequence of cost vectors such that we have $g_t := |L_t(1) - L_t(2)| = g$. First, suppose $t < T$ and $g > 0$. Then there are two cases for ℓ: either the gap $g_{t+1} := |L_{t+1}(1) - L_{t+1}(2)|$ goes up and $g_{t+1} = g_t + 1$, or it goes down and $g_{t+1} = g_t - 1$. This together with Proposition 1 and the fact that $g_t = g > 0$ implies

$$\mathrm{Regret}(T) - \mathrm{Regret}(t) = \mathrm{Regret}(T) - \mathrm{Regret}(t+1) + p(t+1, g_t)(g_{t+1} - g_t)$$

$$= \begin{cases} \mathrm{Regret}(T) - \mathrm{Regret}(t+1) + p(t+1, g), & \text{if } g_{t+1} = g + 1 \\ \mathrm{Regret}(T) - \mathrm{Regret}(t+1) - p(t+1, g), & \text{if } g_{t+1} = g - 1. \end{cases}$$

By taking the maximum over all possible cost vectors with gap g at round t we obtain (12). Now suppose $t < T$ and $g = 0$. In this case, suppose without loss of generality that 1 is the expert to whom $\mathcal{A}_p$ assigns mass $p(t,0)$ (recall that the strategy $\mathcal{A}_p$ breaks ties arbitrarily when the gap is 0). Proposition 1 together with the fact that $\ell_t \in \mathcal{L} = \{(1,0)^\mathsf{T}, (0,1)^\mathsf{T}\}$

$$\mathrm{Regret}(T) - \mathrm{Regret}(t) = \mathrm{Regret}(T) - \mathrm{Regret}(t+1) + \ell_t^\mathsf{T} x_t$$

$$= \begin{cases} \mathrm{Regret}(T) - \mathrm{Regret}(t+1) + p(t+1, g) & \text{if } \ell_t(1) = 1, \\ \mathrm{Regret}(T) - \mathrm{Regret}(t+1) + 1 - p(t+1, g) & \text{if } \ell_t(1) = 0. \end{cases}$$

Since the gap on round $t+1$ is certainly 1 in this case, taking the maximum over all the adversaries with gap 0 on round t yields (13). ∎

For the sake of convenience, we redefine $V_p[t,g]$ for all t, g such that $V_p[t,g] = -\infty$ to, instead, be the value given by the equations from the above theorem.[8] This does not affect any of our results and makes it less cumbersome to design and analyze the algorithm.

---

8. There will be places where this definition requires access to undefined or "out-of-bounds" entries (such as for entries with gap T and time $t < T$). In such cases, we set such undefined/out-of-bounds values to 0.

A.2. Picking Optimal Probabilities

We are interested in a function $p^*\colon [T] \times \{0, \ldots, T-1\} \to [0, 1]$, if any, that minimizes $V_p[0, 0]$. To see that there is indeed such a function, note that we can formulate the problem of minimizing $V_p[0, 0]$ as a linear program using Theorem 14 to design the constraints. Such a linear program is certainly bounded (the regret is always between 0 and T) and feasible. Thus, let $p^*\colon [T] \times \{0, \ldots, T-1\} \to [0, 1]$ be a function that attains $\min_p V_p[0, 0]$ and define $V^* := V_{p^*}$. The next proposition shows that V^* can be computed recursively and show how to obtain p^* from V^*.

Theorem 15 *For each $t, g \in \{0, \ldots, T\}$*

$$
\begin{aligned}
V^*[t, g] &= 0 & &\text{if } t = T \text{ or } g = T, \\
V^*[t, g] &= \frac{1}{2}(V^*[t+1, g+1] + V^*[t+1, g-1]) & &\text{if } t < T \text{ and } 0 < g < T, \\
V^*[t, 0] &= V^*[t+1, 1] + \frac{1}{2} & &\text{if } t < T.
\end{aligned}
$$

Furthermore, if we define $p^*\colon \{0, \ldots, T\}^2 \to [0, 1]$ by

$$
p^*(t, g) := \begin{cases} \frac{1}{2}(V_{p^*}[t, g-1] - V_{p^*}[t, g+1]), & \text{if } g > 0, \\ \frac{1}{2} & \text{if } g = 0, \end{cases} \qquad \forall t \in [T], \forall g \in \{0, \ldots, T-1\},
$$

then $V_{p^*} = V^*$.

Proof Let us show that p^* as defined in the statement of the theorem attains $\inf_p V_p[0, 0]$, where the infimum ranges over all functions from $[T] \times \{0, \ldots, T-1\}$ to $[0, 1]$. Note that smaller values of any entry of V_{p^*} can only make $V_{p^*}[0, 0]$ smaller. Thus, showing that $p^*(t+1, g)$ minimizes $V_{p^*}[t, g]$ for all $t, g \in \{0, \ldots, T-1\}$ (given that $V_{p^*}[t', g']$ is fixed for $t' \geq t+1$ and $g' \in \{0, \ldots, T\}$) suffices to that p^* minimizes[9] $V_{p^*}[0.0]$. Moreover, by Theorem 14 and the definition of p^* we have that V_{p^*} obeys the formulas on the statement of this theorem. Thus, we only need to show that this choice of p^* indeed minimizes the entries of V_{p^*}.

Let us first show that

$$V_{p^*}[t, g-1] - V_{p^*}[t, g+1] \in [0, 1] \text{ for all } g \in \{1, \ldots, T-1\} \text{ and } t \in \{0, \ldots, T\}. \tag{14}$$

For $t = T$, since $V_{p^*}[T, \cdot] \equiv 0$, the above claim clearly holds. Let $t \in \{0, \ldots, T-1\}$ and $g \in \{1, \ldots, T-1\}$. If $g = 1$ we have

$$
\begin{aligned}
V_{p^*}[t, g-1] - V_{p^*}[t, g+1] &= V_{p^*}[t+1, g] + \frac{1}{2} - \frac{1}{2}(V_{p^*}[t+1, g] + V_{p^*}[t+1, g+2]) \\
&= \frac{1}{2}(V_{p^*}[t+1, g] - V_{p^*}[t+1, g+2]) + \frac{1}{2}.
\end{aligned}
$$

---

9. One may fear that choosing $p^*(t+1, g)$ to minimize $V_{p^*}[t, g]$ might increase other entries, making this argument invalid. However, note that $V_{p^*}[t, g]$ depends only on $p^*(t+1, g)$ and entries $V_{p^*}[t', g']$ with $t' > t$. Thus, it is not hard to see that by proceeding from higher to smaller values of $t \in \{0, \ldots, T\}$, we can in fact pick $p^*(t+1, g)$ to minimize $V_{p^*}[t, g]$.

The last term above, by the induction hypothesis, is in $[0, 1]$. Similarly, if $T - 1 \geq g \geq 2$ we have

$$V_{p^*}[t, g - 1] - V_{p^*}[t, g + 1]$$
$$= \frac{1}{2}(V_{p^*}[t + 1, g] + V_{p^*}[t + 1, g - 2]) - \frac{1}{2}(V_{p^*}[t + 1, g] + V_{p^*}[t + 1, g + 2])$$
$$= \frac{1}{2}(V_{p^*}[t + 1, g - 2] - V_{p^*}[t + 1, g]) + \frac{1}{2}(V_{p^*}[t + 1, g] - V_{p^*}[t + 1, g + 2]),$$

and the last term is in $[0, 1]$ by the induction hypothesis. This completes the proof of (14).

Let $t \in [T]$ and $g \in \{0, \dots, T - 1\}$. Let us now show that $p^*(t, g)$ minimizes $V_{p^*}[t, g]$ given that entries of the form $V_{p^*}[t', g']$ for $t' \geq t + 1$ and $g' \in \{0, \dots, T\}$ are fixed. For $g = 0$, Theorem 14 shows that $V_{p^*}[t, 0] = V_{p^*}[t + 1, 1] + \max\{1 - \alpha, \alpha\}$ for some $\alpha \in [0, 1]$, which is minimized when $\alpha = 1/2 = p^*(t, 0)$. For $g > 0$, Theorem 14 tells us that

$$V_{p^*}[t, g] = \max\{V_{p^*}[t + 1, g + 1] + p^*(t + 1, g), V_{p^*}[t + 1, g - 1] - p^*(t + 1, g)\}.$$

Since $p^*(t + 1, g) \geq 0$ and $V^*[t + 1, g + 1] \leq V^*[t + 1, g - 1]$ by (14), $p^*(t + 1, g)$ certainly minimizes $V^*[t, g]$ since it makes both terms in the maximum equal. Finally, (14) guarantees that[10] $p^*(t + 1, g) \in [0, 1]$. ∎

A.3. Discrete Backwards Heat Equation

Interestingly, the potential function V^* for Cover's optimal algorithm satisfies the discrete backwards heat equation when the gap is not . For simplicity, let us focus on the simpler case with $t \in [T]$ and gap $g \in \{1, \dots, T - 1\}$. Then, taking $V[t, T + 1] = 0$ we have

$$V_t^*[t, g] = V^*[t, g] - V^*[t - 1, g]$$
$$= V^*[t, g] - \frac{1}{2}(V^*[t, g + 1] + V[t, g - 1]) \qquad \text{(by Theorem 15)}$$
$$= -\frac{1}{2}(-2V^*[t, g] + V^*[t, g + 1] + V[t, g - 1])$$
$$= -\frac{1}{2}V_{gg}^*[t, g].$$

The same holds for the case where the gap is zero, but we need to extend V^* for when the gap is -1. Namely, set $V[t, -1] = V[t, 1] + 1$ for $t < T$. This guarantees that $p(t, 0) = 1/2 = (1/2)(V[t, -1] - V[t, 1])$ and $V^*[t, 0] = 1/2(V[t + 1, 1] + V[t + 1, -1])$, so the cases with zero gap agree with the formulas for non-zero gaps in Theorem 15. Interestingly, one may verify that $p^*(t, g)$ also satisfies the discrete BHE by setting $p(t, -g) = 1 - p(t, g)$ for $g \geq 0$.

A.4. Connecting the Regret with Random Walks

As argued before, to give an upper-bound on the regret of $\mathcal{A}_{p^*}$, where p^* is as in Theorem 15, we need only to bound the value of $V_{p^*}[0, 0] = V^*[0, 0]$. Interestingly, the entries of V^* have a strong connection to random walks, and this helps us give an upper-bound on the value of $V^*[0, 0]$. In the

---

10. In fact, it guarantees that $p^*(t, g) \leq 1/2$. Intuitively this makes sense since we want to give more probability to the current best/leading expert than to the lagging expert.

next theorem and for the remainder of the text, a **random walk** (of length $t \in \mathbb{N}$ starting at $g \in \mathbb{Z}_{\geq 0}$) is a sequence of random variables $(S_i)_{i=0}^t$ where $S_i := g + X_1 + \cdots + X_i$ for each $i \in \{0, \ldots, t\}$ and $\{X_j\}_{j \in [t]}$ are i.i.d. random variables taking values in $\{\pm 1\}$. If we do not specify a starting point of a random walk, take it to be 0. We say that S_t is **symmetric** if $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = -1) = 1/2$. Moreover, a **reflected random walk** (of length t) is the sequence of random variables $(|S_i|)_{i \in \{0, \ldots, t\}}$ where $(S_i)_{i=0}^t$ is a random walk. Finally, we say that a random walk $(S_i)_{i=0}^t$ **passes through** $g \in \mathbb{N}$ if the event $\{S_i = g\}$ happens for some $i \in \{0, \ldots, t\}$.

The following lemma gives numeric bounds on the expected number of passages through 0 of a symmetric random walk. Its proof boils down to careful applications of Stirling's formula and can be found in Appendix B.

Lemma 16 *Let the random variable $Z_T(0)$ be the number of passages through 0 of a reflected symmetric random walk of length T. Then,*

$$\sqrt{\frac{2T}{\pi}} + \frac{3}{5} \leq \mathbf{E}[Z_T(0)] \leq 1 + \sqrt{\frac{2T}{\pi}}.$$

We are now in position to prove an upper-bound on the performance of $\mathcal{A}_{p^*}$.

Theorem 17 *For every $r, g \in \mathbb{N}$, let the random variable $Z_r(g)$ be the number of passages through 0 of a reflected symmetric random walk of length r starting at position g. Then $V^*[t, g] = \frac{1}{2}\mathbf{E}[Z_{T-t-1}(g)]$ for every $t, g \in \mathbb{N}$. In particular,*

$$V^*[0, 0] = \frac{1}{2}\mathbf{E}[Z_{T-1}(0)] \leq \sqrt{\frac{T}{2\pi}}$$

Proof Let us show that $V^*[t, g] = (1/2)\mathbf{E}[Z_{T-t-1}(g)]$ for all $t, g \in \{0, \ldots, T\}$ by induction on $T - t$. For $t = T$ we have $Z_{T-t-1}(g) = 0$. Assume $t = T - 1$ and let $g \in \{0, \ldots, T\}$. If $g > 0$ we have $(1/2)Z_0(g) = 0 = V^*[T-1, g]$. If $g = 0$, we have $(1/2)Z_0 = 1/2 = V^*[T-1, 0]$. Suppose now $t < T - 1$. If $g = T$, then we have $V^*[t, g] = 0 = (1/2)Z_{T-t-1}(g)$. Now let us look at the case $T > g > 0$. By Theorem 15 and by the induction hypothesis, we have

$$\begin{aligned}
&V^*[t, g] \\
&= \frac{1}{2}(V^*[t+1, g+1] + V^*[t+1, g-1]) \\
&= \frac{1}{2}\left(\frac{1}{2}\mathbf{E}[Z_{T-t-2}(g+1)] + \frac{1}{2}\mathbf{E}[Z_{T-t-2}(g-1)]\right) \\
&= \frac{1}{2}\Big(\mathbf{P}(S_{T-t-2} = S_{T-t-1} + 1)\mathbf{E}[Z_{T-t-2}(g+1)] \\
&\qquad + \mathbf{P}(S_{T-t-2} = S_{T-t-1} - 1)\mathbf{E}[Z_{T-t-2}(g-1)]\Big) \\
&= \frac{1}{2}\mathbf{E}[Z_{T-t}(g)].
\end{aligned}$$

Similarly, for the case when $g = 0$ we have

$$V^*[t, 0] = V^*[t+1, 1] + \frac{1}{2} = \frac{1}{2}\mathbf{E}[Z_{T-t-2}(1) + 1] = \frac{1}{2}\mathbf{E}[Z_{T-t-1}(0)].$$

In particular, we have $V^*[0, 0] = (1/2)\mathbf{E}[Z_T(0)]$ and Lemma 16 gives us the desired numerical bound. $\blacksquare$

A.5. Lower Bound on the Optimal Regret

In the previous section we showed that Cover's algorithm suffers regret at most $\sqrt{T/(2\pi)} + O(1)$. In fact, by the definition of V_p (see (10)) and V^* we have that $V^*[0,0]$ is the minimum regret algorithms of the form $\mathcal{A}_p$, where $p\colon [T] \times \{0,\dots,T-1\} \to [0,1]$ is some function, suffer in the worst-case scenario. However, this does not tell us whether more general player strategies can do better or not. The next theorem shows that *any* player strategy suffers, in the worst case, at least $\sqrt{T/(2\pi)} - O(1)$ regret. The proof of the theorem boils down to lower-bounding the expected regret of a random adversary that plays uniformly from $\mathcal{L} = \{(0,1)^\mathsf{T}, (1,0)^\mathsf{T}\}$.

Theorem 18 *Let $\mathcal{A}$ be a player strategy for a 2-experts game with $T \in \mathbb{N}$ rounds. Then, there is $\ell \in \mathcal{L}^T$ such that*

$$\mathrm{Regret}(T, \mathcal{A}, \mathcal{B}_\ell) \geq \sqrt{\frac{T}{2\pi}} - \sqrt{\frac{1}{2\pi}} - \frac{1}{5}$$

Proof Let $\{\tilde{\ell}_t\}_{t=1}^T$ be i.i.d. random variables such that $\tilde{\ell}_t$ is equal to a vector in $\mathcal{L} = \{(0,1)^\mathsf{T}, (1,0)^\mathsf{T}\}$ chosen uniformly at random and let $\tilde{\mathcal{B}}$ be the (randomized) oblivious adversary that plays $\tilde{\ell}_t$ at round t. We shall show that

$$\mathbf{E}[\mathrm{Regret}(T, \mathcal{A}, \tilde{\mathcal{B}})] \geq \sqrt{\frac{T}{2\pi}} - \sqrt{\frac{1}{2\pi}} - \frac{1}{5}, \tag{15}$$

which implies the existence of a deterministic adversary as described in the statement. For each $t \in \{0,\dots,T\}$, let the random variable $\tilde{g}_t$ be the gap between experts due to the costs of $\tilde{\mathcal{B}}$ on round t, define $x_t := \mathcal{A}(\tilde{\ell}_1,\dots,\tilde{\ell}_{t-1})$, and set $p_t := x_t(i_t)$ where $i_t \in [2]$ is a lagging expert on round t. It is worth to already note that $(\tilde{g}_t)_{t=0}^T$ is a reflected random walk of length T. By Proposition 1 we have

$$\mathbf{E}[\mathrm{Regret}(T)] = \sum_{t=1}^T \mathbf{E}\big[[\tilde{g}_{t-1} > 0]p_t \cdot (\tilde{g}_t - \tilde{g}_{t-1})\big] + \sum_{t=1}^T \mathbf{E}\big[[\tilde{g}_{t-1} = 0]\tilde{\ell}_t^\mathsf{T} x_t\big],$$

where we recall that for any predicate P we have $[P]$ equals 1 if P is true, and equals 0 otherwise. First, let us show that

$$\mathbf{E}\big[[\tilde{g}_{t-1} > 0]p_t \cdot (\tilde{g}_t - \tilde{g}_{t-1})\big] = 0, \qquad \forall t \in [T]. \tag{16}$$

For each $t \in \{0,\dots,T\}$, define $\mathbf{E}_t[\cdot] := \mathbf{E}[\cdot | \tilde{\ell}_1,\dots,\tilde{\ell}_t]$, that is, $\mathbf{E}_t$ is the conditional expectation given the choices of the random adversary on rounds $1,\dots,t$. Let $t \in [T]$. On the event $\{\tilde{g}_{t-1} > 0\}$, one can see that $\tilde{g}_t - \tilde{g}_{t-1}$ is independent of $\tilde{\ell}_1,\dots,\tilde{\ell}_{t-1}$ and is uniformly distributed on $\{\pm 1\}$. This together with the fact that p_t is a function of $\tilde{\ell}_1,\dots,\tilde{\ell}_{t-1}$ implies

$$\begin{aligned}
\mathbf{E}\big[[\tilde{g}_{t-1} > 0]p_t \cdot (\tilde{g}_t - \tilde{g}_{t-1})\big] &= \mathbf{E}\Big[\mathbf{E}_{t-1}\big[[\tilde{g}_{t-1} > 0]p_t \cdot (\tilde{g}_t - \tilde{g}_{t-1})\big]\Big] \\
&= \mathbf{E}\Big[p_t \mathbf{E}_{t-1}\big[[\tilde{g}_{t-1} > 0](\tilde{g}_t - \tilde{g}_{t-1})\big]\Big] \\
&= \mathbf{E}\Big[p_t \mathbf{E}\big[[\tilde{g}_{t-1} > 0](\tilde{g}_t - \tilde{g}_{t-1})\big]\Big] \\
&= \mathbf{E}[p_t \cdot 0] = 0.
\end{aligned}$$

This ends the proof of (16). Let us now show that

$$\sum_{t=1}^T \mathbf{E}\big[[\tilde{g}_{t-1} = 0]\tilde{\ell}_t^\mathsf{T} x_t\big] = \frac{1}{2}\mathbf{E}[Z_{T-1}(0)]. \tag{17}$$

For each $t \in [T]$, since x_t is a function of $\tilde{\ell}_1, \ldots, \tilde{\ell}_{t-1}$ and $\tilde{\ell}_t$ is independent of $\tilde{\ell}_1, \ldots, \tilde{\ell}_{t-1}$, we have

$$
\begin{aligned}
\mathbf{E}\big[[\tilde{g}_{t-1} = 0]\tilde{\ell}_t^\mathsf{T} x_t\big] &= \mathbf{E}\Big[\mathbf{E}_{t-1}\big[[\tilde{g}_{t-1} = 0]\tilde{\ell}_t^\mathsf{T} x_t\big]\Big] \\
&= \mathbf{E}\Big[[\tilde{g}_{t-1} = 0]\mathbf{E}_{t-1}\big[\tilde{\ell}_t\big]^\mathsf{T} x_t\Big] \\
&= \mathbf{E}\Big[[\tilde{g}_{t-1} = 0]\mathbf{E}\big[\tilde{\ell}_t\big]^\mathsf{T} x_t\Big] \\
&= \mathbf{E}\Big[[g_{t-1} = 0]\Big(\frac{1}{2}x_t(1) + \frac{1}{2}x_t(2)\Big)\Big] = \frac{1}{2}\mathbf{P}(\tilde{g}_{t-1} = 0).
\end{aligned}
$$

Thus,

$$
\sum_{t=1}^{T} \mathbf{E}\big[[\tilde{g}_{t-1} = 0]\tilde{\ell}_t^\mathsf{T} x_t\big] = \frac{1}{2}\sum_{t=1}^{T} \mathbf{P}(\tilde{g}_{t-1} = 0) = \frac{1}{2}\mathbf{E}\Big[\sum_{t=1}^{T} \mathbb{1}_{\{\tilde{g}_{t-1}=0\}}\Big] = \frac{1}{2}\mathbf{E}[Z_{T-1}(0)].
$$

This completes the proof of (17) and the desired numerical lower-bound is given by Lemma 16. ∎

Appendix B. On the Passages Through Zero of a Symmetric Random Walk

In this section we shall prove Lemma 16, which bounds the expected number of passages through 0 of a symmetric random walk. First, we need a simple corollary of Stirling's formula (which we state here for convenience) to bound binomial terms.

Theorem 19 (Stirling's Formula, Robbins, 1955) *For any $n \in \mathbb{N}$ we have*

$$
\sqrt{2\pi n}\Big(\frac{n}{e}\Big)^n e^{1/(12n+1)} < n! < \sqrt{2\pi n}\Big(\frac{n}{e}\Big)^n e^{1/(12n)}.
$$

Corollary 20 *For any $n \in \mathbb{N}$ we have*

$$
\frac{2^{2n}}{\sqrt{\pi n}}\Big(1 - \frac{2}{15n}\Big) \leq \binom{2n}{n} \leq \frac{2^{2n}}{\sqrt{\pi n}}.
$$

Proof Let $n \in \mathbb{N}$. For the upper-bound, we have

$$
\begin{aligned}
\binom{2n}{n} &= \frac{2n!}{(n!)^2} \\
&< \frac{\sqrt{2\pi n} \cdot (2n)^{2n} \cdot e^{-2n} \cdot e^{1/24n}}{2\pi n \cdot n^{2n} \cdot e^{-2n} \cdot e^{2/(12n+1)}} \\
&= \frac{2^{2n}\sqrt{2}}{\sqrt{2\pi n}} \cdot \exp\Big(\frac{1}{24n} - \frac{2}{12n+1}\Big) \\
&\leq \frac{2^{2n}}{\sqrt{\pi n}} \cdot \exp\Big(-\frac{1}{24n}\Big) \leq \frac{2^{2n}}{\sqrt{\pi n}}.
\end{aligned}
$$

Similarly, for the lower-bound we have

$$
\begin{aligned}
\binom{2n}{n} &= \frac{2n!}{(n!)^2} \\
&> \frac{\sqrt{2\pi n} \cdot (2n)^{2n} \cdot e^{-2n} \cdot e^{1/(24n+1)}}{2\pi n \cdot n^{2n} \cdot e^{-2n} \cdot e^{2/12n}} \\
&= \frac{2^{2n}\sqrt{2}}{\sqrt{2\pi n}} \cdot \exp\left(\frac{1}{24n} - \frac{1}{6n}\right) \\
&= \frac{2^{2n}}{\sqrt{\pi n}} \cdot \exp\left(-\frac{4}{30n}\right) \\
&\geq \frac{2^{2n}}{\sqrt{\pi n}}\left(1 - \frac{2}{15n}\right), && \text{(Since } e^{-x} \geq 1 - x \text{ for } x \geq 0\text{).}
\end{aligned}
$$

∎

We are now ready to prove Lemma 16, which we restate for convenience.

Lemma 21 *Let the random variable $Z_T(0)$ be the number of passages through 0 of a reflected symmetric random walk of length T. Then,*

$$
\sqrt{\frac{2T}{\pi}} + \frac{3}{5} \leq \mathbf{E}[Z_T(0)] \leq 1 + \sqrt{\frac{2T}{\pi}}.
$$

Proof Let $\{S_t\}_{t=0}^T$ be a symmetric random walk and define $X_i \coloneqq S_i - S_{i-1}$ for every $i \in [T]$. Note that

$$
\mathbf{P}(|S_t| = 0) = \mathbf{P}(S_t = 0) = \mathbf{P}(|\{X_i = 1 : i \in [t]\}| = |\{X_i = -1 : i \in [t]\}|)
$$
$$
= \begin{cases} 0 & \text{if } t \text{ is odd,} \\ \binom{t}{t/2} 2^{-t} & \text{if } t \text{ is even.} \end{cases}
$$

Therefore,

$$
\mathbf{E}[Z_T(0)] = \sum_{t=0}^T \mathbf{P}(S_t = 0) = \sum_{k=0}^{\lfloor T/2 \rfloor} \mathbf{P}(S_{2k} = 0) = \sum_{k=0}^{\lfloor T/2 \rfloor} \binom{2k}{k} \frac{1}{2^{2k}}.
$$

Using Corollary 20, a consequence of Stirling's approximation to the factorial function, we can show upper- and lower-bounds to the above quantity. Namely, for the upper-bound we have

$$
\sum_{k=0}^{\lfloor T/2 \rfloor} \binom{2k}{k} \frac{1}{2^{2k}} \leq 1 + \sum_{k=1}^{\lfloor T/2 \rfloor} \frac{2^{2k}}{\sqrt{\pi k}} \frac{1}{2^{2k}} = 1 + \frac{1}{\sqrt{\pi}}\left(\sum_{k=1}^{\lfloor T/2 \rfloor} \frac{1}{\sqrt{k}}\right) \leq 1 + \frac{1}{\sqrt{\pi}}\left(\int_0^{T/2} \frac{1}{\sqrt{x}}\,\mathrm{d}x\right)
$$
$$
= 1 + \left(2\sqrt{\frac{k}{\pi}}\right)\Bigg|_{k=0}^{T/2} = 1 + \sqrt{\frac{2T}{\pi}}.
$$

We proceed similarly for the lower-bound. By setting $\beta := \sum_{k=1}^{\infty} k^{-3/2}$ we get

$$
\sum_{k=0}^{\lfloor T/2 \rfloor} \binom{2k}{k} \frac{1}{2^{2k}} \geq 1 + \sum_{k=1}^{\lfloor T/2 \rfloor} \frac{2^{2k}}{\sqrt{\pi k}} \left(1 - \frac{2}{15k}\right) \frac{1}{2^{2k}} = 1 + \frac{1}{\sqrt{\pi}} \left(\sum_{k=1}^{\lfloor T/2 \rfloor} \frac{1}{\sqrt{k}} - \sum_{k=1}^{\lfloor T/2 \rfloor} \frac{2}{15k^{3/2}} \right)
$$

$$
\leq 1 + \frac{1}{\sqrt{\pi}} \left(\int_0^{T/2} \frac{1}{\sqrt{x}} \, dx - \frac{2}{15} \sum_{k=1}^{\infty} k^{-3/2} \right) = \left(2\sqrt{\frac{k}{\pi}} \right) \Bigg|_{k=0}^{T/2} - \frac{2}{15}\beta + 1
$$

$$
= \sqrt{\frac{2T}{\pi}} - \frac{2}{15}\beta + 1.
$$

To conclude the proof, some simple calculations yield

$$
\beta = \sum_{k=1}^{\infty} \frac{1}{k^{3/2}} = 1 + \sum_{k=2}^{\infty} \frac{1}{k^{3/2}} \leq 1 + \int_1^{\infty} \frac{1}{x^{3/2}} \, dx = 1 + \left(-\frac{2}{\sqrt{x}} \right) \Bigg|_{x=1}^{\infty} = 3.
$$

■

Appendix C. Missing Proofs for Section 4

Lemma 22 *For all $t \in [0, T)$ and $g \in \mathbb{R}_{\geq 0}$ we have $\overset{*}{\Delta} Q(t, g) = 0$.*

Proof Fix $t \in [0, T)$ and $g \in \mathbb{R}_{\geq 0}$. Then,

$$
\partial_t Q(t, g) = \frac{1}{2} \partial_t \operatorname{erfc}\left(\frac{g}{\sqrt{2(T-t)}} \right) = -\frac{1}{\sqrt{\pi}} \exp\left(-\frac{g^2}{2(T-t)} \right) \partial_t \left(\frac{g}{\sqrt{2(T-t)}} \right)
$$

$$
= -\frac{1}{\sqrt{\pi}} \exp\left(-\frac{g^2}{2(T-t)} \right) \frac{g}{(2(T-t))^{3/2}}.
$$

Similarly,

$$
\partial_g Q(t, g) = \frac{1}{2} \partial_g \operatorname{erfc}\left(\frac{g}{\sqrt{2(T-t)}} \right) = -\frac{1}{\sqrt{\pi}} \exp\left(-\frac{g^2}{2(T-t)} \right) \partial_g \left(\frac{g}{\sqrt{2(T-t)}} \right)
$$

$$
= -\frac{1}{\sqrt{\pi}} \exp\left(-\frac{g^2}{2(T-t)} \right) \frac{1}{\sqrt{2(T-t)}}
$$

and

$$
\partial_{gg} Q(t, g) = -\frac{1}{\sqrt{2\pi(T-t)}} \partial_g \exp\left(-\frac{g^2}{2(T-t)} \right)
$$

$$
= \frac{1}{\sqrt{2\pi(T-t)}} \exp\left(-\frac{g^2}{2(T-t)} \right) \frac{g}{T-t} = -2\partial_t Q(t, g),
$$

as desired.

■

Let us now prove (6).

23

Lemma 23 *For all $t \in (-\infty, T)$ and $g \in \mathbb{R}$, we have*

$$R(t, g) = \frac{g}{2} \operatorname{erfc}\left(\frac{g}{\sqrt{2(T-t)}}\right) - \sqrt{\frac{T-t}{2\pi}} \exp\left(-\frac{g^2}{2(T-t)}\right) + \sqrt{\frac{T}{2\pi}}.$$

Proof Fix $(t, g) \in (-\infty, T) \times \mathbb{R}$. Using that $\int_0^g \operatorname{erfc}(x) \, \mathrm{d}x = g \operatorname{erfc}(g) - \frac{1}{\sqrt{\pi}} e^{-g^2} + \frac{1}{\sqrt{\pi}}$ (Olver et al., 2010, Section 7.7(i)), we have

$$\int_0^g Q(t, x) \, \mathrm{d}x = \frac{\sqrt{2(T-t)}}{2} \int_0^{g/\sqrt{2(T-t)}} \operatorname{erfc}(y) \, \mathrm{d}y$$

$$= \frac{\sqrt{2(T-t)}}{2} \left(x \operatorname{erfc}(x) - \frac{e^{-x^2}}{\sqrt{\pi}} \right) \Bigg|_{x=0}^{x=g/\sqrt{2(T-t)}}$$

$$= \frac{g}{2} \operatorname{erfc}\left(\frac{g}{\sqrt{2(T-t)}}\right) - \sqrt{\frac{T-t}{2\pi}} \exp\left(-\frac{g^2}{2(T-t)}\right) + \sqrt{\frac{T-t}{2\pi}}$$

for $g \geq 0$. Similarly, using that $\frac{\mathrm{d}}{\mathrm{d}x} \operatorname{erfc}(x) = -\frac{2}{\sqrt{\pi}} e^{-x^2}$ we have

$$\frac{1}{2} \int_0^t \partial_g Q(s, 0) \, \mathrm{d}s = \frac{1}{2} \int_0^t \left(-\frac{1}{\sqrt{2\pi(T-s)}} \right) \mathrm{d}s = \sqrt{\frac{T-t}{2\pi}} - \sqrt{\frac{T}{2\pi}}.$$

Plugging both equations into the definition of R concludes the proof. ∎

Appendix D. Missing Proofs for Section 5

Let us begin by proving a crucial condition on the function q defined on Section 5.

Lemma 24 *We have $q(t, 0) = 1/2$ for all $t \in [T]$.*

Proof The claim follows directly from the definition of q for $t = T$. Let $t \in \{0, \ldots, T-1\}$. From the definition of q, we have

$$q(t, 0) = R_g(t, 0) = \frac{1}{2}(R(t, 1) - R(t, -1))$$

$$= \frac{1}{2}\left(\int_0^1 Q(t, x) \, \mathrm{d}x - \frac{1}{2} \int_0^t \partial_g Q(s, 0) \, \mathrm{d}s - \int_0^{-1} Q(t, x) \, \mathrm{d}x + \frac{1}{2} \int_0^t \partial_g Q(s, 0) \, \mathrm{d}s \right)$$

$$= \frac{1}{2}\left(\int_0^1 Q(t, x) \, \mathrm{d}x - \int_0^{-1} Q(t, x) \, \mathrm{d}x \right)$$

$$= \frac{1}{2}\left(\int_0^1 Q(t, x) \, \mathrm{d}x + \int_0^1 Q(t, -x) \, \mathrm{d}x \right)$$

$$= \frac{1}{2} \int_0^1 \left(Q(t, x) + Q(t, -x) \right) \mathrm{d}x.$$

Moreover, note that for any $z \in \mathbb{R}$ we have

$$\text{erfc}(-z) = 1 - \frac{2}{\sqrt{\pi}} \int_0^{-z} e^{-x^2} \, dx = 1 + \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2} \, dx = 2 - \text{erfc}(z). \qquad (18)$$

Therefore, $Q(t, -x) = 1 - Q(t, x)$ for any $x \in \mathbb{R}$ and

$$q(t, 0) = \frac{1}{2} \int_0^1 (Q(t, x) + Q(t, -x)) \, dx = \frac{1}{2} \int_0^1 1 \, dx = \frac{1}{2}.$$

∎

This section contains the proofs on the bounds on r_t and r_{gg}. We start by bounding r_t.

Lemma 25 *For any $t \in (-\infty, T)$ and $g \in \mathbb{R}$ we have*

$$r_t(t, g) \leq \frac{\sqrt{2}}{8\sqrt{\pi}} \cdot \frac{1}{(T-t)^{3/2}} \qquad \text{and} \qquad r_{gg}(t, g) \leq \frac{2\sqrt{2}}{3\sqrt{\pi}} \cdot \frac{1}{(T-t)^{3/2}}.$$

Proof Fix $t \in (-\infty, T)$ and $g \in \mathbb{R}$. Note that R is continuously differentiable with respect to its first argument on $[t-1, t]$ and two times differentiable on $(t-1, t)$. Thus, by Taylor's Theorem, there is $t' \in (t-1, t)$ such that

$$R(t-1, g) = R(t, g) + (-1)\partial_t R(t, g) + (-1)^2 \frac{\partial_{tt} R(t', g)}{2},$$

where $\partial_{tt} R(t', g)$ denotes the second derivative of R with respect to its first argument at (t', g). Therefore,

$$r_t(t, g) = \partial_t R(t, g) - (R(t, g) - R(t-1, g)) = \frac{\partial_{tt} R(t', g)}{2}.$$

Thus, to bound $r_t(t, g)$ we need only to bound $(1/2)\partial_{tt} R(t', g)$. Computing the derivatives yields

$$\partial_t R(t', g) = \frac{1}{2\sqrt{2\pi(T-t')}} \exp\left(-\frac{g^2}{2(T-t')}\right)$$

and

$$\begin{aligned}
\partial_{tt} R(t', g) &= \frac{\sqrt{2}}{8\sqrt{\pi}(T-t')^{5/2}} \exp\left(-\frac{g^2}{2(T-t')}\right) (T - t' - g^2) \\
&\leq \frac{\sqrt{2}}{8\sqrt{\pi}(T-t')^{5/2}} \exp\left(-\frac{g^2}{2(T-t')}\right) (T - t') \\
&\leq \frac{\sqrt{2}}{8\sqrt{\pi}(T-t')^{3/2}} \qquad\qquad\qquad\qquad \text{(Since } T - t' > 0\text{),} \\
&\leq \frac{\sqrt{2}}{8\sqrt{\pi}(T-t)^{3/2}} \qquad\qquad\qquad\qquad \text{(Since } T - t' > T - t\text{).}
\end{aligned}$$

To bound r_{gg}, we will need to be slightly more careful. First, we will need the following simple lemma about Lipschitz continuity of $x \in \mathbb{R} \mapsto xe^{x^2}$.

Lemma 26 *Let $K > 0$ and define $f(\alpha) := \alpha e^{-\alpha^2/K}$ for every $\alpha \in \mathbb{R}$. Then f is 2-Lipschitz continuous.*

Proof Let $\alpha \in \mathbb{R}$. First, note that $f'(\alpha) = e^{-\alpha^2/K}(1 - 2\alpha^2/K)$. Therefore, using the fact that $e^{\beta} \geq 1 + \beta$ for any $\beta \in \mathbb{R}$ we have

$$|f'(\alpha)| = \frac{1}{\exp\left(\frac{\alpha^2}{K}\right)}\left|1 - \frac{2\alpha^2}{K}\right| \leq \frac{1}{\left|1 + \frac{\alpha^2}{K}\right|}\left|1 - \frac{2\alpha^2}{K}\right| \leq 2\frac{\left|1 - \frac{\alpha^2}{K}\right|}{\left|1 + \frac{\alpha^2}{K}\right|} \leq 2.$$

∎

We are now ready to bound r_{gg}. Fix $t \in (-\infty, T)$ and $g \in \mathbb{R}$. Moreover, denote by $\partial_g^{(3)} R$ the third partial derivative of R with respect to its third argument. By Taylor's Theorem, there are $g'_+ \in (g, g+1)$ and $g'_- \in (g-1, g)$ such that

$$R(t, g+1) = R(t, g) + \partial_g R(t, g) + \frac{1}{2}\partial_{gg} R(t, g) + \frac{1}{3!}\partial_g^{(3)} R(t, g'_+) \quad \text{and}$$

$$R(t, g-1) = R(t, g) - \partial_g R(t, g) + \frac{1}{2}\partial_{gg} R(t, g) - \frac{1}{3!}\partial_g^{(3)} R(t, g'_-).$$

Therefore,

$$r_{gg}(t, g) = \partial_{gg} R(t, g) - (R(t, g+1) + R(t, g-1) - 2R(t, g)) = \frac{1}{3!}(\partial_g^{(3)} R(t, g'_-) - \partial_g^{(3)} R(t, g'_+)).$$

Let $g' \in \mathbb{R}$. To compute the partial derivatives, first note that

$$\partial_g R(t, g') = Q(t, g') = \frac{1}{2}\operatorname{erfc}(g'/\sqrt{2(T-t)}).$$

Thus, one may check that

$$\partial_{gg} R(t, g') = -\frac{1}{2}\frac{2}{\sqrt{\pi}}\exp\left(-\frac{(g')^2}{2(T-t)}\right)\frac{1}{\sqrt{2(T-t)}}$$

$$= -\frac{1}{\sqrt{2\pi(T-t)}}\exp\left(-\frac{(g')^2}{2(T-t)}\right)$$

and $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (19)

$$\partial_g^{(3)} R(t, g') = \frac{1}{\sqrt{2\pi(T-t)}}\exp\left(-\frac{(g')^2}{2(T-t)}\right)\frac{2g'}{2(T-t)}$$

$$= \frac{1}{\sqrt{2\pi}}\frac{g'}{(T-t)^{3/2}}\exp\left(-\frac{(g')^2}{2(T-t)}\right).$$

By Lemma 26, we know that $\partial_g^{(3)} R(t, \cdot)$ is Lipschtiz continuous with Lipschitz constant $2(2\pi)^{-1/2}(T-t)^{-3/2}$. Therefore,

$$r_{gg}(t, g) = \frac{1}{3!}(\partial_g^{(3)} R(t, g'_-) - \partial_g^{(3)} R(t, g'_+)) \leq \frac{2}{3\sqrt{2\pi}(T-t)^{3/2}}|g'_- - g'_+| \leq \frac{2\sqrt{2}}{3\sqrt{\pi}(T-t)^{3/2}}.$$

∎

Appendix E. Extending the Regret Analysis for General Costs

In Section 5, we relied on the fact the gap values were in $\{+1, -1\}$. This assumption was fundamental for the version of the discrete Itô's Formula that we have uses (see the assumption on $g_0, \ldots, g_T$ in the statement of Theorem 9). It was also required by Proposition 1 to connect the regret with the "discrete stochastic integral". To extend the upper-bound on the regret of the algorithm from Section 5 to general costs we will follow the same techniques used by Harvey et al. to extend the guarantees of their algorithm to general costs: we shall use a more general version of the discrete Itô's formula, concavity of R with respect to its second argument, and a lemma relating the per-round regret to terms that appear in the more general version of the discrete Itô's formula (see Harvey et al., 2020b, Section 3.3 for details on these arguments).

As in the work of Harvey et al., we will rely on a more general version of the discrete Itô's formula that holds for general $[0, 1]$ costs. The main issue with this general formula is that more work is needed to relate it to the regret of our player strategy.

Theorem 27 (General Discrete Itô's Formula, Harvey et al., 2020a, Lemma 3.13) *Let $f \colon \mathbb{R}^2 \to \mathbb{R}$ be a function and let $g_0, g_1, \ldots, g_T \in \mathbb{R}$. Then,*

$$f(T, g_T) - f(0, g_0) = \sum_{t=1}^{T} \Big(f(t, g_t) - \frac{f(t, g_{t-1} + 1) + f(t, g_{t-1} - 1)}{2} \Big)$$
$$+ \sum_{t=1}^{T} \big(\tfrac{1}{2} f_{gg}(t, g_{t-1}) + f_t(t, g_{t-1}) \big).$$

Fix $T \in \mathbb{N}$, fix gaps $g \in \mathbb{R}^T$, and set $g_0 := 0$. For the remainder of this section all results will be regarding a game of $\mathcal{A}_q$ against an oblivious adversary with gap sequence $g_0, g_1, \ldots, g_T \in \mathbb{R}_{\geq 0}$.

For every $t \in \{0, \ldots, T\}$, define the **per-round regret** (at round t) by

$$\Delta_{\text{Regret}}(t) := \text{Regret}(t) - \text{Regret}(t - 1).$$

Our goal in this section is to prove the following lemma.

Lemma 28 *For every $t \in [T]$ we have*

$$\Delta_{\text{Regret}}(t) \leq R(t, g_t) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2}, \qquad \forall t \in [T]. \tag{20}$$

Combining the above lemma with Theorem 27 and the fact that R satisfies (BHE) yields

$$\text{Regret}(T) = \sum_{t=1}^{T} \Delta_{\text{Regret}}(t) \leq R(T-1, g_{T-1}) + \Delta_{\text{Regret}}(T) + \sum_{t=1}^{T-1} \big(\tfrac{1}{2} r_{gg}(t, g_{t-1}) + r_t(t, g_{t-1}) \big).$$

Since $q(T, g) = [g = 0](1/2)$ for any $g \in \{0, \ldots, T-1\}$, we have $\Delta_{\text{Regret}}(T) \leq 1/2$. At this point, the exact same proof of Theorem 13 applies and we obtain the same regret bound. Thus, it only remains to prove Lemma 28. In order to prove Lemma 28, we will use the following result from Harvey et al. (2020a).

Proposition 29 (Harvey et al., 2020a, Lemma 3.14) *Let g_{t-1} and g_t be the values of the gap on rounds $t-1$ and t, respectively, and let $q(t, g_{t-1})$ be the probability mass put in the worst expert at round t by the player (with $q(t, 0) = 1/2$). For all $t \geq 1$,*

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

$$\Delta_{\mathrm{Regret}}(t) = q(t, g_{t-1})(g_t - g_{t-1}).$$

2. *If a best expert at time $t-1$ remains a best expert at time t, then $g_t + g_{t-1} \leq 1$ and*

$$\Delta_{\mathrm{Regret}}(t) = g_t - q(t, g_{t-1})(g_t + g_{t-1}).$$

We shall also make use of the following fact about concave function.

Lemma 30 *Let $f \colon \mathbb{R} \to \mathbb{R}$ be a concave function and let $\alpha < \beta$ be real numbers. Then $f(x) \geq \min\{f(\alpha), f(\beta)\}$.*

Proof [Proof of Lemma 28] To prove (20), we will consider each one of the cases from Proposition 29 separately.

Case 1. In this case, (20) is equivalent to

$$0 \leq -q(t, g_{t-1})(g_t - g_{t-1}) + R(t, g_t) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2}. \tag{21}$$

Since the first term in the right-hand side of the above inequality is linear in g_t and since $R(t, \cdot)$ is concave (by (19) we know that $\partial_{gg} R(t, \cdot)$ is negative everywhere), we conclude that the whole right-hand side is concave as a function of g_t. Thus, by Fact 30 it suffices to prove the above inequality for $g_t \in \{g_{t-1} - 1, g_{t-1} + 1\}$ to prove that it holds for $g_t \in [g_{t-1} - 1, g_{t-1} + 1]$. But for $g_t \in \{g_{t-1} - 1, g_{t-1} + 1\}$ the right-hand side of (20) becomes exactly $q(t, g_{t-1})(g_t - g_{t-1})$.

Case 2. In this case, (20) is equivalent to

$$0 \leq -g_t + q(t, g_{t-1})(g_t + g_{t-1}) + R(t, g_t) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2}. \tag{22}$$

Again, the right-hand side of the above inequality is concave as a function of g_t. Since $g_t \geq 0$ and $g_t + g_{t-1} \leq 1$, we know that $g_t \in [0, 1 - g_{t-1}]$. Thus, it suffices to prove the above inequality for $g_t \in \{0, 1 - g_{t-1}\}$. For $g_t = 0$ we have

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

and in the previous case we showed that the above is non-negative for all $g_t \in [g_{t-1} - 1, g_{t-1} + 1]$. Since $g_{t-1} \leq 1$ in this case, we have in particular that the above holds for $g_t = 0$. Suppose now that $g_t = 1 - g_{t-1}$. Since $q(t, g_{t-1}) = R_g(t, g_{t-1})$, we have that (22) is equivalent to

$$\begin{aligned}
0 \leq & -g_t + q(t, g_{t-1})(g_t + g_{t-1}) + R(t, g_t) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2} \\
= & \; g_{t-1} - 1 + q(t, g_{t-1}) + R(t, 1 - g_{t-1}) - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2} \\
= & \; g_{t-1} - 1 + \frac{R(t, g_{t-1} + 1) - R(t, g_{t-1} - 1)}{2} + R(t, 1 - g_{t-1}) \\
& - \frac{R(t, g_{t-1} + 1) + R(t, g_{t-1} - 1)}{2} \\
= & \; g_{t-1} - 1 + R(t, 1 - g_{t-1}) - R(t, g_{t-1} - 1).
\end{aligned}$$

28

By the definition of R and since $\mathrm{erfc}(-z) = 2 - \mathrm{erfc}(z)$ for all $z \in \mathbb{R}$ (see (18)), we have

$$
\begin{aligned}
R(t, 1 - g_{t-1}) - R(t, g_{t-1} - 1) &= \frac{1 - g_{t-1}}{2} \, \mathrm{erfc}\left(\frac{1 - g_{t-1}}{\sqrt{2(T - t)}} \right) - \frac{g_{t-1} - 1}{2} \, \mathrm{erfc}\left(\frac{g_{t-1} - 1}{\sqrt{2(T - t)}} \right) \\
&= \left(\frac{1 - g_{t-1}}{2} \right) \left(\mathrm{erfc}\left(\frac{1 - g_{t-1}}{\sqrt{2(T - t)}} \right) + \mathrm{erfc}\left(\frac{g_{t-1} - 1}{\sqrt{2(T - t)}} \right) \right) \\
&= \left(\frac{1 - g_{t-1}}{2} \right) 2 = 1 - g_{t-1}.
\end{aligned}
$$

This concludes the proof of (22). ∎