Potential-based hedging algorithms

Yoav Freund

November 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 ϵ 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 an algorithm with second order bounds that characterize easy sequences.

1 Introduction

Online prediction with expert advise has been studied extensively over the years and the number of publications in the area is vast (see e.g. [20, 10, 16, 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 this game.

DTOL (Figure 1) 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.

For $i = 1, \ldots, T$

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

Figure 1: Decision theoretic online learning

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

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

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

A potential function is an increasing function $\phi : \mathbb{R} \to \mathbb{R}$. Potential based learing algorithm are designed to bound the the average potential $\mathbf{E}_{R \sim \mathbf{\Psi}} [\phi(R)]$. Potential functions have long been used to design and analyze online learning algorithms. The novelty here is that we consider minimizing the average potential as a goal in itself.

Definition 2 (Average potential bound)**.** *A distribution over he reals $\mathbf{\Psi}$ satisfies the average potential function ϕ if*

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

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

The next theorem identifies a one to one relationship between simultanous bounds and average potential bounds.

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

The proof of the theorem is in Appendix A.

Simultanous regret bounds are more intuitive than potential functions. On the other hand potent, but average potential bounds lend themselves to analysis and optimization. In

given the potential function at the end of the game $\phi(T, R)$ and the strategies used by the learner and the adversary, we can define a potential $\phi(T - 1, R)$ such that the average potentials are equal:

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

In Section 5.1 we that this equation can be used to create a potential function of all iterations, and how it can be used to find the optimal strategies for both learner and adversary.

Our ultimate result are min-max optimal bounds. However, there seem to be no matching min-max strategies for the original DTOL. To achieve min-max optimality we extend the game by enlarging the set of choices available to the adversary. As the learner does not get additional choices, 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 organizes as follows.

TBD

2 related work

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, 17] are member of this set.

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

[1] The analysis described here builds on a long line of work. Including the Binomial Weights algorithm and it's variants [5, 1, 2] as well as drifting games [19, 11].

Zero-order bounds on the regret [14] 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 ϵ-regret of the algorithm as the regret with respect to the best (smallest-loss) ϵ-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

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

A few things are worth noting. First, as $|l_j^i| \leq 1$, $\mathrm{Var}_j \leq 1$ 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.

3 Main Results

1. **Unifrm regret bound** There exists an online learning algorithm such that for any $\nu > 0$ (set in advance) and any t, ϵ (holds uniformly) the following regret bound holds.

$$R_\epsilon \leq \sqrt{(t+\nu)\left(\ln(t+\nu) + 2 \ln \frac{1}{\epsilon}\right)} \tag{4}$$

2. **Second order bound**

3. **optimality of Brownian motion** For any potential function in $\mathcal{P}^4$ the min/max value of any state (t, R) is attained by Brownian motion on the part of the adversary for any $s \geq t$.

4 Preliminaries

We define some notation that will be used in the rest of the paper.

Positivity We require that potential functions have positive derivatives for a range of degree. To that end we use the following definition:

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

For the integer time game we assume that $\phi(\cdot) \in \mathcal{P}^2$. Later on, in section 6.2, we will further restrict our potential functions to be in $\mathcal{P}^4$.

Divisibility: To reach optimality we need the set of actions to be arbitrarily divisible. Intuitively, We replace the finite set of actions with a continuous mass, so that each set of actions can be partitioned into two parts of equal weight. Formally, we define the set of actions to be a probability space (Ω, σ, μ) such that $\omega \in \Omega$ is a particular action. We require that the space is *arbitrarily divisible*, which means that for any $s \in \sigma$, there exist a partition $u, v \in \sigma$ such that $u \cup v = s, u \cap v = \emptyset$, and $\mathbf{P}[u] = \mathbf{P}[v] = \frac{1}{2}\mathbf{P}[s]$.

State: The *state* of a game at iteration i, denoted $\mathbf{\Psi}(i)$, is a random variable that maps each action $\omega \in \Omega$ to the cumulative regret of ω at time i: R_ω^i. The sequence of cumulative regrets corresponding to action ω is the *path* of ω:

$$S_\omega = (R_\omega^1, R_\omega^2, \ldots, R_\omega^N) \tag{5}$$

Generalized binomial distribution We denote by $\mathbb{B}(n, s)$ the distribution over the reals defined by $\sum_{i=1}^n X_i$ where X_i are iid binary random variables which attain the values $-s, +s$ with equal probabilities.

Expected value shorthand: Suppose P is a distribution over the reals, and $f : \mathbb{R} \to \mathbb{R}$, we use the following short-hand notation for the expected value of f under the distribution P:

$$P \odot f \doteq \mathbf{E}_{x \sim P}[f(x)]$$

We define the *score* at iteration i as the average potential with respect to the state:

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

Note that in this short-hand notation we suppress the variable with respect to which the integration is defined, which will always be R.

Convolution: Let A, B be two independent random variables. We define the convolution $A \oplus B$ to be the distribution of $x + y$. A constant a corresponds to the point mass distribution concentrated at a. For convenience we define $A \ominus B = A \oplus (-B)$

5 Integer time game

The integer time game is described in Figure 2. The integer time game generalizes the decision theoretic online learning problem [13] in the following ways:

1. The goal of the learner in DTOL is to guarantee an upper bounds on the regret. The learner's goal in the integer time game is to minimize the final score. From theorem 1 we know that if we set the potential as $\phi(R) = \frac{1}{G(x)}$ then the two conditions are equivalent, allowing us to focus on the score.

2. The number of iterations T is given as input, as is the potential function at the end: $\phi(T, R)$.

3. The actions are assumed to be *divisible.* For our purposes it is enough to assume that any action can be split into two equal weight parts.

The key to the potential based analysis is that usiing the predefined final potential we can define potential functions and scores for all iterations $1, \ldots, T-1$. This is explained in the next subsection.

5.1 Defining potential Functions for all iterations

The potential game defines the *final* potential function $\phi(T)$, at the end of the game. We will now show, that we can extend the definition of a potential function to all iterations of the game.

A single action defines a path S_ω (as defined in (5)). Fixing the strategies of the learner and the adversary determines a distribution $\mathcal{D}$ over paths. We describe two equivalent ways to define $\phi_{P,Q}(i, R)$ for $i < T$

Initialization:

- input: T : The number of iterations.
- Final iteration potential function: $\phi(T,R) \in \mathcal{P}^2$
- $\mathbf{\Psi}(1) = \delta(0)$ is the initial state of the game which is a point mass distribution at 0.

For $i = 1, 2, \ldots, T$:

1. The learner chooses a non-negative random variable over Ω that is the *weight function* $P(i,R)$ such that $\mathbf{\Psi}(i) \odot P(i) = 1$
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
$$B(i,R) \doteq \mathbf{E}_{l\sim Q(i,R)}\left[l\right] \tag{6}$$
4. the average loss is
$$\ell(i) = \mathbf{\Psi}(i,R) \odot \left(P(i,R)B(i,R)\right) \tag{7}$$
5. The state is updated.
$$\mathbf{\Psi}(i+1) = \mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[R \oplus Q(i,R)\right] \oplus -\ell(i) \tag{8}$$
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.

The final score is calculated: $\Phi(T) = \mathbf{\Psi}(T) \odot \phi(T)$.
The goal of the learner is to minimize this score, the goal of the adversary is to maximize it.

Figure 2: The integer time game

1. **Using conditional expectation** We can define the potential on iteration i based on the fixed potential at iteration T. Using the definition of prefix sum given in Equation (**??**, we can write this conditional expectation as
$$\forall i = 1, \ldots, T, R \quad \phi_{P,Q}(i,R) = \mathbf{E}_{\omega\sim\mathcal{D}|R^i_\omega=R}\left[\phi(T, R^T_\omega)\right] \tag{9}$$
2. **Using backward induction** It is sometimes convenient to compute the the potential for time i from the potential at time $i+1$:
$$\forall i = 1, \ldots, T-1, R \quad \phi_{P,Q}(i,R) = \mathbf{E}_{\omega\sim\mathcal{D}|M^i_\omega=R}\left[\phi_{P,Q}(i+1, R^{i+1}_\omega)\right] \tag{10}$$
by using backwards induction: $i = T-1, T-2, \ldots, 1$ we can compute the potential for all iterations.
We use Equations (6,7) and marginalizing over R to express Equation (10) in terms of the single step strategies:
$$\forall i = 1, \ldots, T-1, R \quad \phi_{P,Q}(i,R) \doteq \mathbf{E}_{r\sim[(R-\ell(i))\oplus Q(i,R)]}\left[\phi_{P,Q}(i+1, r)\right] \tag{11}$$

The score at iteration i is defined as $\Phi(i) = \mathbf{\Psi}(i) \odot \phi(i)$. The scores are all different expressions for calculating the expected final potential for the fixed strategies Q, P. Therefor the scores are all equal, as expressed in the following theorem:

Theorem 2. *Assuming $P(i,R), Q(i,R)$ are fixed for all $i = 1, \ldots, T-1$, then*
$$\mathbf{\Psi}(T) \odot \phi(T) = \Phi(T) = \Phi(T-1) = \cdots = \Phi(1) = \phi_{P,Q}(0,0)$$

A few things worth noting:

1. $\phi_{P,Q}(i,R)$ is the the final expected potential given that the paths starts at (i,R) and that the strategies used by both players in iterations $i,\ldots,T$ are fixed. Note also that which strategies were used in iterations $1,\ldots,i-1$ is of no consequence. The effect of past choices is captured by the state $\mathbf{\Psi}(i)$.
2. The final expected potential is equal to $\phi(0,0)$ which is the potential at the common starting point: $i=1$, $R=0$.

5.2 Upper and Lower potentials

Next,we vary the strategies of one side or the other to define upper and lower potentials.

$$\exists P, \quad \forall Q, \quad \forall 1 \leq i \leq T, \;\; \forall R \in \mathbb{R}, \;\; \phi^{\downarrow}_{P}(i,R) \geq \phi_{P,Q}(i,R) \tag{12}$$

$$\exists Q, \quad \forall P, \quad \forall 1 \leq i \leq T, \;\; \forall R \in \mathbb{R}, \;\; \phi^{\uparrow}_{Q}(i,R) \leq \phi_{P,Q}(i,R) \tag{13}$$

In words, $\phi^{\downarrow}_{P}$ is an upper bound on the potential that is guaranteed by the learner strategy P while $\phi^{\uparrow}_{Q_D}$ is a lower bound that is guaranteed by the adversarial strategy Q.

Following the same argument as the one leading to Theorem 2. We define upper and lower scores $\Phi^{\downarrow}_{P}(i), \Phi^{\uparrow}_{Q}(i)$ such that

$$\mathbf{\Psi}_P(T) \odot \phi(T) = \Phi^{\downarrow}_{P}(T) = \Phi^{\downarrow}_{P}(T-1) = \cdots = \Phi^{\downarrow}_{P}(0) = \phi^{\downarrow}_{P}(0,0) \tag{14}$$

and

$$\mathbf{\Psi}_Q(T) \odot \phi(T) = \Phi^{\uparrow}_{Q}(T) = \Phi^{\uparrow}_{Q}(T-1) = \cdots = \Phi^{\uparrow}_{Q}(0) = \phi^{\uparrow}_{Q}(0,0) \tag{15}$$

Our ultimate goal is to find strategies P and Q such that

$$\forall i, R, \quad \phi^{\uparrow}_{Q}(i,R) = \phi^{\downarrow}_{P}(i,R) \tag{16}$$

in particular, $\Phi^{\uparrow}_{Q}(0) = \phi^{\uparrow}_{Q}(0,0) = \phi^{\downarrow}_{P}(0,0) = \Phi^{\downarrow}_{P}(0)$. This means that Q, P are a min/max pair of strategies and that $\Phi^{\uparrow}_{Q}(0) = \Phi^{\downarrow}_{P}(0)$ define the min/max value of the game.

We do not achieve this for the integer game described in the next section. To achieve min/max optimality we extend the integer time game to the discrete time game (section 6) and to the continuous time game (7).

5.3 Strategies for the integer time game

We assume that $\phi(T) \in \mathcal{P}^2$, in other words, the final potential is positive, increasing and convex.

We define a particular adversarial strategy

$$Q_I(i,R) = \begin{cases} +1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases} \tag{17}$$

and a particular learner strategy

$$P^I(i,R) = \frac{1}{Z} \frac{\phi(i,R+2) - \phi(i,R-2)}{2} \tag{18}$$

Where Z is a normalization factor

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

We next give upper and lower bounds on the final average potential based on these strategies.

Theorem 3. *Let $\phi_T \in \mathcal{P}^2$, for any iteration $0 \leq i \leq T$ and initial regret $R_0 \in \mathbb{R}$ we define $\mathbf{\Psi}(i, R_0)$ to contain all paths that are equal to R_0 on iteration i. We consider the final score $\Phi(T)$ starting from state $\mathbf{\Psi}(i, R_0)$ and using a particular strategy*

- *The adversarial strategy (17) starting from $\mathbf{\Psi}(i, R_0)$. Guarantees a final potential*

$$\Phi(T) \geq \mathbf{E}_{R \sim R_0 \oplus \mathbb{B}(T-i,1)}\left[\phi(T, R)\right]$$

- *There learner strategy (18) guarantees*

$$\Phi(T) \leq \mathbf{E}_{R \sim R_0 \oplus \mathbb{B}(T-i,2)}\left[\phi(T, R)\right]$$

The next Lemma is the main part of the proof ot Theorem (5). We use the backward induction from Theorem (2) To compute upper and lower potentials (Equations (12,13)) for Strategies (17) and (18)

The last iteration of the game: $i = T$ is the first step of the backward induction. The uper and lower bounds are both set equal to the first step in the backward induction we define

$$\phi^{\uparrow}_{Q_I}(T, R) = \phi^{\downarrow}_{P^I}(T, R) = \phi(T, R)$$

Lemma 4. *If $\phi(i, R) \in \mathcal{P}^2$*

1. *The adversarial strategy (Eqn (17)) guarantees the lower potential*

$$\phi^{\uparrow}_{Q_I}(i, R) = \frac{\phi^{\uparrow}_{Q_I}(i, R+1) + \phi^{\uparrow}_{Q_I}(i, R-1)}{2} \tag{19}$$

2. *The learner strategy (Eqn (18)) guarantees the upper potential*

$$\phi^{\downarrow}_{P^I}(i, R) = \frac{\phi^{\downarrow}_{P^I}(i, R+2) + \phi^{\downarrow}_{P^I}(i, R-2)}{2} \tag{20}$$

Proof.

1. By symmetry adversarial strategy (17) guarantees that the aggregate loss (7) is zero regardless of the choice of the learner: $\ell(i) = 0$. Therefor the state update (8) is equivalent to the symmetric random walk:

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

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

$$\phi^{\uparrow}_{Q_I}(i-1, R) = \frac{\phi^{\uparrow}_{Q_I}(i, R-1) + \phi^{\uparrow}_{Q_I}(i, R+1)}{2} \tag{21}$$

   As this adversarial strategy is oblivious to the learner's strategy, it guarantees that the average value at iteration i is *equal* to the average of the lower value at iteration i.

2. Plugging learner's strategy (18) into equation (7) we find that

$$\ell(i) = \frac{1}{Z_i} \mathbf{E}_{R \sim \mathbf{\Psi}(i)}\left[\left(\phi^{\downarrow}_{P^I}(i, R+2) - \phi^{\downarrow}_{P^I}(i, R-2)\right) B(i, R)\right] \tag{22}$$

   Consider the score at iteration i when the learner's strategy is P^* and the adversarial strategy Q is arbitrary

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

As $\phi(i,\cdot)$ is convex and as $(y-\ell(i)) \in [-2,2]$,

$$\phi^{\downarrow}_{PI}(i-1,R+y) \leq \frac{\phi^{\downarrow}_{PI}(i,R+2)+\phi^{\downarrow}_{PI}(i,R-2)}{2} + (y-\ell(i))\frac{\phi^{\downarrow}_{PI}(i,R+2)-\phi^{\downarrow}_{PI}(i,R-2)}{2} \tag{24}$$

Combining the equations (22) and (23) we find that

$$\begin{aligned}
\Phi_{P^*,Q}(i,R) &= \mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\mathbf{E}_{y\sim Q(i)(R)}\left[\phi^{\downarrow}_{PI}(i,R+y-\ell(i))\right]\right] && (25)\\
&\leq \mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\frac{\phi^{\downarrow}_{PI}(i,R+2)+\phi^{\downarrow}_{PI}(i,R-2)}{2}\right] && (26)\\
&+ \mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\mathbf{E}_{y\sim Q(i)(R)}\left[(y-\ell(i))\frac{\phi^{\downarrow}_{PI}(i,R+2)-\phi^{\downarrow}_{PI}(i,R-2)}{2}\right]\right] && (27)
\end{aligned}$$

The final step is to show that the term (27) is equal to zero. As $\ell(i)$ is a constant with respect to R and y the term (27) can be written as:

$$\begin{aligned}
&\mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\mathbf{E}_{y\sim Q(i)(R)}\left[(y-\ell(i))\frac{\phi^{\downarrow}_{PI}(i,R+2)-\phi^{\downarrow}_{PI}(i,R-2)}{2}\right]\right] && (28)\\
=\ &\mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[B(i,R)\frac{\phi^{\downarrow}_{PI}(i,R+2)-\phi^{\downarrow}_{PI}(i,R-2)}{2}\right] && (29)\\
-\ &\ell(i)\mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\frac{\phi^{\downarrow}_{PI}(i,R+2)-\phi^{\downarrow}_{PI}(i,R-2)}{2}\right] && (30)\\
=\ &0 && (31)
\end{aligned}$$

□

Theorem 5 follows directly from Lemma 6

6 From integer to discrete time

The upper and lower bound on the final score given in Theorem 5 do not match. $\mathbf{E}_{R\sim R_0\oplus\mathbb{B}(T-i,1)}\left[\phi(T,R)\right] < \mathbf{E}_{R\sim R_0\oplus\mathbb{B}(T-i,2)}\left[\phi(T,R)\right]$. In other words, the strategies (17,18) are not a min/max pair.[2]

To close this gap we extend the integer time game into a new game we call the discrete time game (Fig. 3). The descrete time game increases the options available to the adversary, but not to the learner. As the integer step game is a special case of the new game, any upper potential that can be guaranteed by the learner in the discrete time game is also an upper potnetial for the discrete time game.

In the integer time game the loss of each action is in the range $[-1,+1]$, in the discrete time game the adversary chooses, on iteration i a step size $0 < s_i \leq 1$ which restricts the losses to the range $[-s_i,+s_i]$. Note that by always choosing $s_i = 1$, the adversary can choose to play the integer time game.

We make two additional alterations to the integer time game in order to keep the game fair. An unfair game is one where one side always wins. We list the alterations and then justify them.

1. **real-valued time** In the integer time game we use an integer to indicate the iteration number: $i = 1,2,\ldots,T$. In the discrete time game we use an positive real value, which we call "time" and use the update rule $t_{i+1} = t_i + s_i^2$, and define the final time, which is used in the regret bound, to be $\mathcal{T} = \sum_{i=0}^{T} s_i^2$

[2] There might be other (pure) strategies for the integer game that are a min/max pair, we conjecture that is not the case, and seek a extension of the game that would yield min/max strategies.

Initialization: $t_0 = 0$

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

1. If $t_i = \mathcal{T}$ the game terminates.
2. The adversay chooses a *step size* $0 < s_i \leq \min(\sqrt{1 - t_i}, 1)$, 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 $[-s_i, +s_i]$: $Q(t, \cdot) : \mathbb{R} \to \Delta^{[-s_i, +s_i]}$
5. The aggregate loss is calculated:

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

   Such that $|\ell(t_i)| \leq s_i^2$

6. The state is updated.

$$\boldsymbol{\Psi}(t_i) = \mathbf{E}_{R \sim \boldsymbol{\Psi}(t_i)}\left[Q(t_i, R) \oplus (R - \ell(t_i))\right]$$

   Where $\oplus$ is a convolution as defined in the preliminaries.

Upon termination, the final value is calculated:

$$\Phi(\mathcal{T}) = \boldsymbol{\Psi}(\mathcal{T}) \odot \phi(\mathcal{T})$$

Figure 3: The discrete time game

2. **Bounded average loss** We restrict the average loss to a range much smaller than $[-s_i, +s_i]$, specifically: $|\ell(i)| \leq s_i^2$

Note that both of these conditions hold trivially when $s_i = 1$

1. **Justification of real-valued time** To justify these choices we consider the following adversarial strategy for the discrete time game:

$$Q_D[s, p](t, R) = \begin{cases} +s & \text{w.p. } p \\ -s & \text{w.p. } 1 - p \end{cases} \tag{33}$$

   From Equation (15) we get that the initial score,

$$\Phi^{\uparrow}_{Q_D}(0) = \Phi^{\uparrow}_{Q_D}(T) = \boldsymbol{\Psi}_{Q_D}(T) \odot \phi(T)$$

   On the other hand, we know that $\boldsymbol{\Psi}_{Q_D}(T) = \mathbb{B}(T, s)$. Suppose T is large enough that the normal approximation for the binomial can be used. Let $\mathcal{N}(\mu, \sigma^2)$ be the normal distribution with mean μ and variance σ^2.

$$\lim_{T \to \infty} \Phi^{\uparrow}_{Q_D}(0) = \mathcal{N}(0, Ts^2) \odot \phi(T)$$

   Recall that $\phi(T)$ is a fixed strictly convex function. It is not hard to see that if $Ts^2 \to 0$ minimizes $\Phi^{\uparrow}_{Q_D}(0)$ and makes it equal to to $\phi(T, 0)$, which means that the learner wins, while if $Ts^2 \to \infty$, $\Phi^{\uparrow}_{Q_D}(0) \to \infty$ which means that the adversary wins. In order to keep the game balanced keep Ts^2 constant as we let $s \to 0$. We achieve that by defining the real-valued discrete time as $t_j = \sum_{i=0}^{j-1} s_i^2$.

2. **Justification of bounding average loss** Suppose the game is played for T itertions and that the adversary uses the strategy $Q_D\left[s, \frac{1}{2}+\epsilon\right](t,R)$ and that $s=\frac{1}{\sqrt{T}}$. In this case the loss of the learner in iteration i is $\ell(i) = 2s\epsilon$ and the total loss is

$$L_\ell^T = \sum_{i=0}^{T-1} \ell(i) = T2\epsilon s = \frac{2\epsilon}{s}$$

.

If ϵ is kept constant as $s \to 0$ then $\lim_{T\to\infty} L_\ell^T = \infty$, biasing the game towards the adversary. On the other hand, if $\epsilon = s^\alpha$ for $\alpha < 1$ then $L_\ell^T \to 0$, biasing the game towards the learner. To keep the game balanced we have to set $\epsilon = cs$ for some constant c. Withoutloss of generality we set $c=1$.

Generaliziing this to the game where the adversary can choose a different s_i in each iteration we get the constraint $|\ell(i)| \leq s_i^2$

6.1 Strategies for discrete time game

Now that the game is defined, we define a sequence of strategies for the adversary, indexed by k and a single strategy for the learner. These strategies are scaled versions of the strategies for the integer time game (17,18).

Adversarial strategy $Q_{D(k)}$ is a scaled version of the integer strategy (Eqn (17)):

$$Q_{D(k)} = \begin{cases} +2^{-k} & \text{w.p. } \frac{1}{2} \\ -2^{-k} & \text{w.p. } \frac{1}{2} \end{cases} \tag{34}$$

The learner chooses the distribution over action after the adversary chooses the step size, which is s_k on all iterations. Recall that $t_{i+1} = t_i + s_k^2$.

$$P^d(t_i, R) = \frac{1}{Z^{1d}} \frac{\phi(t_{i+1}, R + s_k(1+s_k)) - \phi(t_{i+1}, R - s_k(1+s_k))}{2} \tag{35}$$
$$\text{where } Z^{1d} = \mathbf{E}_{R\sim\mathbf{\Psi}(t_{i+1})}\left[\frac{\phi(t_{i+1}, R + s_k(1+s_k)) - \phi(t_{i+1}, R - s_k(1+s_k))}{2}\right]$$

The learner strategy P^d is identical to P^I (Eqn (18)) when $s_k = 1$, but is different when $s_k < 1$. In particular when $s_k \to 0$, $s_k(1+s_k) \to s_k$, which, as we show below, reduces the gap between the upper and lower potentials to zero.

Theorem 5. *Let $\phi_T \in \mathcal{P}^2$, fix the step size $s_k = 2^{-k}$, let $\mathcal{T}$ be an integer multiple of s_k define $T = \frac{\mathcal{T}}{s_k^2}$ and let $t_i = is_k^2$. For any iteration $0 \leq i < T$ and initial regret $R_0 \in \mathbb{R}$ we define $\mathbf{\Psi}(t_i, R_0)$ to contain all paths that are equal to R_0 on iteration i. We consider the final score $\Phi(\mathcal{T})$ starting from state $\mathbf{\Psi}(t_i, R_0)$ and using a particular strategy*

- *The adversarial strategy (Eqn (34)) starting from $\mathbf{\Psi}(t_i, R_0)$. Guarantees a final potential*

$$\Phi(\mathcal{T}) \geq \mathbf{E}_{R\sim R_0\oplus\mathbb{B}(T-i, s_k)}\left[\phi(\mathcal{T}, R)\right]$$

- *There learner strategy (35) guarantees*

$$\Phi(\mathcal{T}) \leq \mathbf{E}_{R\sim R_0\oplus\mathbb{B}(T-i, s_k(1+s_k))}\left[\phi(\mathcal{T}, R)\right]$$

The next Lemma is the main part of the proof ot Theorem (5). We use the backward induction from Theorem (2) To compute upper and lower potentials (Equations (12,13)) for Strategies (17) and (18)

The last iteration of the game: $i = T$ is the first step of the backward induction. The uper and lower bounds are both set equal to the first step in the backward induction we define

$$\phi_{Q_I}^{\uparrow}(T, R) = \phi_{P^I}^{\downarrow}(T, R) = \phi(T, R)$$

Lemma 6. *If $\phi(i,R) \in \mathcal{P}^2$*

1. *The adversarial strategy (Eqn (17)) guarantees the lower potential*

$$\phi^{\uparrow}_{Q_I}(i,R) = \frac{\phi^{\uparrow}_{Q_I}(i,R+1) + \phi^{\uparrow}_{Q_I}(i,R-1)}{2} \tag{36}$$

2. *The learner strategy (Eqn (18)) guarantees the upper potential*

$$\phi^{\downarrow}_{P^I}(i,R) = \frac{\phi^{\downarrow}_{P^I}(i,R+2) + \phi^{\downarrow}_{P^I}(i,R-2)}{2} \tag{37}$$

$$\begin{aligned}
\phi^{\downarrow}_{P^d}(t_i,R) &= \mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\mathbf{E}_{y\sim Q(t_i,R)}\left[\phi^{\downarrow}_{P^d}(t_i,R+y-\ell(t_i))\right]\right] && (38)\\
&\leq \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\frac{\phi^{\downarrow}_{P^d}(i,R+s_k(1+s_k)) + \phi^{\downarrow}_{P^d}(i,R-s_k(1+s_k))}{2}\right] && (39)\\
&+ \mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\mathbf{E}_{y\sim Q(i)(R)}\left[(y-\ell(i))\frac{\phi^{\downarrow}_{P^d}(i,R+s_k+s_k^2) + \phi^{\downarrow}_{P^d}(i,R-s_k-s_k^2)}{2}\right]\right] && (40)
\end{aligned}$$

(Eqn 18). We follow the same line of argument as the second part of the proof of Lemma 6 to give a recursion for the upper potential. The citical difference between the integer game is and the dicrete game is that in the discrete game $\ell(t_i) \leq s_i^2$ which implies that $(y - \ell(t_i)) \in [-s_k(1+s_k), s_k(1+s_k)]$. This yields

Following the same line of argument as the first part of the proof of Lemma 6 we consider the time points: $t_i = is_k^2 = i2^{-2k}\mathcal{T}$ for $i = 0, 1, \ldots, 2^{2k}$.

$$\phi^{\uparrow}_{Q_D}(t_{i-1},R) = \frac{\phi^{\uparrow}_{Q_D}(t_i,R-s_k) + \phi^{\uparrow}_{Q_D}(t_i,R+s_k)}{2} \tag{41}$$

ToDo: Follow the proof of Lemma 6.Derive upper and lower scores for the two strategies. Show that they converge to the same thing as $k \to \infty$.

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 $\mathbf{\Psi}(t_i)$ is transformed to $\mathbf{\Psi}(t_i)$. To choose it's strategy, the learner needs to assign value possible states $\mathbf{\Psi}(t_i)$. 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.

Note that the learner chooses a distribution *after* the adversary set the value of s_i. The discrete time version of P^1

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.

Theorem 7.
Assume $\phi(\mathcal{T},R) \in \mathcal{P}^4$ and let $A = N(0,\sqrt{\mathcal{T}}) \odot \phi(\mathcal{T})$. then for any $\epsilon > 0$

- *There exists a strategy for the adversary that guarantees, against any learner, that $\Phi(T) \geq A - \epsilon$*
- *There exists a strategy for the learner that guarantees, against any adversary, that $\Phi(T) \leq A + \epsilon$.*

6.2 The adversary prefers smaller steps

As noted before, if the adversary chooses $s_i = 1$ for all i the game reduces the 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 somewhat 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 $s_i \to 0$, as it turns out, this limit is well defined and corrsponds to Brownian motion, also known as Wiener process.

Consider a sequence of adversarial strategies S_k indexed by $k = 0, 1, 2,.$ The adversarial strategy S_k is corresponds to always choosing $s_i = 2^{-k}$, and repeating $Q^{1/2}_{\pm 2^{-k}}$ for $T2^{2k}$ iterations. This corresponds to the distribution created by a random walk with $T2^{2k}$ time steps, each step equal to $+2^{-k}$ or -2^{-k} with probabilities $1/2, 1/2$. Note that in order to preserve the variance, halving the step size requires incresing 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 8. *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 $0 \le t \le T$, 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 it's consequence for the online learning problem. We can restrict Theorem 8 for the case $t = 0$,$R = 0$ in which case we get an increasing sequence:

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

The limit of the strategies S_k as $k \to \infty$ is the well studied Brownian or Wiener process. We will discuss this connection in Section **??**.

We now go back to proving Theorem 8. 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 advesarial strategies S_k and S_{k+1} at a particular time point $0 \le t \le T$ 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 , $\phi(t', R)$ 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 9. *If $\phi(t', R)$ as function of R is in $\mathcal{P}^4$, then*

- $\phi_k(t, R) < \phi_{k+1}(t, R)$
- *Both $\phi_k(t, R)$ and $\phi_{k+1}(t, R)$ are in $\mathcal{P}^4$.*

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

- the difference between the values is

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

- Our goal is to show that the RHS is positive. Therefor we can divide it by the positive constant $\frac{2}{3}s^4$, and call the resulting function f:

$$g(R)=\frac{1}{24s^4}(\phi(t,R+2s)-4\phi(t,R+s)+6\phi(t,R)-4\phi(t,R-s)+\phi(t,R-2s)) \tag{42}$$

The function $g(R)$ has a special form called "divided difference" that has been extensively studied [18, 3, 9]. and is closely related to to derivatives of different orders. Using this connection and the fact that $\phi(\cdot,R)\in\mathcal{P}^4$ we prove the following lemma:

The following lemma states that $g(R)>0$ for all R.

Lemma 10. *Fix $t>0$ and $s>0$, let $\phi(t,R)\in\mathcal{P}^4$ as a function of R, and let*

$$g(R)\doteq\frac{1}{24s^2}(\phi(t,R+2s)-4\phi(t,R+s)+6\phi(t,R)-4\phi(t,R-s)+\phi(t,R-2s)) \tag{43}$$

then $\forall R,\ g(R)>0$

The proof is given in appendix B

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 therefor retain their continuity and convexity properties.

□

Proof. of Theorem 8
The proof is by double induction over k and over t. For a fixed k we take a finite backward induction over $t=T-2^{-2k},T-2\times 2^{-2k},T-3\times 2^{-2k},\cdots,0$. 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 9.

The theorem follows by forward induction on k.

□

7 Brownian motion and min/max strategies

In the previous section we described a sequence of adversarial strategies $S_1,S_2,\ldots$ and a learner strategy such that ....

described a sequence of adversarial strategies

It is well known that the limit of random walks where $s\to 0$ and $\Delta t=s^2$ is the the Brownian or Wiener process (see [15]).

An alternative characterization of Brownian Process is

$$\mathbf{P}\left[X_{t+\Delta t}=x_1|X_t=x_0\right]=e^{-\frac{(x_1-x_0)^2}{2\Delta t}}$$

The backwards recursion that defines the value function is the celebrated Backwrds Kolmogorov Equation with no drift and unit variance

$$\frac{\partial}{\partial t}\phi(t,R)+\frac{1}{2}\frac{\partial^2}{\partial R^2}\phi(t,R)=0 \tag{44}$$

Given a final value function with a strictly positive fourth derivative we can use Equation (44) to compute the value function for all $0\le t\le T$. We will do so in he next section.

8 Stable potential functions and anytime strategies

The potential functions, $\phi(t,R)$ is a solution of PDE (44):

$$\frac{\partial}{\partial t}\phi(t,R) + \frac{1}{2}\frac{\partial^2}{\partial r^2}\phi(t,R) = 0 \tag{45}$$

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 $\phi(t,R)$ regarless 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 44, 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 ϵ we choose $\eta = \sqrt{\frac{\ln(1/\epsilon)}{t}}$ we get the regret bound that holds for any $t > 0$

$$R_\epsilon \leq \sqrt{2t \ln \frac{1}{\epsilon}} \tag{46}$$

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

The NormalHedge value is

$$\phi_{\text{NH}}(R,t) = \begin{cases} \frac{1}{\sqrt{t+\nu}} \exp\left(\frac{R^2}{2(t+\nu)}\right) & \text{if } R \geq 0 \\ \frac{1}{\sqrt{t+\nu}} & \text{if } R < 0 \end{cases} \tag{47}$$

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:

$$R_\epsilon \leq \sqrt{(t+\nu)\left(\ln(t+\nu) + 2\ln\frac{1}{\epsilon}\right)} \tag{48}$$

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

9 The continuous time game and bounds for easy sequences

In Section 6 we have shown that the integer time game has a natural extension to a setting where $\Delta t_i = s_i^2$. We also demonstrated sequences of adversarial strategies $S_1, S_2, \ldots$ such that $\sup_{k\to\infty} \phi^{\uparrow}_{Q_k}(0,R) =$

We characterized the optimal adversarial strategy for the discrete time game (Section **??**), 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 $\mathbf{\Psi}(i) = \mathbf{\Psi}(i+1)$. 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 (**??**) For concreteness we include the learner's strategy, which is the limit of the strategy in the discrete game when $s_i \to 0$.

Set $t_1 = 0$
Fix maximal step $0 < s < 1$
On iteration $i = 1, 2, \ldots$

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

$$P^{cc}(t,R) = \frac{1}{Z^{cc}} \left.\frac{\partial}{\partial r}\right|_{r=R} \phi(t,r) \text{ where } Z^{cc} = \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\left.\frac{\partial}{\partial r}\right|_{r=R} \phi(t,r)\right] \tag{49}$$

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

$$\ell(t_i) = \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[P^{cc}(t_i,R)B(t_i,R)\right], \quad \text{where } B(t_i,R) \doteq \mathbf{E}_{y\sim Q(t_i,R)}\left[y\right] \tag{50}$$

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

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

$$\Delta t_i = \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[H(t_i,R)\ \ \mathbf{E}_{y\sim Q(t_i,R)}\left[(y - \ell(t_i))^2\right]\right] \tag{51}$$

   Where

$$H(t_i,R) = \frac{1}{Z^H} \left.\frac{\partial^2}{\partial r^2}\right|_{r=R} \phi(t_i,r) \text{ and } Z^H = \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\left.\frac{\partial^2}{\partial r^2}\right|_{r=R} \phi(t_i,r)\right] \tag{52}$$

6. The state is updated.

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

Figure 4: The continuous time game and learner strategy

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, τ)-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 τ 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 τ controls the error term.

Theorem 11. *Let $\phi \in \mathcal{P}^{\infty}$ be a potential function that satisfies the Kolmogorov backward equation (44). Fix the total time τ and let G_n be an $(n, \sqrt{\frac{\tau}{n}}, \tau)$-bounded game. Let $n \to \infty$.*

Then

$$\Phi(\mathbf{\Psi}(\tau)) \leq \Phi(\mathbf{\Psi}(0)) + O\left(\frac{1}{\sqrt{n}}\right)$$

The proof is given in appendix C

If we define

$$V_n = t_n = \sum_{i=1}^{n} \Delta t_i = \sum_{i=1}^{n} \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\mathbf{E}_{y \sim Q(t_i, R)} \left[H(t_i, R)((y - \ell(t_i))^2) \right] \right] \tag{53}$$

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

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 Boor. 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. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.

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

[15] Mark Kac. Random walk and the theory of brownian motion. *The American Mathematical Monthly*, 54(7P1):369–391, 1947.

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

[17] Haipeng Luo and Robert E Schapire. Achieving all with no parameters: Adanormalhedge. In *Conference on Learning Theory*, pages 1286–1304. PMLR, 2015.

[18] Tiberiu Popoviciu. Sur certaines inégalités qui caractérisent les fonctions convexes. *Analele Stiintifice Univ. "Al. I. Cuza", Iasi, Sectia Mat*, 11:155–164, 1965.

[19] Robert E Schapire. Drifting games. *Machine Learning*, 43(3):265–291, 2001.

[20] Volodimir G Vovk. Aggregating strategies. *Proc. of Computational Learning Theory, 1990*, 1990.

A Proof of Theorem 1

Proof.

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

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

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

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

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

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

  $$\begin{aligned}
  \mathbf{E}_{R\sim\mathbf{\Psi}}[\phi(R)] &= \mathbf{E}_{R\sim\mathbf{\Psi}}[\phi(R)\mathbf{P}_{R'\sim\mathbf{\Psi}}[\phi(R') \geq \phi(R)]] && (54)\\
  &\leq \mathbf{E}_{R\sim\mathbf{\Psi}}[\phi(R)\mathbf{P}_{R'\sim\mathbf{\Psi}}[R' \geq R]] && (55)\\
  &< \mathbf{E}_{R\sim\mathbf{\Psi}}[\phi(R)B(R)] && (56)\\
  &= \mathbf{E}_{R\sim\mathbf{\Psi}}\left[\frac{B(R)}{B(R)}\right] = \mathbf{E}_{R\sim\mathbf{\Psi}}[1] = 1 && (57)
  \end{aligned}$$

□

B Divided differences of a function

the lhs has the form A function $\phi(T, R)$ that satisfies inequality **??** is said to be *4'th order convex* (see details in in [3]).

Following[3] we give a brief review of divided differences and of n-convexity.

Let $f : [a, b] \to \mathbb{R}$ be a function from the segment $[a, b]$ to the reals.

Definition 5 (n'th order divided difference of a function)**.** *The n'th order divided different of a function $f : [a, b] \to \mathbb{R}$ at mutually distinct and ordered points $a \leq x_0 < x_1 < \cdots < x_n \leq b$ defined recursively by*

$$[x_i; f] = f(x_i),\ i \in 0, \ldots n,$$

$$[x_0, \ldots, x_n; f] = \frac{[x_1, \ldots, x_n; f] - [x_0, \ldots, x_{n-1}; f]}{x_n - x_0}$$

Definition 6 (n-convexity)**.** *A function $f : [a,b] \to \mathbb{R}$ is said to be n-convex $n \geq 0$ if and only if for all choices of $n+1$ distinct points: $a \leq x_0 < x_1 < \cdots < x_n \leq b$, $[x_0, \ldots, x_n; f] \geq 0$ holds.*

n-convexity is has a close connection to the sign of $f^{(n)}$ - the n'th derivative of f, this connection was proved in 1965 by popoviciu [18].

Theorem 12. *If $f^{(n)}$ exists then f is n-convex if and only if $f^{(n)} \geq 0$.*

The next lemma states that the function $g(R) > 0$ as defined in Equation (42).

Proof. **of Lemma (10)**

Fix t and define $f(x) = \phi(t,x)$. Let $(x_0, x_1, x_2, x_3, x_4) = (R-2s, R-s, R, R+s, R+2s)$

Using this notation we can rewrite $g(R)$ in the form

$$h(x_0, x_1, x_2, x_3, x_4) = \frac{1}{24s^4}(f(x_4) - 4f(x_3) + 6f(x_2) - 4f(x_1) + f(x_0)) \tag{58}$$

k

Is the 4-th order divided difference of $\phi(t, \cdot)$

1.

$$[x_i; f] = f(x_i)$$

2.

$$[x_i, x_{i+1}; f] = \frac{f(x_{i+1}) - f(x_i)}{s}$$

3.

$$[x_i, x_{i+1}, x_{i+2}; f] = \frac{\frac{f(x_{i+2})-f(x_{i+1})}{s} - \frac{f(x_{i+1})-f(x_i)}{s}}{2s} = \frac{f(x_{i+2}) - 2f(x_{i+1}) + f(x_i)}{2s^2}$$

4.

$$\begin{aligned} [x_i, x_{i+1}, x_{i+2}, x_{i+3}; f] &= \frac{\frac{f(x_{i+3})-2f(x_{i+2})+f(x_{i+1})}{2s^2} - \frac{f(x_{i+2})-2f(x_{i+1})+f(x_i)}{2s^2}}{3s} \\ &= \frac{f(x_{i+3}) - 3f(x_{i+2}) + 3f(x_{i+1}) - f(x_i)}{6s^3} \end{aligned}$$

5.

$$\begin{aligned} [x_i, x_{i+1}, x_{i+2}, x_{i+3}, x_{i+4}; f] &= \frac{\frac{f(x_{i+4})-3f(x_{i+3})+3f(x_{i+2})-f(x_{i+1})}{6s^3} - \frac{f(x_{i+3})-3f(x_{i+2})+3f(x_{i+1})-f(x_i)}{6s^3}}{4s} \\ &= \frac{f(x_{i+4}) - 4f(x_{i+3}) + 6f(x_{i+2}) - 4f(x_{i+1}) + f(x_i)}{24s^4} \end{aligned}$$

□

C Proof of Theorem 11

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 $\mathbf{\Psi}$ be a distribution over $\mathbb{R}$. If $\mathbf{E}_{x\sim\mathbf{\Psi}}[f(x)]$ is well-defined (and finite) , then $\mathbf{E}_{x\sim\mathbf{\Psi}}[h(x)f'(x)]$ is well defined (and finite) as well.*

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

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

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

$$f(x_0+\Delta x, y_0+\Delta y) = f(x_0, y_0) + \left\{ \frac{\partial}{\partial x}\Big|_{\substack{x,\,y\,=\\ x_0,\,y_0}} f(x,y) \right\} \Delta x + \left\{ \frac{\partial}{\partial y}\Big|_{\substack{x,\,y\,=\\ x_0,\,y_0}} f(x,y) \right\} \Delta y \tag{59}$$

$$+ \quad \frac{1}{2}\left\{ \frac{\partial^2}{\partial x^2}\Big|_{\substack{x,\,y\,=\\ x_0,\,y_0}} f(x,y) \right\} \Delta x^2 + \left\{ \frac{\partial^2}{\partial x \partial y}\Big|_{\substack{x,\,y\,=\\ x_0,\,y_0}} f(x,y) \right\} \Delta x \Delta y + \frac{1}{2}\left\{ \frac{\partial^2}{\partial y^2}\Big|_{\substack{x,\,y\,=\\ x_0,\,y_0}} f(x,y) \right\} \Delta y^2 \tag{60}$$

$$+ \quad \frac{1}{6}\left\{ \frac{\partial^3}{\partial x^3}\Big|_{\substack{x,\,y\,=\\ x_0\,+\,t\Delta x,\,y_0\,+\,t\Delta y}} f(x,y) \right\} \Delta x^3 + \frac{1}{2}\left\{ \frac{\partial^3}{\partial x^2 \partial y}\Big|_{\substack{x,\,y\,=\\ x_0\,+\,t\Delta x,\,y_0\,+\,t\Delta y}} f(x,y) \right\} \Delta x^2 \Delta y \tag{61}$$

$$+\frac{1}{2}\left\{ \frac{\partial^3}{\partial x \partial y^2}\Big|_{\substack{x,\,y\,=\\ x_0\,+\,t\Delta x,\,y_0\,+\,t\Delta y}} f(x,y) \right\} \Delta x \Delta y^2 + \frac{1}{6}\left\{ \frac{\partial^3}{\partial y^3}\Big|_{\substack{x,\,y\,=\\ x_0\,+\,t\Delta x,\,y_0\,+\,t\Delta y}} f(x,y) \right\} \Delta y^3 \tag{62}$$

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

Proof. of Lemma 14 Let $F : [0,1] \to \mathbb{R}$ 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:

$$\frac{d^n}{dt^n}F(t) = \sum_{m=1}^{n} \binom{n}{m} \frac{\partial^n}{\partial x^m \partial y^{n-m}} f(x_0 + t\Delta x, y_0 + t\Delta y)\Delta x^m \Delta y^{n-m} \tag{63}$$

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

$$f(x_0 + \Delta x, y_0 + \Delta y) = F(1) = F(0) + \frac{d}{dt}F(0) + \frac{1}{2}\frac{d^2}{dt^2}F(0) + \frac{1}{6}\frac{d^3}{dt^3}F(t') \tag{64}$$

Where $0 \le t' \le 1$. Using Eqn (63) to expand each term in Eqn. (64) completes the proof. □

Proof. of Theorem 11
We prove the claim by an upper bound on the increase of potential that holds for any iteration $1 \le i \le n$:

$$\Phi(\mathbf{\Psi}(t_{i+1})) \le \Phi(\mathbf{\Psi}(t_i)) + a s_i^3 \text{ for some constant } a > 0 \tag{65}$$

Summing inequality (65) over all iterations we get that

$$\Phi(\mathbf{\Psi}(T)) \le \Phi(\mathbf{\Psi}(0)) + c\sum_{i=1}^{n} s_i^3 \le \Phi(\mathbf{\Psi}(0)) + as\sum_{i=1}^{n} s_i^2 = \Phi(\mathbf{\Psi}(0)) + asT \tag{66}$$

From which the statement of the theorem follows.

We now prove inequality (65). 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

$$\phi(t_i + \Delta t_i, R_i + r_i) = \phi(t_i, R_i) \tag{67}$$

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

$$+ \left\{ \frac{\partial}{\partial \tau}\bigg|_{\substack{\tau,\, \rho\, = \\ t_i,\, R}} \phi(\tau, \rho) \right\} \Delta t_i \tag{69}$$

$$+ \frac{1}{2} \left\{ \frac{\partial^2}{\partial \rho^2}\bigg|_{\substack{\tau,\, \rho\, = \\ t_i,\, R}} \phi(\tau, \rho) \right\} r_i^2 \tag{70}$$

$$+ \left\{ \frac{\partial^2}{\partial r \partial \tau}\bigg|_{\substack{\tau,\, \rho\, = \\ t_i,\, R}} \phi(\tau, \rho) \right\} r_i \Delta t_i \tag{71}$$

$$+ \frac{1}{2} \left\{ \frac{\partial^2}{\partial \tau^2}\bigg|_{\substack{\tau,\, \rho\, = \\ t_i,\, R}} \phi(\tau, \rho) \right\} \Delta t_i^2 \tag{72}$$

$$+ \frac{1}{6} \left\{ \frac{\partial^3}{\partial \rho^3}\bigg|_{\substack{\tau,\, \rho\, = \\ t_i\, +\, g\Delta t_i,\, R_i\, +\, g r_i}} \phi(\tau, \rho) \right\} r_i^3 \tag{73}$$

$$+ \frac{1}{2} \left\{ \frac{\partial^3}{\partial \rho^2 \partial \tau}\bigg|_{\substack{\tau,\, \rho\, = \\ t_i\, +\, g\Delta t_i,\, R_i\, +\, g r_i}} \phi(\tau, \rho) \right\} r_i^2 \Delta t_i \tag{74}$$

$$+ \frac{1}{2} \left\{ \frac{\partial^3}{\partial \rho \partial \tau^2}\bigg|_{\substack{\tau,\, \rho\, = \\ t_i\, +\, g\Delta t_i,\, R_i\, +\, g r_i}} \phi(\tau, \rho) \right\} r_i \Delta t_i^2 \tag{75}$$

$$+ \frac{1}{6} \left\{ \frac{\partial^3}{\partial \tau^3}\bigg|_{\substack{\tau,\, \rho\, = \\ t_i\, +\, g\Delta t_i,\, R_i\, +\, g r_i}} \phi(\tau, \rho) \right\} \Delta t_i^3 \tag{76}$$

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

By assumption ϕ satisfies the Kolmogorov backward equation:

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

Combining this equation with the exchangability of the order of partial derivative (Clairiaut's Theorem) we can substitute all partial derivatives with respect to τ with partial derivatives with respect to ρ 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

$$
\begin{align}
\phi(t_i + \Delta t_i, R_i + r_i) &= \phi(t_i, R_i) \tag{77}\\
&+ \left\{ \left.\frac{\partial}{\partial \rho}\right|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} r_i \tag{78}\\
&+ \left\{ \left.\frac{\partial^2}{\partial \rho^2}\right|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} \left(\frac{r_i^2}{2} - \Delta t_i \right) \tag{79}\\
&- \left\{ \left.\frac{\partial^3}{\partial \rho^3}\right|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} r_i \Delta t_i \tag{80}\\
&+ \frac{1}{2} \left\{ \left.\frac{\partial^4}{\partial \rho^4}\right|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} \Delta t_i^2 \tag{81}\\
&+ \frac{1}{6} \left\{ \left.\frac{\partial^3}{\partial \rho^3}\right|_{\substack{\tau,\,\rho\,=\\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \phi(\tau,\rho) \right\} r_i^3 \tag{82}\\
&- \frac{1}{2} \left\{ \left.\frac{\partial^4}{\partial \rho^4}\right|_{\substack{\tau,\,\rho\,=\\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \phi(\tau,\rho) \right\} r_i^2 \Delta t_i \tag{83}\\
&+ \frac{1}{2} \left\{ \left.\frac{\partial^5}{\partial \rho^5}\right|_{\substack{\tau,\,\rho\,=\\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \phi(\tau,\rho) \right\} r_i \Delta t_i^2 \tag{84}\\
&- \frac{1}{6} \left\{ \left.\frac{\partial^6}{\partial \rho^6}\right|_{\substack{\tau,\,\rho\,=\\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \phi(\tau,\rho) \right\} \Delta t_i^3 \tag{85}
\end{align}
$$

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 (78):** $r_i = 2s_i \frac{r_i}{2s_i} = 2s_i h_1(r_i)$.
- **For (79):** $r_i^2 - \frac{1}{2}\Delta t_i = 4s_i^2 \frac{r_i^2 - \frac{1}{2}\Delta t_i}{4s_i^2} = 4s_i^2 h_2(r_i, \Delta t_i)$
- **For (80):** $r_i \Delta t_i = 2s_i^3 \frac{r_i \Delta t_i}{2s_i^3} = 2s_i^3 h_3(r_i, \Delta t_i)$
- **For (81):** $\Delta t_i^2 = s_i^4 \frac{\Delta t_i^2}{s_i^4} = s_i^3 h_4(\Delta t_i)$
- **For (82):** $r_i^3 = 8s_i^3 \frac{r_i^3}{8s_i^3} = 8s_i^3 h_5(r_i, \Delta t_i)$
- **For (83):** $r_i^2 \Delta t_i = 4s_i^4 \frac{r_i^2 \Delta t_i}{4s_i^4} = 4s_i^3 h_6(r_i, \Delta t_i)$
- **For (84):** $r_i \Delta t_i^2 = 2s_i^5 \frac{r_i \Delta t_i^2}{2s_i^5}$

- **For (85):** $\Delta t_i^3 = s_i^6 \frac{\Delta t_i^3}{s_i^6}$

We therefor get the simplified equation

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

and therefor

$$
\begin{aligned}
\phi(t_i+\Delta t_i, R+r) &= \phi(t_i,R) + \left\{ \frac{\partial}{\partial r}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} r \\
&+ \left\{ \frac{\partial^2}{\partial r^2}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} (r^2 - \Delta t_i) + O(s^3) \qquad (86)
\end{aligned}
$$

Our next step is to consider the expected value of (86) wrt $R \sim \mathbf{\Psi}(t_i)$, $y \sim Q(t_i,R)$ for an arbitrary adversarial strategy Q.

We will show that the expected potential does not increase:

$$
\mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\mathbf{E}_{y\sim Q(t_i,R)}\left[\phi(t_i+\Delta t_i, R+y-\ell(t_i))\right]\right] \leq \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\phi(t_i,R)\right] \qquad (87)
$$

Plugging Eqn (86) into the LHS of Eqn (87) we get

$$
\begin{aligned}
&\mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\mathbf{E}_{y\sim Q(t_i,R)}\left[\phi(t_i+\Delta t_i, R+y-\ell(t_i))\right]\right] && (88)\\
&= \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\phi(t_i,R)\right] && (89)\\
&+ \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\mathbf{E}_{y\sim Q(t_i,R)}\left[\left\{ \frac{\partial}{\partial r}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\}(y-\ell(t_i))\right]\right] && (90)\\
&+ \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\mathbf{E}_{y\sim Q(t_i,R)}\left[\left\{ \frac{\partial^2}{\partial r^2}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\}((y-\ell(t_i))^2-\Delta t_i)\right]\right] && (91)\\
&+ O(s^3) && (92)
\end{aligned}
$$

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.[3]

To prove inequality (65), we need to show that the terms 90 and 91 are smaller or equal to zero.

[3] I need to clean this up and find an argument that the expected value for mixed derivatives is also finite.

Term (90) is equal to zero:
As $\ell(t_i)$ is a constant relative to R and y, and $\left\{ \frac{\partial}{\partial r}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\}$ is a constant with respect to y we can rewrite (90) as

$$\mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\left\{ \frac{\partial}{\partial r}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} \mathbf{E}_{y\sim Q(t_i,R)}\left[y\right]\right] - \ell(t_i)\mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\left\{ \frac{\partial}{\partial r}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\}\right] \quad (93)$$

Combining the definitions of $\ell(t)$ (50) and and the learner's strategy P^{cc} (49) we get that

$$\ell(t_i) \quad = \quad \mathbf{E}_{R\sim\mathbf{\Psi}t_i}\left[\frac{1}{Z}\left\{ \frac{\partial}{\partial r}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} \mathbf{E}_{y\sim Q(i,R)}\left[y\right]\right] \text{ where } Z = \mathbf{E}_{R\sim\mathbf{\Psi}t_i}\left[\frac{1}{Z}\left\{ \frac{\partial}{\partial r}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\}\right] (94)$$

Plugging (94) into (93) and recalling the requirement that $\ell(t_i) < \infty$ we find that term (90) is equal to zero.

Term (91) is equal to zero:
As Δt_i is a constant relative to y, we can take it outside the expectation and plug in the definition of Δt_i (51)

$$\mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\mathbf{E}_{y\sim Q(t_i,R)}\left[Q(t_i,R)\right]\left\{ \frac{\partial^2}{\partial r^2}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\}(y-\ell(t_i))^2 - \Delta t_i\right] = \Delta t_i - \Delta t_i = 0 \quad (95)$$

Where $G(t_i, R)$ is defined in Equation (**??**) We find that (91) is zero.

Finally (92) is negligible relative to the other terms as $s \to 0$. □