

# Potential-based hedging algorithms

Yoav Freund

September 13, 2022

## Abstract

We study regret-minimizing online algorithms based on potential functions. First, we show that any algorithm with a regret bound that holds for any  $\epsilon$  is equivalent to a potential minimizing algorithm and vice versa. Second we should a min-max learning algorithm for known horizon. We show a regret bound that is close to optimal when the horizon is not known. Finally we give bounds on the learning algorithm in the context of Ito processes.

## 1 Introduction

Online prediction with expert advice has been studied extensively over the years and the number of publications in the area is vast (see e.g. [18, 10, 14, 4, 6]).

Here we focus on a simple variant of online prediction with expert advice called *the decision-theoretic online learning game* (DTOL) [12], we consider the *signed* version of the online game.

DTOL is a repeated zero sum game between a *learner* and an *adversary*. The adversary controls the losses of  $N$  actions, while the learner controls a distribution over the actions.

### Decision theoretic online learning

For  $t = 1, \dots, T$

1. The learner chooses a weight function  $w_j^t$  over the actions  $j \in \{1, \dots, N\}$ .
2. The adversary chooses an *instantaneous loss* for each of the  $N$  actions:  
 $l_j^t \in [-1, +1]$  for  $j \in \{1, \dots, N\}$ .
3. The *cumulative loss of action  $j$*  at time  $0 \leq t \leq T$  is  $L_j^t = \sum_{s=1}^t l_j^s$ .
4. The learner incurs an *instantaneous average loss* defined as  $\ell^t = \frac{\sum_{j=1}^N w_j^t l_j^t}{\sum_{j=1}^N w_j^t}$ .
5. The *cumulative loss of the learner* is  $L_\ell^t = \sum_{s=1}^t \ell^s$ .
6. The *cumulative regret* of the learner with respect to action  $j$  is  $R_j^t = L_\ell^t - L_j^t$ .

The goal of the learner is to perform almost as well as the best actions. Specifically, we sort the regrets in decreasing order  $R_1^t \geq R_2^t \geq \dots \geq R_k^t \geq \dots$  and define  $R_k^t$  to as the regret relative to the  $\epsilon = k/M$  top percentile, denote  $R_\epsilon^t$ . Our goal is to find algorithms that guarantee small upper bounds on  $R_\epsilon^t$ . Known bounds have the form  $c\sqrt{t \ln 1/\epsilon}$ , but the algorithm has to be tuned based on prior knowledge of  $\epsilon$ . We seek algorithms with regret bounds that hold *simultaneously* for all values of  $\epsilon$ . In other words algorithms that do not need to know  $\epsilon$  or  $t$  ahead of time.

Formally, we say that the distribution  $\Psi$  satisfied a simultaneous bound  $B$  if

**Definition 1** (Simultaneous regret bound). *Let  $G : \mathbb{R} \rightarrow [0, 1]$  be a non-increasing function which maps regret bounds to probabilities. A distribution over regrets  $\Psi$  is simultaneously bound by  $G$  if*

$$\forall r \in \mathbb{R} \quad \mathbf{P}_{\rho \sim \Psi(T)} [\rho \geq r] \leq G(r)$$

A potential function is an increasing function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ . Potential based learning algorithms are designed to bound the average potential:

**Definition 2** (Average potential bound). *A distribution over the reals  $\Psi$  satisfies the average potential function  $\phi$  if*

$$\mathbf{E}_{R \sim \Psi} [\phi(R)] \leq 1$$

Where  $\phi : \mathbb{R} \rightarrow \mathbb{R}^+$  is a non decreasing function.

We next show that there is a one to one relationship between simultaneous bounds and average potential bounds.

**Theorem 1.** *A distribution  $\Psi$  is simultaneously bounded by  $B$  if and only if it satisfies the average potential bound with  $\phi(R) = B(R)^{-1}$*

Theorem 1 allows us to focus on potential based algorithms.

As will be explained in Section ?? potential functions can be computed by backwards recursion. In other words, given  $\phi_T(R)$  and fixing the strategies for both learner and adversary we can define a potential function  $\phi_{T-1}(R)$  so that the average potentials are equal:

$$\mathbf{E}_{R \sim \Psi(T-1)} [\phi_{T-1}(R)] = \mathbf{E}_{R \sim \Psi(T)} [\phi_T(R)]$$

We use this equality to compare different strategies for both players.

We are interesting in proving min-max optimal bounds. However, as we will show, there are no matching min-max strategies for the original DTOL. To achieve min-max we extend the game by giving the adversary a larger set of moves. As the learner does not get additional options, the min/max bound for the extended game is an upper bound on the average potential in the original game.

The rest of the paper is organized as follows.

## 2 Main results

Zero-order bounds on the regret [13] depend only on  $N$  and  $T$  and typically have the form

$$\max_j R_j^T < CE\sqrt{T \ln N} \tag{1}$$

for some small constant  $C$  (typically smaller than 2). These bounds can be extended to infinite sets of experts by defining the  $\epsilon$ -regret of the algorithm as the regret with respect to the best (smallest-loss)  $\epsilon$ -percentile of the set of experts.

this replaces the bound (1) with

$$\max_j R_j^T < CE\sqrt{T \ln \frac{1}{\epsilon}} \tag{2}$$

Lower bounds have been proven that match these upper bounds up to a constant. These lower bounds typically rely on constructions in which the losses  $l_j^i$  are chosen independently at random to be either +1 or -1 with equal probabilities.

Several algorithms with refined upper bounds on the regret have been studied. Of those, the most relevant to our work is a paper by Cesa-Bianchi, Mansour and Stoltz [7] on second-order regret bounds. The bound given in Theorem 5 of [7] can be written, in our notation, as:

$$\max_j R_j^T \leq 4\sqrt{V_T \ln N} + 2 \ln N + 1/2 \tag{3}$$

Where

$$\text{Var}_i = \sum_{j=1}^N P_j^i (l_j^i)^2 - \left( \sum_{j=1}^N P_j^i l_j^i \right)^2 \quad \text{and} \quad V_T = \sum_{i=1}^T \text{Var}_i$$

A few things are worth noting. First, as  $|l_j^i| \leq 1$ ,  $\text{Var}_j \leq 1$  and therefor  $V_T \leq T$ . However  $V_T/T$  can be arbitrarily small, in which case inequality 3 provides a tighter bound than 1. Intuitively, we can say that  $V_T$  replaces  $T$  in the regret bound. This paper provides additional support for replacing  $T$  with  $V_T$  and provides lower and upper bounds on the regret involving  $V_T$ .

Initialization:

- input:  $T$  : The length of the game.
- Regret bound:  $G(r)$  the simuotaneous bound on the regret as in Definition 1.
- $\Psi(1) = \delta(0)$  is the initial state of the game which is a point mass distribution at 0.

For  $i = 1, 2, \dots, T - 1$ :

1. The learner chooses a non-negative random variable over  $\Omega$  that is the *weight function*  $P(i, R)$  such that  $\mathbf{E}_{R \sim \Psi(i)} [P(i, R)] = 1$
2. The adversary chooses a function  $Q(i, R)$  that maps  $i, R$  to a distribution over  $[-1, +1]$ . This random variable corresponds to the instantanuous loss of each action at time  $t$ .

3. We define the *bias* at  $(i, R)$  to be

$$B(i, R) \doteq \mathbf{E}_{l \sim Q(i, R)} [l] \tag{4}$$

4. the average loss is

$$\ell(i) = \mathbf{E}_{R \sim \Psi(i)} [P(i, R)B(i, R)] \tag{5}$$

5. The state is updated.

$$\Psi(i + 1) = \mathbf{E}_{R \sim \Psi(i)} [R \oplus Q(i, R)] \oplus -\ell(i) \tag{6}$$

Where  $Q(i, R)$  is the distribution of the losses of experts with respect to which the regret is  $R$  after iteration  $i - 1$ .  $\oplus$  denotes the convolution as defined above.

Figure 1: The integer time game

### 3 Preliminaries

We now formally define the integer time game. Our definition is a generalization of the standard definition in that the set of actions is not finite. Intuitively, one can think of  $N \rightarrow \infty$  while  $\epsilon$  is kept constant.

Formally, we define the set of actions to be a probability space  $(\Omega, \sigma, \mu)$  such that  $\omega \in \Omega$  is a particular action. We require that the space is *arbitrarily divisible*, which means that for any  $s \in \sigma$  and for any  $0 \leq p \leq 1$ , there exist  $u, v \in \sigma$  such that  $u \cup v = s$ ,  $\mathbf{P}[u] = p\mathbf{P}[s]$  and  $\mathbf{P}[v] = q\mathbf{P}[s]$

The *state* of the game at iteration  $i$ , denoted  $\Psi(i)$ , is a random variable which maps each action  $\omega \in \Omega$  to the cumulative loss of that action:  $L_\omega^i$ .

*Convolution:* Let  $A, B$  be two independent random variables. We define the convolution  $A \oplus B$  to be the distribution of  $x + y$ . Suppose  $R \in \mathbb{R}$ , we associate with  $R$  the random variable whose distribution is a point mass at  $R$ .

We describe the integer time game in Figure 1.

The condition on the final state  $\Psi T$  is

$$\forall r \in \mathbb{R} \quad \mathbf{P}_{\rho \sim \Psi(T)} [\rho \geq r] \leq G(r) \quad (7)$$

## 4 Potential Functions and backward induction

Theorem 1 implies that if we define a potential  $\phi(T, R) = \frac{1}{G(r)}$  then condition (7) can be replaced by

$$\mathbf{E}_{R \sim \Psi(T)} [\phi(T, R)] \leq 1 \quad (8)$$

We thereby define the *potential game* to be the integer time game where the goal of the learner is to minimize the LHS of (8) and the goal of the adversary is to maximize the same.

In the the potential game, only the final potential function is known. However, as we will now show, there is a natural way to define a potential function for all iterations.

Fix the strategies for both sides  $P(\cdot, \cdot), Q(\cdot, \cdot)$ . We define the fixed strategies potential  $\phi_{P,Q}(i, R)$  using backward induction. The base case is  $\phi_{P,Q}(T, R) = \phi(T, R)$ .

We assume that  $\phi_{P,Q}(i+1, R)$  is known and compute  $\phi_{P,Q}(i, R)$  We use our knowledge of  $P(i, R)$  and Equations (4,5) to calculate  $\ell(i)$ . Define the potential as:

$$\forall R \quad \phi_{P,Q}(i, R) \doteq \mathbf{E}_{r \sim [(R - \ell(i)) \oplus Q(i, R)]} [\phi_{P,Q}(i+1, r)] \quad (9)$$

For the potentials defined using Equation (9) we have the following theorem:

**Theorem 2.** *Assuming  $P(i, R), Q(i, R)$  are fixed for all  $i = 1, \dots, T-1$ , then*

$$\mathbf{E}_{R \sim \Psi(T)} [\phi(T, R)] = \mathbf{E}_{R \sim \Psi(T-1)} [\phi_{P,Q}(T-1, R)] = \dots = \mathbf{E}_{R \sim \Psi(1)} [\phi_{P,Q}(1, R)] = \phi_{P,Q}(1, 0)$$

*Proof.* xxx □

Theorem (2) justifies calling the distribution  $\Psi(i)$  the *state* of the game,  $\Psi(i)$  determines the final average potential, regardless of what happened before iteration  $i$ .

Next, we vary the strategies of one side or the other define the upper and lower potential functions as:

$$\phi_P^\downarrow(i, R) \doteq \sup_Q \phi_{P,Q}(i, R), \quad \phi_Q^\uparrow(i, R) \doteq \inf_P \phi_{P,Q}(i, R) \quad (10)$$

In Words,  $\phi_P^\downarrow$  is smallest upper bound on the potential that can be guaranteed by the learner strategy  $P$  while  $\phi_Q^\uparrow$  is the largest lower bound that can be guaranteed by the adversarial strategy  $Q$ . Our ultimate goal is to find a game for which

$$\forall i, R : \inf_P \phi_P^\downarrow(i, R) = \inf_P \sup_Q \phi_{P,Q}(i, R) = \sup_Q \inf_P \phi_{P,Q}(i, R) = \sup_Q \phi_Q^\uparrow(i, R) \quad (11)$$

Which means that the game has a min/max value, (but there might not be min/max strategies, instead, there are sequences of strategies that get closer and closer to min/max.

## 5 Stuff

We assume that the function is *strictly positive of degree  $k$* , which is defined as follows:

**Definition 3** (Strict Positivity of degree  $k$ ). *A function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is strictly positive of degree  $k$ , denoted  $f \in \mathcal{P}^k$  if the derivatives of orders 0 to  $k$ :  $f(x), \frac{d}{dx}f(x), \dots, \frac{d^k}{dx^k}f(x)$  exist and are strictly positive.*

A simple lemma connects an upper bound on any score function in  $\mathcal{P}^1$  with a bound on the regret.

**Lemma 3.** Let  $\phi(T, R)(\cdot) \in \mathcal{P}^1$ ,  $\Phi(T) \leq U$  and  $\epsilon = \mathbf{P}_{R \sim \Psi^t} [R > R']$  be the probability of the set of actions with respect to which the regret is larger than  $R'$ . Then

then the regret of the algorithm relative to the top  $\epsilon$  of the strategies is upper bounded by

$$\phi(T, R)(R') \leq U/\epsilon$$

For the rest of the paper we will focus on strategies for upper and lower bounding  $\Phi(T)$ . We will come back to the relationship between the bounds on  $\Phi(T)$  and the regret at the end of the paper, where we will argue that NormalHedge is, in a sense, the best possible potential function.

We will first consider the standard setting where time corresponds to the natural and is equal to the iteration number  $t_i = i$ . Later we expand the game the  $t_i$  can take on any real value in  $[0, T]$  under the condition that  $t_{i-1} \leq t_i \leq t_{i-1} + 1$ .

A common approach to design online learning algorithms is to define a *potential function*. The potential function is a continuous positive and increasing function of the regret  $R_j^i$ . A popular potential function is the exponential function which has the form  $\phi(R_j^i) = \exp(\eta R_j^i)$  where  $\eta > 0$  is the *learning rate*. The central quantity in the design and analysis of potential based online-learning algorithms is the *average potential* or *score*, defined as:

$$\Phi^t = \frac{1}{N} \sum_{j=1}^N \phi_j^t$$

Proving regret bounds are based on combining the upper and lower bounds on the score

- **Upper bound:** If the cumulative loss of the best expert is small then the score is small.
- **lower bound:** If the cumulative loss of the learning algorithm is large then the score is large.

Most of the papers on potential based online algorithms consider one or a few potential functions. Most common is the exponential potential, but others have been considered [6]. A natural question is what is the difference between potential functions and whether some potential function is “best”.

In this paper we consider a large set of potential functions, specifically, potential functions that are strictly positive and have strictly positive derivatives of orders up to four. The exponential potential and the NormalHedge potential [8, 15] are member of this set.

To analyze these potential functions we define a slightly different game, which we call a “potential game”. In this game the primary goal of the learner is not to minimize regret, rather, it is to minimize the final score  $\Phi^T$ . To do so we define potential functions for intermediate steps:  $0 \leq t < T$ .<sup>1</sup>

There are two potential functions one corresponding to the learner’s strategy, denoted  $\phi_P^\downarrow$ , which defines an upper bound on  $\Phi^T$  and one to the adversary’s strategy, denoted  $\phi_Q^\uparrow$ , which defines a lower bound on  $\Phi^T$ . We show that good strategies for the learner and for the adversary correspond to two slightly different random walks. However, the upper bound and lower bounds derived this way do not match. To find min-max matching strategies, we need to move to variable time steps.

## 6 Variable time steps

The second order bounds of [7] show that, in some contexts, the time parameter  $t$  can be replaced by a variance-type parameter  $V_T = \sum_{i=1}^T \text{Var}_i$ . This corresponds to a significant tightening of the bound. For example if instead of  $l_j^i \in [-\epsilon, +\epsilon]$ , then  $V_T = T\epsilon^2$ , and if we set  $T = \epsilon^{-2}$  we get an upper bound on the regret that does not depend on  $T$ . The idea is to replace  $T$  with  $V_T$  in the definition of the upper and lower potentials. Which is what we do in Section 8.

<sup>1</sup>The analysis described here builds on a long line of work. Including the Binomial Weights algorithm and its variants [5, 1, 2] as well as drifting games [17, 11].

In that variant of the potential game the adversary chooses the “real-time” step at each iteration. Surprisingly, it turns out that the adversary’s preference is to make the time steps as small as possible (but not zero). Combining the random walk strategy with the infinitesimal time step, we arrive at a strategy that is defined by Brownian motion, or a Wiener process. Even more surprisingly, the strategy of the learning at this limit corresponds to the same Brownian motion. In other words, we have a min-max characterization of DTOL.

**The rest of the paper is organized as follows** In Section 3 we provide some basic definitions. In Section 7 we describe the potential game in the standard integer time case and prove some upper and lower bounds for this case. In Section 8 we define a discrete time game, where time advances are real valued and finite. The core of our argument, which is based on the theory of “divided-differences” appears in Section 8.3. In Section ?? we define the continuous time limit of the game. In Section 10 we show that optimal any-time potential functions correspond to the solution of the Kolmogorov Backwards differential equation, and show the Exponential weights and NormalHedge are solutions for that Equation. Finally, in Section 11 we argue that NormalHedge is the best possible potential function.

## 7 Integer time game

We start with the standard setup in which time corresponds to the natural numbers,  $t_i = i$ . To simplify notation we will use the iteration number  $i$  instead of time.

Connecting this back to decision theoretic online learning (DTOL []). The state  $\Psi(i)$  corresponds to the distribution over the regret values of the experts on iteration  $i$ . Note that here the set of experts is allowed to be un-countably infinite. In particular the adversary can assign to the experts with regret  $x$  at iteration  $t$  an arbitrary distribution of losses in the range  $[-1, +1]$ .

The final score is the mean of the potential according to the final state, as given in Equation (??)

$$\Phi(T) = \mathbf{E}_{R \sim \Psi(T)} [\phi(T, R)(R)] \quad (12)$$

The goal of the learner is to minimize the final score and the goal of the adversary is to maximize it.

### 7.1 Results for integer time game

We prove two results for the integer game. One is a strategy for the adversary that guarantees a lower bound on  $\Phi(T)$  for any strategy of the learner, the other is a strategy for the learner that guarantees an upper bound on  $\Phi(T)$ .

We use  $B(n, s)$  to denote the binomial distribution that assigns probability  $\binom{n}{i} 2^{-n}$  to the point  $(2i - n)s$  for  $i = 0, \dots, n$

**Theorem 4.**

- *There exists a strategy for the adversary such that for any strategy of the learner,*

$$\Phi(T) \geq \mathbf{E}_{R \sim B(T, 1)} [\phi(T, R)]$$

- *There exists a strategy for the learner such that for any strategy of the adversary,*

$$\Phi(T) \leq \mathbf{E}_{R \sim B(T, 1+c)} [\phi(T, R)]$$

## 7.2 Analysis of the integer time game

The final potential  $\phi(T, R)(R)$  is part of the setup of the game. We will extend the notion of the potential to earlier steps of the game.

More specifically, we associate with any learner strategy an *upper potential* for the game and with any adversarial strategy a *lower potential* for the game. When the upper and lower potentials are equal, the associated strategies are min/max optimal.

We describe the upper potential, the lower potential description is the same other than the learner and the adversary are switched. Let  $P$  denote a learner strategy and  $Q$  denote an adversarial strategy. A legal strategy is any mapping from sequences of actions (of both players) at previous steps to an action for the current step. We will denote the set of all legal strategies for the learner by  $\mathcal{L}$ , and for the adversary by  $\mathcal{A}$ . Let  $P$  and  $Q$  be strategies for learner and adversary respectively and  $\Psi(i) = \delta(R)$  defines a game that starts at iteration  $i$  with a state that is a point mass at  $R$ . Then  $\Phi(P, Q, i, R)$  is the final score if the game starts at iteration  $i$  with state  $\Psi(i) = \delta(R)$  and the players use strategies  $P, Q$  respectively.

The upper potential for learner strategy  $P$  is defined as

$$\phi_P^\downarrow(P, i, R) = \sup_{Q \in \mathcal{A}} \Phi(P, Q, i, R)$$

## 7.3 Strategies for integer time game

Consider the states at two consecutive iterations  $\Psi(i-1), \Psi(i)$ . Suppose that  $\phi(i, R)$ , the value function for iteration  $i$  is fixed. We say that  $\phi_Q^\uparrow(i-1, R)$  is a lower bound on the value at iteration  $i-1$  if there exists an adversarial strategy  $Q^*$  such that for any strategy of the learner  $P$ ,

$$\mathbf{E}_{R \sim \Psi(i-1)} [\phi_Q^\uparrow(i-1)] \leq \mathbf{E}_{R \sim \Psi(i)} [\phi(i)]$$

Similarly,  $\phi_P^\downarrow(i-1, R)$  is an upper potential if there exists a learner strategy  $P^*$  such that for any adversarial strategy  $Q$ ,

$$\mathbf{E}_{R \sim \Psi(i-1)} [\phi_P^\downarrow(i-1)] \geq \mathbf{E}_{R \sim \Psi(i)} [\phi(i)]$$

**Lemma 5.** Suppose  $\phi_Q^\uparrow(i, R)$  is strictly convex with respect to  $R$ . Let  $\mathcal{Q}(i-1, R, B)$  be the set of adversarial strategies  $Q(i-1, R)$  that have bias  $B = B(i-1, R) = \mathbf{E}_{y \sim Q(i-1, R)} [y]$  then the strategy in  $\mathcal{Q}(i-1, R, B)$  that is best for the adversary is

$$Q^p(i-1, R) = \begin{cases} +1 & \text{w.p. } \frac{1+B}{2} \\ -1 & \text{w.p. } \frac{1-B}{2} \end{cases} \quad (13)$$

$$\phi_Q^\uparrow(i-1, R) = p\phi_Q^\uparrow(i, R+1) + (1-p)\phi_Q^\uparrow(i, R-1) \quad (14)$$

which is strictly higher than  $\phi_Q^\uparrow(i-1, R)$  for any other distribution in  $\mathcal{Q}(i-1, R, B)$ . In addition,  $\phi_Q^\uparrow(i-1, R)$  is strictly convex.

In the next lemma we describe strategies for the adversary and the learner and prove upper and lower bounds on the potential that they guarantee.

**Lemma 6.** If  $\phi(i, R)$  is CIC then

1. The adversarial strategy

$$Q^{1/2}(i-1, R) = \begin{cases} -1 & \text{w.p. } 1/2 \\ +1 & \text{w.p. } 1/2 \end{cases} \quad (15)$$

Guarantees the lower potential

$$\phi_Q^\uparrow(i-1, R) = \frac{\phi(i, R+1) + \phi(i, R-1)}{2} \quad (16)$$

2. The learner strategy:

$$P^1(i-1, R) = \frac{1}{Z} \frac{\phi(i, R+1+c) - \phi(i, R-1-c)}{2} \quad (17)$$

Where  $Z$  is a normalization factor

$$Z = \mathbf{E}_{R \sim \Psi(i)} \left[ \frac{\phi(i, R+1+c) - \phi(i, R-1-c)}{2} \right]$$

guarantees the upper potential

$$\phi_P^\downarrow(i-1, R) = \frac{\phi(i, R+1+c) + \phi(i, R-1-c)}{2} \quad (18)$$

*Proof.* 1. By symmetry adversarial strategy (15) guarantees that the aggregate loss (5) is zero regardless of the choice of the learner:  $\ell(t) = 0$ . Therefor the state update (6) is equivalent to the symmetric random walk:

$$\Psi(i) = \frac{1}{2}((\Psi(i-1) \oplus 1) + (\Psi(i-1) \ominus 1))$$

Which in turn implies that if the adversary plays  $Q^*$  and the learner plays an arbitrary strategy  $P$

$$\phi_Q^\uparrow(i-1, R) = \frac{1}{2}(\phi(i, R-1) + \phi(i, R+1)) \quad (19)$$

As this adversarial strategy is oblivious to the strategy, it guarantees that the average value at iteration  $i$  is equal to the average of the lower value at iteration  $i-1$ .

2. Plugging learner's strategy (17) into equation (5) we find that

$$\ell(i-1) = \frac{1}{Z_{i-1}} \mathbf{E}_{R \sim \Psi(i-1)} [(\phi(i, R+1+c) - \phi(i, R-1-c))B(i-1, R)] \quad (20)$$

Consider the average value at iteration  $i-1$  when the learner's strategy is  $P^*$  and the adversarial strategy is arbitrary  $Q$ :

$$\Phi_{P^*, Q}(i-1, R) = \mathbf{E}_{R \sim \Psi(i-1)} [\mathbf{E}_{y \sim Q(i-1)(R)} [\phi(i, R+y-\ell(i-1))]] \quad (21)$$

As  $\phi(i, \cdot)$  is convex and as  $(y - \ell(i-1)) \in [-1-c, 1+c]$ ,

$$\phi(i, R+y) \leq \frac{\phi(i, R+1+c) + \phi(i, R-1-c)}{2} + (y - \ell(i)) \frac{\phi(i, R+1+c) - \phi(i, R-1-c)}{2} \quad (22)$$

Combining the equations (20) and (21) we find that

$$\Phi_{P^*, Q}(i-1, R) = \mathbf{E}_{R \sim \Psi(i-1)} [\mathbf{E}_{y \sim Q(i-1)(R)} [\phi(i, R+y-\ell(i-1))]] \quad (23)$$

$$\leq \mathbf{E}_{R \sim \Psi(i-1)} \left[ \frac{\phi(i, R+1+c) + \phi(i, R-1-c)}{2} \right] \quad (24)$$

$$+ \mathbf{E}_{R \sim \Psi(i-1)} \left[ \mathbf{E}_{y \sim Q(i-1)(R)} \left[ (y - \ell(i-1)) \frac{\phi(i, R+1+c) - \phi(i, R-1-c)}{2} \right] \right] \quad (25)$$

The final step is to show that the term (25) is equal to zero. As  $\ell(i-1)$  is a constant with respect to  $R$  and  $y$  the term (25) can be written as:

$$\mathbf{E}_{R \sim \Psi(i-1)} \left[ \mathbf{E}_{y \sim Q(i-1)(R)} \left[ (y - \ell(i-1)) \frac{\phi(i+1, R+1) - \phi(i+1, R-1)}{2} \right] \right] \quad (26)$$

$$= \mathbf{E}_{R \sim \Psi(i-1)} \left[ B(i-1, R) \frac{\phi(i, R+1+c) - \phi(i, R-1-c)}{2} \right] \quad (27)$$

$$- \ell(i) \mathbf{E}_{R \sim \Psi(i-1)} \left[ \frac{\phi(i, R+1+c) - \phi(i, R-1-c)}{2} \right] \quad (28)$$

$$= 0 \quad (29)$$



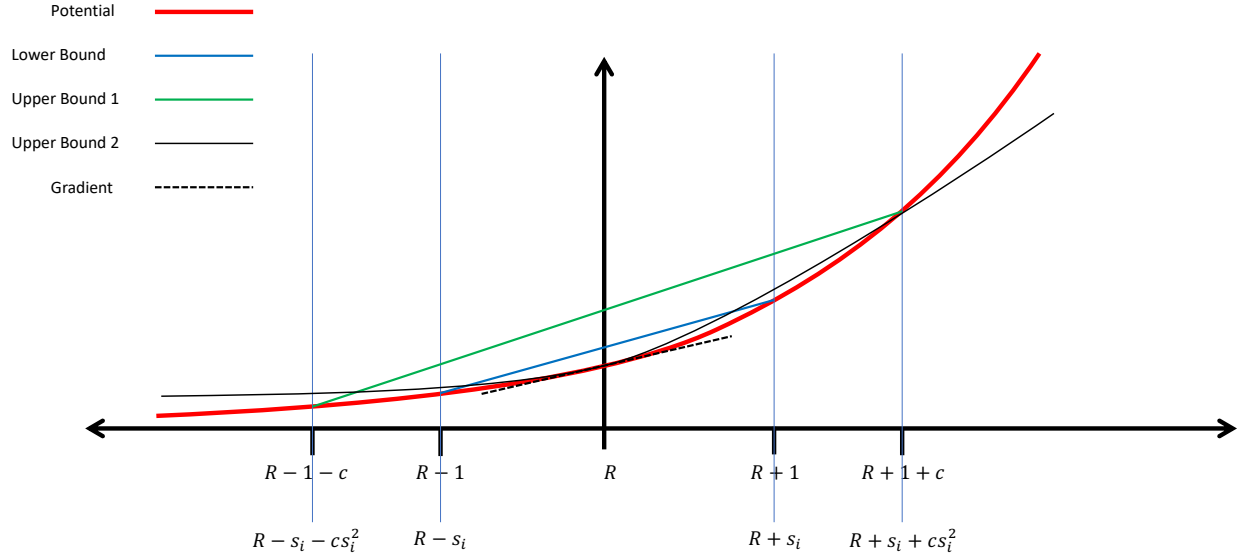


Figure 2: This figure depicts the relationship between the different upper and lower bounds used in the analysis. To aid understanding we describe the elements of the figure twice: once for the integer game, and once for the continuous time game.

**Integer time game:** let the current iteration be  $i$  and let the current regret be  $R$ . Let  $r$  be the regret at iteration  $i + 1$ , we have that  $R - 1 - c \leq r \leq R + 1 + c$ . The potential at iteration  $i + 1$  is  $\phi(i + 1, r)$  (the red curve). The lower bound (blue line) corresponds to the adversarial strategy:  $Q^{1/2}(i, R)$ . The first-order learner strategy:  $P^1(i, R)$  corresponds to the green line. The second-order learner strategy:  $P^2(i, R)$  corresponds to the black curve.

**Continuous time game:** let the current iteration be  $i$ , the current time be  $t_i$  and the current regret be  $R$ . Let  $0 < s_i \leq 1$  be the step size chosen by the adversary, so that the next time is  $t_{i+1} = t_i + s_i^2$ . Let  $r$  be the regret at iteration  $i + 1$ , we have that  $R - s_i - cs_i^2 \leq r \leq R + s_i + cs_i^2$ . The potential at iteration  $i + 1$  is  $\phi(t_{i+1}, r)$  (the red curve). The lower bound (blue line) corresponds to the adversarial strategy:  $Q_{\pm s_i}^{1/2}(t_i, R)$ . The first-order learner strategy:  $P^{1c}(t_i, R)$  corresponds to the green line. The second-order learner strategy:  $P^{2c}(t_i, R)$  corresponds to the black curve. Observe that when  $s_i \rightarrow 0$  the ratio  $\frac{s_i}{s_i + cs_i^2}$  converges to 1, and the upper and gap between the green and blue lines converges to zero.

□

We find that the lower bound corresponds to an unbiased random walk with step size  $\pm 1$ . The upper bound also corresponds to an unbiased random walk with step size  $\pm(1 + c)$ . The natural setting in the natural game is  $c = 1$ , which means that there is a significant difference between the upper and lower bounds. As we show in the next section, this gap converges to zero in the continuous time setting, and the upper and lower bounds match, making the strategies for both sides min-max optimal.

Note also that the adversarial strategy the aggregate loss  $\ell(t)$  is always zero, regardless of the strategy of the learner, and state progression is independent of the learner's choices.

## 7.4 A learner strategy with a variance-dependent bound

As shown in Lemma 5, the adversary always prefers mixed strategies that assign zero probability for all steps other than  $\pm 1$ . Suppose, however, that the adversary is not worst-case optimal and chooses steps whose length is less than one. The following lemma gives a slightly different strategy for the learner, which

guarantees a tighter bound for this case.

**Lemma 7.** *The learner strategy:*

$$P^2(i-1, R) = \frac{1}{Z} \left. \frac{\partial}{\partial r} \right|_{r=R} \phi(i, r) \quad (30)$$

Where  $Z$  is a normalization factor

$$Z = \mathbf{E}_{R \sim \Psi(i)} \left[ \left. \frac{\partial}{\partial r} \right|_{r=R} \phi(i, r) \right]$$

guarantees the following upper potential against any adversarial strategy  $Q$

$$\phi_P^\downarrow(i-1, R) = \phi(i, R) + b(i, R) \mathbf{E}_{l \sim Q(i, R)} [l^2] \quad (31)$$

where  $b(i, R) = \phi(i, R+1+c) - \phi(i, R) - (1+c) \left. \frac{\partial}{\partial r} \right|_{r=R} \phi(i, r)$

We compare the bound for  $P^2$  to the bound for  $P^1$  given in Lemma 6. We find that when the adversary is optimal:  $\mathbf{E}_{l \sim Q(i, R)} [l^2] = 1$  then the bound for  $P^1$  is better than the bound for  $P^2$ , on the other hand, when  $\mathbf{E}_{l \sim Q(i, R)} [l^2]$  is close to zero,  $P^2$  is better than  $P^1$ .

## 8 From integer to discrete time

We start with motivation for using time that is indexed by real values rather than the natural numbers. We distinguish between two notions of time. The first notion of time is a counter that counts the iterations of the game, we will call this counter the *iteration counter* and denote it by  $i = 0, 1, \dots$ . The second, more interesting notion of time is the time that appears in the regret bounds, we denote this time by  $t_i$  where  $i$  is the iteration number. We restrict the time increments  $\Delta t_i = t_i - t_{i-1}$  to the range  $0 \leq \Delta t_i \leq 1$ . The magnitude of  $\Delta t_i$  corresponds to the *hardness* of iteration  $i$ .  $\Delta t_i = 0$  corresponds to the case where the losses of all of the strategies are equal to a common value  $-1 \leq a \leq 1$ . In this case the aggregate loss  $\ell = a$ , the state does not change:  $\Psi(i-1) = \Psi(i)$ ,  $\Delta t_i = 0$  and the instantaneous regret is zero. On the other hand  $\Delta t_i = 1$  corresponds to the adversarial strategy  $Q^{1/2}(t-1, R)$  (Eqn. 15) which maximizes the instantaneous regret.

We introduce an additional step to the integer game. Before the learner makes it's choice, the adversary chooses a real number  $0 \leq s_i \leq 1$ , by doing so, the adversary commits that all of the instantaneous expert losses at that step be in the range  $[-s_i, s_i]$ . The time step is defined to be  $\Delta t_i = s_i^2$ .

First, note that the adversary in this game is at least as powerful as the adversary in the integer game. This is because the adversary can always choose  $s_i = 1$ , effectively reducing the game to the integer game.

Next, we justify the choice  $\Delta t_i = s_i^2$ . Our argument is that any significantly different choice would give the game to one side or the other. Suppose that  $s_i = 1/k$  on all  $m_k$  iterations of the game. In other words, this is a rescaling of the integer game. Consider the adversarial strategy. The distribution (state) after  $m_k$  iterations is the binomial distribution. with mean zero and variance  $m_k \frac{1}{k^2}$  if  $m_k \gg \frac{1}{k^2}$  then the variance at the end of the game goes to infinity.

By allowing  $\Delta t_i$  to vary from iteration to iteration we get a more refined quantification of the regret and, as we show below, min/max optimality.

To find the relationship between loss magnitude and time increments we compare two adversarial strategies. The first strategy, discussed above, generates losses  $\pm 1$  with equal probability, we denote this strategy by  $Q_{\pm 1}^{1/2}$ . The other strategy, denoted  $Q_{\pm 1/k}^{1/2}$ , generates losses of  $\pm 1/k$  with equal probabilities.

From the adversarial point of view  $Q_{\pm 1/k}^{1/2}$  is worse than  $Q_{\pm 1}^{1/2}$ . So it should correspond to a smaller time increment. But how much smaller? Suppose we start with the initial state  $\Psi(0)$  which is a delta functions at  $R = 0$ . One iteration of  $Q_{\pm 1}^{1/2}$  results in a distribution  $\pm 1$  w.p.  $(1/2, 1/2)$ , which has mean 0 and variance 1.

Suppose we associate  $\Delta t = 1/j$  with a single step of  $Q_{\pm 1/k}^{1/2}$ . Equivalently, we associate  $j$  iterations of  $Q_{\pm 1/k}^{1/2}$  with  $t = 1$ . How should we set  $j$ ? the distribution generated by  $j$  steps is a binomial distribution supported on  $j + 1$  points, so there is no hope of making the two distributions identical. However, as it turns out, it is enough to equalize the mean and the variance of the two distributions. The mean of  $Q_{\pm 1/k}^{1/2}$  is zero for any  $k$ . As for the variances, a single iteration of  $Q_{\pm 1}^{1/2}$  is 1 and a single iteration of  $Q_{\pm 1/k}^{1/2}$  is  $1/k^2$ . It follows that the variance after  $j$  iterations of  $Q_{\pm 1/k}^{1/2}$  is  $j/k^2$ . Equating this variance with that of a single step of  $Q_{\pm 1}^{1/2}$  we get  $j = k^2$  and  $\Delta t = 1/k^2$ .

Note a curious behaviour of the *range* of  $R$  as  $k \rightarrow \infty$  the number of steps increases like  $k^2$  while the size of each step is  $1/k$ . This means that the range of  $R$  is  $[-k, k]$ , which becomes converges to  $(-\infty, +\infty)$  when  $k \rightarrow \infty$ . On the other hand, the variance increases like  $t$ .

Next let's consider effect of reducing the step size on a *biased* strategy  $Q_{\pm 1}^{1/2+\gamma}$  as defined in Eqn (13) for some  $0 \leq \gamma \leq 1/2$ . We now figure out what  $\gamma'$  should be so that the distribution generated by  $k^2$  iterations of  $Q_{\pm 1/k}^{1/2+\gamma'}$  has the same mean as a single iteration of  $Q_{\pm 1}^{1/2+\gamma}$ . The mean of a single iteration of  $Q_{\pm 1}^{1/2+\gamma}$  is  $2\gamma$  while the mean of a single iteration of  $Q_{\pm 1/k}^{1/2+\gamma'}$  is  $2\gamma'/k$ . Therefor to keep the means equal we need to set  $2\gamma'/k = 2\gamma$  or  $\gamma' = \gamma/k$ .

Note that as  $k \rightarrow \infty$ ,  $\gamma' \rightarrow 0$ . This observation motivates scaling the bound on  $\ell(t)$  like  $cs_i^2$  (see the description of the game below.)

This leads to the following formulation of a continuous time game. The game is a generalization of the integer time game in that it reduces to the integer time game if the adversary always chooses  $s_i = 1$ .

In this game we use  $i = 1, 2, 3, \dots$  as the iteration index. We use  $t_i$  to indicate a sequence of real-valued time points.  $t_0 = 0$  and we assume there exists a finite  $n$  such that  $t_n = T$ .

We will later give some particular potential functions for which no a-priori knowledge of the termination condition is needed. The associated bounds will hold for any iteration of the game.

## 8.1 The discrete time game

On iteration  $i = 1, 2, \dots$

1. If  $t_{i-1} = T$  the game terminates.
2. The adversary chooses a *step size*  $0 < s_i \leq 1$ , which advances time by  $t_i = t_{i-1} + s_i^2$ .
3. Given  $s_i$ , the learner chooses a distribution  $P(i)$  over  $\mathbb{R}$ .
4. The adversary chooses a mapping from  $\mathbb{R}$  to distributions over  $[-s_i, +s_i]$ :  $Q(t) : \mathbb{R} \rightarrow \Delta^{[-s_i, +s_i]}$
5. The aggregate loss is calculated:

$$\ell(t_i) = \mathbf{E}_{R \sim \Psi(t_i)} [P(t_i, R)B(t_i, R)], \text{ where } B(t_i, R) \doteq \mathbf{E}_{y \sim Q(t_i, R)} [y] \quad (32)$$

6. the aggregate loss is restricted  $|\ell(t_i)| \leq cs_i^2$ .
7. The state is updated. The expectation below is over distributions. and the notation  $G \oplus R$  means that distribution  $G$  over the reals is shifted by the amount defined by the scalar  $R$ :

$$\Psi(t_i) = \mathbf{E}_{R \sim \Psi(t_{i-1})} [Q(t_i)(R) \oplus (R - \ell(t_i))]$$

When  $t_i = T$  the game is terminated, and the final value is calculated:

$$\Phi(T) = \mathbf{E}_{R \sim \Psi(T)} [\phi(T, R)(R)]$$

## 8.2 Results for the discrete time game

In the discrete time game the adversary has an additional choice, the choice of  $s_i$ . Thus the adversary's strategy includes that choice. There are two constraints on this choice:  $s_i \geq 0$  and  $\sum_{i=1}^n s_i^2 = T$ . Note that even that by setting  $s_i$  arbitrarily small, the adversary can make the number of steps -  $n$  - arbitrarily large. We will therefor not identify a single adversarial strategy but instead consider the supremum over an infinite sequence of strategies.

We use  $N(0, \sigma)$  to denote the normal distribution with mean 0 and std  $\sigma$ .

**Theorem 8.**

let  $A = \mathbf{E}_{R \sim N(0, \sqrt{T})} [\phi(T, R)]$

- For any  $\epsilon > 0$  there exists a strategy for the adversary such that for any strategy of the learner  $\Phi(T) \geq A - \epsilon$
- There exists a strategy for the learner that guarantees, against any adversary  $\Phi(T) \leq A$ .

## 8.3 The adversary prefers smaller steps

As noted before, if the adversary chooses  $s_i = 1$  for all  $i$  the game reduces to the integer time game. The question is whether the adversary would prefer to stick with  $s_i = 1$  or instead prefer to use  $s_i < 1$ . In this section we give a surprising answer to this question – the adversary always prefers a smaller value of  $s_i$  to a larger one. This leads to a preference for  $s_i \rightarrow 0$ , as it turns out, this limit is well defined and corresponds to Brownian motion, also known as Wiener process.

Consider a sequence of adversarial strategies  $S_k$  indexed by  $k = 0, 1, 2, \dots$ . The adversarial strategy  $S_k$  corresponds to always choosing  $s_i = 2^{-k}$ , and repeating  $Q_{\pm 2^{-k}}^{1/2}$  for  $T2^{2k}$  iterations. This corresponds to the distribution created by a random walk with  $T2^{2k}$  time steps, each step equal to  $\pm 2^{-k}$  with probabilities  $1/2, 1/2$ . Note that in order to preserve the variance, halving the step size requires increasing the number of iterations by a factor of four.

Let  $\phi(S_k, t, R)$  be the value associated with adversarial strategy  $S_k$ , time  $t$  (divisible by  $2^{-2k}$ ) and location  $R$ . We are ready to state our main theorem.

**Theorem 9.** *If the final value function has a strictly positive fourth derivative:*

$$\frac{d^4}{dR^4} \phi(T, R)(R) > 0, \forall R$$

*then for any integer  $k > 0$  and any  $0 \leq t \leq T$ , such that  $t$  is divisible by  $2^{-2k}$  and any  $R$ ,*

$$\phi(S_{k+1}, t, R) > \phi(S_k, t, R)$$

Before proving the theorem, we describe its consequence for the online learning problem. We can restrict Theorem 9 for the case  $t = 0, R = 0$  in which case we get an increasing sequence:

$$\phi(S_1, 0, 0) < \phi(S_2, 0, 0) < \dots < \phi(S_k, 0, 0) <$$

The limit of the strategies  $S_k$  as  $k \rightarrow \infty$  is the well studied Brownian or Wiener process. The backwards recursion that defines the value function is the celebrated Backwards Kolmogorov Equation with zero drift and unit variance

$$\frac{\partial}{\partial t} \phi(t, R) + \frac{1}{2} \frac{\partial^2}{\partial R^2} \phi(t, R) = 0 \quad (33)$$

Given a final value function with a strictly positive fourth derivative we can use Equation (33) to compute the value function for all  $0 \leq t \leq T$ . We will do so in the next section.

We now go back to proving Theorem 9. The core of the proof is a lemma which compares, essentially, the value recursion when taking one step of size 1 to four steps of size  $1/2$ .

Consider the adversarial strategies  $S_k$  and  $S_{k+1}$  at a particular time point  $0 \leq t \leq T$  such that  $t$  is divisible by  $\Delta t = 2^{-2k}$  and at a particular location  $R$ . Let  $t' = t + \Delta t$ , and fix a value function for time  $t'$ ,  $\phi(t', R)$  and compare between two values at  $R, t$ . The first value denoted  $\phi_k(t, R)$  corresponds to  $S_k$ , and consists of a single random step of  $\pm 2^{-k}$ . The other value  $\phi_{k+1}(t, R)$  corresponds to  $S_{k+1}$  and consists of four random steps of size  $\pm 1/2$ .

**Lemma 10.** *If  $\phi(t', R)$  is, as a function of  $R$  continuous, strictly convex and with a strictly positive fourth derivative. Then*

- $\phi_k(t, R) < \phi_{k+1}(t, R)$
- Both  $\phi_k(t, R)$  and  $\phi_{k+1}(t, R)$  are continuous, strictly convex and with a strictly positive fourth derivative.

*Proof.* Recall the notations  $\Delta t = 2^{-2k}$   $t' = t + \Delta t$  and  $s = 2^{-k}$ . We can write out explicit expressions for the two values:

- For strategy  $S_0$  the value is

$$\phi_k(t, R) = \frac{\phi(t', R + s) + \phi(t', R - s)}{2}$$

- For strategy  $S_1$  the value is

$$\phi_{k+1}(t, R) = \frac{1}{16}(\phi(t', R + 2s) + 4\phi(t', R + s) + 6\phi(t', R) + 4\phi(t', R - s) + \phi(t', R - 2s))$$

We want to show that  $\phi_1(T - 1, R) > \phi_0(T - 1, R)$  for all  $R$ , in other words we want to characterize the properties of  $\phi(T, R)$  that would guarantee that

$$\phi_1(t, R) - \phi_0(t, R) = \frac{1}{16}(\phi(t', R + 2) - 4\phi(t', R + 1) + 6\phi(t', R) - 4\phi(t', R - 1) + \phi(t', R - 2)) > 0 \quad (34)$$

Inequalities of this form have been studied extensively under the name “divided differences” [16, 3, 9]. A function  $\phi(T, R)$  that satisfies inequality 34 is said to be *4'th order convex* (see details in [3]).

$n$ -convex functions have a very simple characterization:

**Theorem 11.** *Let  $f$  be a function which is differentiable up to order  $n$ , and let  $f^{(n)}$  denote the  $n$ 'th derivative, then  $f$  is  $n$ -convex ( $n$ -strictly convex) if and only if  $f^{(n)} \geq 0$  ( $f^{(n)} > 0$ ).*

We conclude that if  $\phi(t', R)$  has a strictly positive fourth derivative then  $\phi_{k+1}(t, R) > \phi_k(t, R)$  for all  $R$ , proving the first part of the lemma.

The second part of the lemma follows from the fact that both  $\phi_{k+1}(t, R)$  and  $\phi_k(t, R)$  are convex combinations of  $\phi(t, R)$  and therefore retain their continuity and convexity properties. □

*Proof.* of Theorem 9

The proof is by double induction over  $k$  and over  $t$ . For a fixed  $k$  we take a finite backward induction over  $t = T - 2^{-2k}, T - 2 \times 2^{-2k}, T - 3 \times 2^{-2k}, \dots, 0$ . Our inductive claims are that  $\phi_{k+1}(t, R) > \phi_k(t, R)$  and  $\phi_{k+1}(t, R), \phi_k(t, R)$  are continuous, strongly convex and have a strongly positive fourth derivative. That these claims carry over from  $t = T - i \times 2^{-2k}$  to  $t = T - (i + 1) \times 2^{-2k}$  follows directly from Lemma 10.

The theorem follows by forward induction on  $k$ . □

## 8.4 Strategies for the Learner in the discrete time game

The strategies we propose for the learner in the discrete time game are an adaptation of the strategies  $P^1, P^2$  from the integer time game to the case where  $s_i < 1$ .

We start with the high-level idea. Consider iteration  $i$  of the continuous time game. We know that the adversary prefers  $s_i$  to be as small as possible. On the other hand, the adversary has to choose some  $s_i > 0$ . This means that the adversary always plays sub-optimally. Based on  $s_i$  the learner makes a choice and the adversary makes a choice. As a result the current state  $\Psi(t_{i-1})$  is transformed to  $\Psi(t_i)$ . To choose it's strategy, the learner needs to assign value possible states  $\Psi(t_i)$ . How can she do that? By assuming that in the future the adversary will play optimally, i.e. setting  $s_i$  arbitrarily small. While the adversary cannot be optimal, it can get arbitrarily close to optimal, which is brownian motion.

Solving the backwards Kolmogorov equation with the boundary condition  $\phi(T, R)$  yields  $\phi(t, R)$  for any  $R \in \mathbb{R}$  and  $t \in [0, T]$ . We now explain how using this potential function we derive strategies for the learner.

Note that the learner chooses a distribution *after* the adversary set the value of  $s_i$ . The discrete time version of  $P^1$  (Eqn 17) is

$$P^{1d}(t_{i-1}, R) = \frac{1}{Z^{1d}} \frac{\phi(t_i, R + s_{i-1} + cs_{i-1}^2) - \phi(t_i, R - s_{i-1} - cs_{i-1}^2)}{2} \quad (35)$$

where  $Z^{1d} = \mathbf{E}_{R \sim \Psi(t_i)} \left[ \frac{\phi(t_i, R + s_{i-1} + cs_{i-1}^2) - \phi(t_i, R - s_{i-1} - cs_{i-1}^2)}{2} \right]$

Next, we consider the discrete time version of  $P^2$ : (Eqn 30)

$$P^{2d}(t_{i-1}, R) = \frac{1}{Z^{2d}} \frac{\partial}{\partial r} \bigg|_{r=R} \phi(t_{i-1} + s_{i-1}^2, r) \quad (36)$$

where  $Z^{2d} = \mathbf{E}_{R \sim \Psi(t_i)} \left[ \frac{\partial}{\partial r} \bigg|_{r=R} \phi(t_{i-1} + s_{i-1}^2, r) \right]$

## 9 Bounds for easy sequences

In Section 8 we have shown that the integer time game has a natural extension to a setting where  $\Delta t_i = s_i^2$ . We also showed that the adversary gains from setting  $s_i$  as small as possible.

We characterized the optimal adversarial strategy for the discrete time game (Section 8.1), which corresponds to the adversary choosing the loss to be  $s_i$  or  $-s_i$  with equal probabilities. A natural question at this point is to characterize the regret when the adversary is not optimal, or the sequences are “easy”.

To see that such an improvement is possible, consider the following *constant* adversary. This adversary associates the same loss to all experts on iteration  $i$ , formally,  $Q(i, R) = l$ . In this case the average loss is also equal to  $l$ ,  $\ell(i) = l$  which means that all of the instantaneous regrets are  $r = l - \ell(t_i) = 0$ , which, in turn, implies that  $\Psi(i) = \Psi(i+1)$ . As the state did not change, it makes sense to set  $t_{i+1} = t_i$ , rather than  $t_{i+1} = t_i + s_i^2$ .

We observe two extremes for the adversarial behaviour. The constant adversary described above for which  $t_{i+1} = t_i$ , and the random walk adversary described earlier, in which each expert is split into two, one half with loss  $-s_i$  and the other with loss  $+s_i$ . In which case  $t_{i+1} = t_i + s_i^2$  which is the maximal increase in  $t$  that the adversary can guarantee. The analysis below shows that these are two extremes on a spectrum and that intermediate cases can be characterized using a variance-like quantity.

We define a variant of the discrete time game (8.1) For concreteness we include the learner's strategy, which is the limit of the strategy in the discrete game when  $s_i \rightarrow 0$ .

### 9.0.1 The continuous time game and a learner strategy

Set  $t_1 = 0$

Fix maximal step  $0 < s < 1$

On iteration  $i = 1, 2, \dots$

1. If  $t_i = T$  the game terminates.
2. Given  $t_{i-1}$ , the learner chooses a distribution  $P(i)$  over  $\mathbb{R}$ :

$$P^{cc}(t, R) = \frac{1}{Z^{cc}} \frac{\partial}{\partial r} \Big|_{r=R} \phi(t, r) \text{ where } Z^{cc} = \mathbf{E}_{R \sim \Psi(t_i)} \left[ \frac{\partial}{\partial r} \Big|_{r=R} \phi(t, r) \right] \quad (37)$$

3. The adversary chooses a *step size*  $0 < s_i \leq s$  and a mapping from  $\mathbb{R}$  to distributions over  $[-s_i, +s_i]$ :  
 $Q(t) : \mathbb{R} \rightarrow \Delta^{[-s_i, +s_i]}$
4. The aggregate loss is calculated:

$$\ell(t_i) = \mathbf{E}_{R \sim \Psi(t_i)} [P^{cc}(t_i, R) B(t_i, R)], \text{ where } B(t_i, R) \doteq \mathbf{E}_{y \sim Q(t_i, R)} [y] \quad (38)$$

the aggregate loss is restricted to  $|\ell(t_i)| \leq cs_i^2$ .

5. Increment  $t_{i+1} = t_i + \Delta t_i$  where

$$\Delta t_i = \mathbf{E}_{R \sim \Psi(t_i)} \left[ \mathbf{E}_{y \sim Q(t_i, R)} \left[ \left\{ \frac{\partial^2}{\partial r^2} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R}} \right\} ((y - \ell(t_i))^2) \right] \right] \quad (39)$$

6. The state is updated. The expectation below is over distributions. and the notation  $G \oplus R$  means that distribution  $G$  over the reals is shifted by the amount defined by the scalar  $R$ :

$$\Psi(t_{i+1}) = \mathbf{E}_{R \sim \Psi(t_i)} [Q(t_i)(R) \oplus (R - \ell(t_i))]$$

Our characterization applies to the limit where the  $s_i$  are small. Formally, we define

**Definition 4.** We say that an instance of the discrete time game is  $(n, s, \tau)$ -bounded if it consists of  $n$  iterations and  $\forall 0 < i \leq n, s_i < s$  and  $\sum_{j=1}^n s_j^2 = \tau$

Note that  $\tau > t_n$  and that  $\tau$  depends only on the ranges  $s_i$  while  $t_n$  depends on the variance.  $t_n = T$  is the dominant term in the regret bound, while  $\tau$  controls the error term.

**Theorem 12.** Let  $\phi \in \mathcal{P}^\infty$  be a potential function that satisfies the Kolmogorov backward equation (33). Assume a sequence of  $(n_i, s_i, \tau)$ -bounded games where  $\tau$  is a constant  $n \rightarrow \infty$  and  $s = \sqrt{\frac{\tau}{n}}$

Then

$$\Phi(\Psi(t_{i+1})) \leq \Phi(\Psi(t_i)) + O(s\tau)$$

The proof is given in appendix B

If we define

$$V_n = t_n = \sum_{i=1}^n \Delta t_i = \sum_{i=1}^n \mathbf{E}_{R \sim \Psi(t_i)} [\mathbf{E}_{y \sim Q(t_i, R)} [Q(t_i, R)(y - \ell(t_i))^2]] \quad (40)$$

We can use  $V_n$  instead of  $T$  giving us a variance based bound.

## 10 Two self-consistent potential functions

The potential functions,  $\phi(t, R)$  is a solution of PDE (33):

$$\frac{\partial}{\partial t}\phi(t, R) + \frac{1}{2} \frac{\partial^2}{\partial R^2}\phi(t, R) = 0 \quad (41)$$

under a boundary condition  $\phi(T, R) = \phi(T, R)(R)$ , which we assume is in  $\mathcal{P}^4$

So far, we assumed that the game horizon  $T$  is known in advance. We now show two value functions where knowledge of the horizon is not required. Specifically, we call a value function  $\phi(t, R)$  *self consistent* if it is defined for all  $t > 0$  and if for any  $0 < t < T$ , setting  $\phi(T, R)$  as the final potential and solving for the Kolmogorov Backward Equation yields  $\phi(t, R)$  regardless of the time horizon  $T$ .

We consider two solutions to the PDE, the exponential potential and the NormalHedge potential. We give the form of the potential function that satisfies Kolmogorov Equation 33, and derive the regret bound corresponding to it.

**The exponential potential function** which corresponds to exponential weights algorithm corresponds to the following equation

$$\phi_{\text{exp}}(R, t) = e^{\sqrt{2}\eta R - \eta^2 t}$$

Where  $\eta > 0$  is the learning rate parameter.

Given  $\epsilon$  we choose  $\eta = \sqrt{\frac{\ln(1/\epsilon)}{t}}$  we get the regret bound that holds for any  $t > 0$

$$R_\epsilon \leq \sqrt{2t \ln \frac{1}{\epsilon}} \quad (42)$$

Note that the algorithm depends on the choice of  $\epsilon$ , in other words, the bound does *not* hold for all values of  $\epsilon$  at the same time.

**The NormalHedge value** is

$$\phi_{\text{NH}}(R, t) = \begin{cases} \frac{1}{\sqrt{t+\nu}} \exp\left(\frac{R^2}{2(t+\nu)}\right) & \text{if } R \geq 0 \\ \frac{1}{\sqrt{t+\nu}} & \text{if } R < 0 \end{cases} \quad (43)$$

Where  $\nu > 0$  is a small constant. The function  $\phi_{\text{NH}}(R, t)$ , restricted to  $R \geq 0$  is in  $\mathcal{P}^4$  and is a constant for  $R \leq 0$ .

The regret bound we get is:

$$R_\epsilon \leq \sqrt{(t+\nu) \left( \ln(t+\nu) + 2 \ln \frac{1}{\epsilon} \right)} \quad (44)$$

This bound is slightly larger than the bound for exponential weights, however, the NormalHedge bound holds simultaneously for all  $\epsilon > 0$  and the algorithm requires no tuning.

## 11 NormalHedge yields the fastest increasing potential

Up to this point, we considered any continuous value function with strictly positive derivatives 1-4. We characterized the min-max strategies for any such function. It is time to ask whether value functions can be compared and whether there is a “best” value function. In this section we give an informal argument that NormalHedge is the best function. We hope this argument can be formalized.

We make two observations. First, the min-max strategy for the adversary does not depend on the potential function! (as long as it has strictly positive derivatives). That strategy corresponds to the brownian process.

Second, the argument used to show that the regret relative to  $\epsilon$ -fraction of the expert is based on two arguments



- The average value function does not increase with time.
- The (final) value function increases rapidly as a function of  $R$

The first item is true by construction. The second argument suggests the following partial order on value functions. Let  $\phi_1(t, R), \phi_2(t, R)$  be two value functions such that

$$\lim_{R \rightarrow \infty} \frac{\phi_1(t, R)}{\phi_2(t, R)} = \infty$$

then  $\phi_1$  *dominates*  $\phi_2$ , which we denote by,  $\phi_1 > \phi_2$ .

On the other hand, if the value function increases too quickly, then, when playing against brownian motion, the average value will increase without bound. Recall that the distribution of the brownian process at time  $t$  is the standard normal with mean 0 and variance  $t$ . The question becomes what is the fastest the value function can grow, as a function of  $R$  and still have a finite expected value with respect to the normal distribution.

The answer seems to be NormalHedge (Eqn. 43). More precisely, if  $\epsilon > 0$ , the mean value is finite, but if  $\epsilon = 0$  the mean value becomes infinite.

## References

- [1] Jacob Abernethy, John Langford, and Manfred K Warmuth. Continuous experts and the binning algorithm. In *International Conference on Computational Learning Theory*, pages 544–558. Springer, 2006.
- [2] Jacob Abernethy, Manfred K Warmuth, and Joel Yellin. Optimal strategies from random walks. In *Proceedings of The 21st Annual Conference on Learning Theory*, pages 437–446. Citeseer, 2008.
- [3] Saad Ihsan Butt, Josip Pečarić, and Ana Vukelić. Generalization of popoviciu-type inequalities via fink’s identity. *Mediterranean journal of mathematics*, 13(4):1495–1511, 2016.
- [4] Nicolo 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 (JACM)*, 44(3):427–485, 1997.
- [5] Nicolo Cesa-Bianchi, Yoav Freund, David P Helmbold, and Manfred K Warmuth. On-line prediction and conversion strategies. *Machine Learning*, 25(1):71–110, 1996.
- [6] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.
- [7] Nicolo Cesa-Bianchi, Yishay Mansour, and Gilles Stoltz. Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66(2):321–352, 2007.
- [8] Kamalika Chaudhuri, Yoav Freund, and Daniel J Hsu. A parameter-free hedging algorithm. *Advances in neural information processing systems*, 22, 2009.
- [9] Carl de Boer. Divided differences. *arXiv preprint math/0502036*, 2005.
- [10] Meir Feder, Neri Merhav, and Michael Gutman. Universal prediction of individual sequences. *IEEE transactions on Information Theory*, 38(4):1258–1270, 1992.
- [11] Yoav Freund and Manfred Opper. Drifting games and brownian motion. *Journal of Computer and System Sciences*, 64(1):113–132, 2002.
- [12] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

- [13] Yoav Freund and Robert E Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1-2):79–103, 1999.
- [14] Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. *Information and computation*, 108(2):212–261, 1994.
- [15] Haipeng Luo and Robert E Schapire. Achieving all with no parameters: Adanormalhedge. In *Conference on Learning Theory*, pages 1286–1304. PMLR, 2015.
- [16] Tiberiu Popoviciu. Sur certaines inégalités qui caractérisent les fonctions convexes. *Analele Stiintifice Univ. “Al. I. Cuza”, Iasi, Sectia Mat*, 11:155–164, 1965.
- [17] Robert E Schapire. Drifting games. *Machine Learning*, 43(3):265–291, 2001.
- [18] Volodimir G Vovk. Aggregating strategies. *Proc. of Computational Learning Theory, 1990*, 1990.

## A Proof of Theorem 1

*Proof.* •  $\Psi$  satisfies a simultaneous bound for  $B$  if it satisfies an average potential bound for  $\phi = B^{-1}$

Assume by contradiction that  $\Psi$  does not satisfy the simultaneous bound. In other words there exists  $a \in \mathbb{R}$  such that  $\mathbf{P}_{R \sim \Psi}[R > a] > B(a)$ . From Markov inequality and the fact that  $\phi$  is non decreasing we get

$$\mathbf{E}_{R \sim \Psi}[\phi(R)] \geq \phi(a)\mathbf{P}_{R \sim \Psi}[R > a] > \phi(a)B(a) = \frac{B(a)}{B(a)} = 1$$

but  $\mathbf{E}_{R \sim \Psi}[\phi(R)] > 1$  contradicts the average potential assumption for the potential  $\phi(R) = B(R)^{-1}$

- $\Psi$  satisfies an average potential bound for  $\phi = B^{-1}$  if it satisfies a simultaneous bound for  $B$

As  $\phi$  is a non-decreasing function, and assuming  $R, R'$  are drawn independently at random according to  $\Psi$ :

$$\mathbf{E}_{R \sim \Psi}[\phi(R)] = \mathbf{E}_{R \sim \Psi}[\phi(R)\mathbf{P}_{R' \sim \Psi}[\phi(R') \geq \phi(R)]] \quad (45)$$

$$\leq \mathbf{E}_{R \sim \Psi}[\phi(R)\mathbf{P}_{R' \sim \Psi}[R' \geq R]] \quad (46)$$

$$< \mathbf{E}_{R \sim \Psi}[\phi(R)B(R)] \quad (47)$$

$$= \mathbf{E}_{R \sim \Psi}\left[\frac{B(R)}{B(R)}\right] = \mathbf{E}_{R \sim \Psi}[1] = 1 \quad (48)$$

□

## B Proof of Theorem 12

We start with two technical lemmas

**Lemma 13.** Let  $f(x) \in \mathcal{P}^2$ , i.e.  $f(x), f'(x), f''(x) > 0$  for all  $x \in \mathbb{R}$ , let  $h(x)$  be a uniformly bounded function:  $\forall x, |h(x)| < 1$ . Let  $\Psi$  be a distribution over  $\mathbb{R}$ . If  $\mathbf{E}_{x \sim \Psi}[f(x)]$  is well-defined (and finite), then  $\mathbf{E}_{x \sim \Psi}[h(x)f'(x)]$  is well defined (and finite) as well.

*Proof.* Assume by contradiction that  $\mathbf{E}_{x \sim \Psi}[h(x)f'(x)]$  is undefined. Define  $h^+(x) = \max(0, h(x))$ . As  $f'(x) > 0$ , this implies that either  $\mathbf{E}_{x \sim \Psi}[h^+(x)f'(x)] = \infty$  or  $\mathbf{E}_{x \sim \Psi}[(-h)^+(x)f'(x)] = \infty$  (or both).

Assue wlog that  $\mathbf{E}_{x \sim \Psi}[h^+(x)f'(x)] = \infty$ . As  $f'(x) > 0$  and  $0 \leq h^+(x) \leq 1$  we get that  $\mathbf{E}_{x \sim \Psi}[f'(x)] = \infty$ . As  $f(x+1) \geq f'(x)$  we get that  $\mathbf{E}_{x \sim \Psi}[f(x)] = \infty$  which is a contradiction. □

**Lemma 14.** Let  $f(x, y)$  be a differentiable function with continuous derivatives up to degree three. Then

$$f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0) + \left\{ \frac{\partial}{\partial x} f(x, y) \Big|_{x_0, y_0} \right\} \Delta x + \left\{ \frac{\partial}{\partial y} f(x, y) \Big|_{x_0, y_0} \right\} \Delta y \quad (49)$$

$$+ \frac{1}{2} \left\{ \frac{\partial^2}{\partial x^2} f(x, y) \Big|_{x_0, y_0} \right\} \Delta x^2 + \left\{ \frac{\partial^2}{\partial x \partial y} f(x, y) \Big|_{x_0, y_0} \right\} \Delta x \Delta y + \frac{1}{2} \left\{ \frac{\partial^2}{\partial y^2} f(x, y) \Big|_{x_0, y_0} \right\} \Delta y^2 \quad (50)$$

$$+ \frac{1}{6} \left\{ \frac{\partial^3}{\partial x^3} f(x, y) \Big|_{x_0 + t\Delta x, y_0 + t\Delta y} \right\} \Delta x^3 + \frac{1}{2} \left\{ \frac{\partial^3}{\partial x^2 \partial y} f(x, y) \Big|_{x_0 + t\Delta x, y_0 + t\Delta y} \right\} \Delta x^2 \Delta y \quad (51)$$

$$+ \frac{1}{2} \left\{ \frac{\partial^3}{\partial x \partial y^2} f(x, y) \Big|_{x_0 + t\Delta x, y_0 + t\Delta y} \right\} \Delta x \Delta y^2 + \frac{1}{6} \left\{ \frac{\partial^3}{\partial y^3} f(x, y) \Big|_{x_0 + t\Delta x, y_0 + t\Delta y} \right\} \Delta y^3 \quad (52)$$

for some  $0 \leq t \leq 1$ .

*Proof. of Lemma 14* Let  $F : [0, 1] \rightarrow \mathbb{R}$  be defined as  $F(t) = f(x(t), y(t))$  where  $x(t) = x_0 + t\Delta x$  and  $y(t) = y_0 + t\Delta y$ . Then  $F(0) = f(x_0, y_0)$  and  $F(1) = f(x_0 + \Delta x, y_0 + \Delta y)$ . It is easy to verify that

$$\frac{d}{dt} F(t) = \frac{\partial}{\partial x} f(x(t), y(t)) \Delta x + \frac{\partial}{\partial y} f(x(t), y(t)) \Delta y$$

and that in general:

$$\frac{d^n}{dt^n} F(t) = \sum_{m=1}^n \binom{n}{m} \frac{\partial^n}{\partial x^m \partial y^{n-m}} f(x_0 + t\Delta x, y_0 + t\Delta y) \Delta x^m \Delta y^{n-m} \quad (53)$$

As  $f$  has partial derivatives up to degree 3, so does  $F$ . Using the Taylor expansion of  $F$  and the intermediate point theorem we get that

$$f(x_0 + \Delta x, y_0 + \Delta y) = F(1) = F(0) + \frac{d}{dt} F(0) + \frac{1}{2} \frac{d^2}{dt^2} F(0) + \frac{1}{6} \frac{d^3}{dt^3} F(t') \quad (54)$$

Where  $0 \leq t' \leq 1$ . Using Eqn (53) to expand each term in Eqn. (54) completes the proof.  $\square$

*Proof. of Theorem 12*

We prove the claim by an upper bound on the increase of potential that holds for any iteration  $1 \leq i \leq n$ :

$$\Phi(\Psi(t_{i+1})) \leq \Phi(\Psi(t_i)) + as_i^3 \text{ for some constant } a > 0 \quad (55)$$

Summing inequality (55) over all iterations we get that

$$\Phi(\Psi(T)) \leq \Phi(\Psi(0)) + c \sum_{i=1}^n s_i^3 \leq \Phi(\Psi(0)) + as \sum_{i=1}^n s_i^2 = \Phi(\Psi(0)) + asT \quad (56)$$

From which the statement of the theorem follows.

We now prove inequality (55). We use the notation  $r = y - \ell(i)$  to denote the instantaneous regret at iteration  $i$ .

Applying Lemma 14 to  $\phi(t_{i+1}, R_{i+1}) = \phi(t_i + \Delta t_i, R_i + r_i)$  we get

$$\phi(t_i + \Delta t_i, R_i + r_i) = \phi(t_i, R_i) \quad (57)$$

$$+ \left\{ \frac{\partial}{\partial \rho} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R}} \right\} r_i \quad (58)$$

$$+ \left\{ \frac{\partial}{\partial \tau} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R}} \right\} \Delta t_i \quad (59)$$

$$+ \frac{1}{2} \left\{ \frac{\partial^2}{\partial \rho^2} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R}} \right\} r_i^2 \quad (60)$$

$$+ \left\{ \frac{\partial^2}{\partial r \partial \tau} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R}} \right\} r_i \Delta t_i \quad (61)$$

$$+ \frac{1}{2} \left\{ \frac{\partial^2}{\partial \tau^2} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R}} \right\} \Delta t_i^2 \quad (62)$$

$$+ \frac{1}{6} \left\{ \frac{\partial^3}{\partial \rho^3} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i + g \Delta t_i, R_i + g r_i}} \right\} r_i^3 \quad (63)$$

$$+ \frac{1}{2} \left\{ \frac{\partial^3}{\partial \rho^2 \partial \tau} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i + g \Delta t_i, R_i + g r_i}} \right\} r_i^2 \Delta t_i \quad (64)$$

$$+ \frac{1}{2} \left\{ \frac{\partial^3}{\partial \rho \partial \tau^2} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i + g \Delta t_i, R_i + g r_i}} \right\} r_i \Delta t_i^2 \quad (65)$$

$$+ \frac{1}{6} \left\{ \frac{\partial^3}{\partial \tau^3} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i + g \Delta t_i, R_i + g r_i}} \right\} \Delta t_i^3 \quad (66)$$

for some  $0 \leq g \leq 1$ .

By assumption  $\phi$  satisfies the Kolmogorov backward equation:

$$\frac{\partial}{\partial \tau} \phi(\tau, \rho) = -\frac{1}{2} \frac{\partial^2}{\partial \rho^2} \phi(\tau, \rho)$$

Combining this equation with the exchangability of the order of partial derivative (Clairaut's Theorem) we can substitute all partial derivatives with respect to  $\tau$  with partial derivatives with respect to  $\rho$  using the following equation.

$$\frac{\partial^{n+m}}{\partial \rho^n \partial \tau^m} \phi(\tau, \rho) = (-1)^m \frac{\partial^{n+2m}}{\partial \rho^{n+2m}} \phi(\tau, \rho)$$

Which yields

$$\phi(t_i + \Delta t_i, R_i + r_i) = \phi(t_i, R_i) \quad (67)$$

$$+ \left\{ \frac{\partial}{\partial \rho} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R}} \right\} r_i \quad (68)$$

$$+ \left\{ \frac{\partial^2}{\partial \rho^2} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R}} \right\} \left( \frac{r_i^2}{2} - \Delta t_i \right) \quad (69)$$

$$- \left\{ \frac{\partial^3}{\partial \rho^3} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R}} \right\} r_i \Delta t_i \quad (70)$$

$$+ \frac{1}{2} \left\{ \frac{\partial^4}{\partial \rho^4} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R}} \right\} \Delta t_i^2 \quad (71)$$

$$+ \frac{1}{6} \left\{ \frac{\partial^3}{\partial \rho^3} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i + g\Delta t_i, R_i + gr_i}} \right\} r_i^3 \quad (72)$$

$$- \frac{1}{2} \left\{ \frac{\partial^4}{\partial \rho^4} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i + g\Delta t_i, R_i + gr_i}} \right\} r_i^2 \Delta t_i \quad (73)$$

$$+ \frac{1}{2} \left\{ \frac{\partial^5}{\partial \rho^5} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i + g\Delta t_i, R_i + gr_i}} \right\} r_i \Delta t_i^2 \quad (74)$$

$$- \frac{1}{6} \left\{ \frac{\partial^6}{\partial \rho^6} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i + g\Delta t_i, R_i + gr_i}} \right\} \Delta t_i^3 \quad (75)$$

From the assumption that the game is  $(n, s, T)$ -bounded we get that

1.  $|r_i| \leq s_i + cs_i^2 \leq 2s_i$
2.  $\Delta t_i \leq s_i^2 \leq s^2$

given these inequalities we can rewrite the second factor in each term as follows, where  $|h_i(\cdot)| \leq 1$

- **For (68):**  $r_i = 2s_i \frac{r_i}{2s_i} = 2s_i h_1(r_i)$ .
- **For (69):**  $r_i^2 - \frac{1}{2} \Delta t_i = 4s_i^2 \frac{r_i^2 - \frac{1}{2} \Delta t_i}{4s_i^2} = 4s_i^2 h_2(r_i, \Delta t_i)$
- **For (70):**  $r_i \Delta t_i = 2s_i^3 \frac{r_i \Delta t_i}{2s_i^3} = 2s_i^3 h_3(r_i, \Delta t_i)$
- **For (71):**  $\Delta t_i^2 = s_i^4 \frac{\Delta t_i^2}{s_i^4} = s_i^4 h_4(\Delta t_i)$
- **For (72):**  $r_i^3 = 8s_i^3 \frac{r_i^3}{8s_i^3} = 8s_i^3 h_5(r_i, \Delta t_i)$
- **For (73):**  $r_i^2 \Delta t_i = 4s_i^4 \frac{r_i^2 \Delta t_i}{4s_i^4} = 4s_i^4 h_6(r_i, \Delta t_i)$
- **For (74):**  $r_i \Delta t_i^2 = 2s_i^5 \frac{r_i \Delta t_i^2}{2s_i^5}$

- **For (75):**  $\Delta t_i^3 = s_i^6 \frac{\Delta t_i^3}{s_i^6}$

We therefor get the simplified equation

$$\begin{aligned}
\phi(t_i + \Delta t_i, R_i + r_i) &= \phi(t, R) + \left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} r + \left\{ \frac{\partial}{\partial t} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} \Delta t \\
&+ \frac{1}{2} \left\{ \frac{\partial^2}{\partial r^2} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} r^2 \\
&+ \left\{ \frac{\partial^2}{\partial r \partial t} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} r_i \Delta t_i \\
&+ \frac{1}{6} \left\{ \frac{\partial^3}{\partial r^3} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} r_i^3 + O(s^4)
\end{aligned}$$

and therefor

$$\begin{aligned}
\phi(t_i + \Delta t_i, R + r) &= \phi(t_i, R) + \left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} r \\
&+ \left\{ \frac{\partial^2}{\partial r^2} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} (r^2 - \Delta t_i) + O(s^3)
\end{aligned} \tag{76}$$

Our next step is to consider the expected value of (76) wrt  $R \sim \Psi(t_i)$ ,  $y \sim Q(t_i, R)$  for an arbitrary adversarial strategy  $Q$ .

We will show that the expected potential does not increase:

$$\mathbf{E}_{R \sim \Psi(t_i)} [\mathbf{E}_{y \sim Q(t_i, R)} [\phi(t_i + \Delta t_i, R + y - \ell(t_i))]] \leq \mathbf{E}_{R \sim \Psi(t_i)} [\phi(t_i, R)] \tag{77}$$

Plugging Eqn (76) into the LHS of Eqn (77) we get

$$\mathbf{E}_{R \sim \Psi(t_i)} [\mathbf{E}_{y \sim Q(t_i, R)} [\phi(t_i + \Delta t_i, R + y - \ell(t_i))]] \tag{78}$$

$$= \mathbf{E}_{R \sim \Psi(t_i)} [\phi(t_i, R)] \tag{79}$$

$$+ \mathbf{E}_{R \sim \Psi(t_i)} \left[ \mathbf{E}_{y \sim Q(t_i, R)} \left[ \left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} (y - \ell(t_i)) \right] \right] \tag{80}$$

$$+ \mathbf{E}_{R \sim \Psi(t_i)} \left[ \mathbf{E}_{y \sim Q(t_i, R)} \left[ \left\{ \frac{\partial^2}{\partial r^2} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} ((y - \ell(t_i))^2 - \Delta t_i) \right] \right] \tag{81}$$

$$+ O(s^3) \tag{82}$$

Some care is needed here. we need to show that the expected value are all finite. We assume that the expected potential (Eqn (eqn:contin0)) is finite. Using Lemma 13 this implies that the expected value of higher derivatives of  $\frac{\partial}{\partial R} \phi(R)$  are also finite.<sup>2</sup>

To prove inequality (55), we need to show that the terms 80 and 81 are smaller or equal to zero.

---

<sup>2</sup>I need to clean this up and find an argument that the expected value for mixed derivatives is also finite.

**Term (80) is equal to zero:**

As  $\ell(t_i)$  is a constant relative to  $R$  and  $y$ , and  $\left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\}$  is a constant with respect to  $y$  we can rewrite (80) as

$$\mathbf{E}_{R \sim \Psi(t_i)} \left[ \left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} \mathbf{E}_{y \sim Q(t_i, R)} [y] \right] - \ell(t_i) \mathbf{E}_{R \sim \Psi(t_i)} \left[ \left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} \right] \quad (83)$$

Combining the definitions of  $\ell(t)$  (38) and the learner's strategy  $P^{cc}$  (37) we get that

$$\ell(t_i) = \mathbf{E}_{R \sim \Psi(t_i)} \left[ \frac{1}{Z} \left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} \mathbf{E}_{y \sim Q(i, R)} [y] \right] \text{ where } Z = \mathbf{E}_{R \sim \Psi(t_i)} \left[ \frac{1}{Z} \left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} \right] \quad (84)$$

Plugging (84) into (83) and recalling the requirement that  $\ell(t_i) < \infty$  we find that term (80) is equal to zero.

**Term (81) is equal to zero:**

As  $\Delta t_i$  is a constant relative to  $y$ , we can take it outside the expectation and plug in the definition of  $\Delta t_i$  (39)

$$\mathbf{E}_{R \sim \Psi(t_i)} \left[ \mathbf{E}_{y \sim Q(t_i, R)} \left[ \left\{ \frac{\partial^2}{\partial r^2} \phi(\tau, \rho) \Big|_{\tau, \rho = t_i, R} \right\} (y - \ell(t_i))^2 \right] - \Delta t_i \right] = \Delta t_i - \Delta t_i = 0 \quad (85)$$

We find that (81) is zero.

Finally (82) is negligible relative to the other terms as  $s \rightarrow 0$ . □