Potential-based hedging algorithms

Yoav Freund

September 20, 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 bounds on the learning algorithm in the context of Ito processes.

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. [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 *instantanous 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 $R_1^t \geq R_2^t \geq \cdots \geq R_k^t \geq \cdots$ and define R_k^t to as the regret relative to the $\epsilon = k/M$ top percentile, denote R_ϵ^t. 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.

Formally, we say that the distribution over regrets $\mathbf{\Psi}$ satisfied a simultanous bound B if

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

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

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

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

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

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

We next show that there is a one to one relationship between 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 $\phi(R) = B(R)^{-1}$*

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:

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

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 organizes as follows.

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, 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 Φ^T. To do so we define potential functions for intermediate steps: $0 \leq t < T$.[1]

3 Main results

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

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

[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 [17, 11].

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.

4 Preliminaries

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

Our results apply to potential functions with positivity constraint defined as follows.

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.*

Theorem 1 implies that any potential function has strict connectivity of degree 1. We will initially restrict ourselves to potential functions that are strictly convex, i.e. have strict positivity of degree 2. Later on, in section 8.3, we will further restrict our potential functions to have strict positivity of degree 4.

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 $\mathbf{P}[u] = \mathbf{P}[v] = \frac{1}{2}\mathbf{P}[s]$.

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:

$$S_\omega = (R_\omega^1, R_\omega^2, \ldots, R_\omega^T) \tag{7}$$

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:

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

We define the *score* at iteration i as

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

Initialization:

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

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

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

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

Figure 1 describes the integer time game. However it does contain a definition for who wins the game. We consider two definitions:

- The learner wins if the final distribution $\phi(T)$ satisfies the simultanous regret bound G as defined in 1
- The learner wins if the final distribution $\phi(T)$ satisfies $\mathbf{\Psi}(T) \cdot \phi(T) \leq 1$ as defined in 2.

From theorem 1 we know that if we set the potential as $\phi(R) = \frac{1}{G(x)}$ then the two conditions are equivalent.

The simultanous bound is our ultimate goal, the advantage of the potential game is that intermediate potential functions can be defined, thus decoupling iterations of the game.

We therefore fix the length of the game T and the final potential function $\phi(T)$. We define the goal of the learner to minimize $\mathbf{\Psi}(T) \cdot \phi(T)$ and the goal of the adversary to maximize the same.

In the next section we define the intermediate potential functions in general. We then come back to the integer game and show good strategies for each side.

6 Potential Functions and backward induction

In the setup of the potential game, only the *final* potential function, at the end of the game, is defined. However, as we will now show, there is a natural way to define a potential function for all iterations.

An action in our game defines a path S_ω (7). Fixing the strategies of the learner and the adversary implies a fixed distribution over paths.

We define the potential for $\phi_{P,Q}(i,R)$ to be the expected value of the final potential conditioned on the path passing through cumulative regret R on iteration i. Recall that $\mathbf{\Psi} i$ is the probability distribution of R on iteration i. We can therefor calculate the final expected potential by computing the expectation of the conditional expectation with respect to $\phi(t)$. Which is summarized by the following theorem.

Fix the strategies for both the learner $P(\cdot,\cdot)$ and the adversary $Q(\cdot,\cdot)$. We denote the potential function for the fixed strategies by $\phi_{P,Q}(i,R)$.

Fixing the strategies defines a distribution over paths: distribution over paths. The potentials i,R corresponds to the expected final potentail given that the paths

The base case is $\phi_{P,Q}(T,R) = \phi(T,R)$. In the induction 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 $\ell(i)$. We then define

$$\forall R \;\; \phi_{P,Q}(i,R) \doteq \mathbf{E}_{r\sim[(R-\ell(i))\oplus Q(i,R)]}\left[\phi_{P,Q}(i+1,r)\right] \tag{8}$$

For the potentials defined using Equation (8) we have the following theorem:

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

$$\mathbf{\Psi}(T)\cdot\phi(T) = \mathbf{\Psi}(T-1)\cdot\phi_{P,Q}(T-1) = \cdots = \mathbf{\Psi}(1)\cdot\phi_{P,Q}(1) = \phi_{P,Q}(1,0)$$

Note

1. The final expected potential is equal to $\phi(1,0)$ which is the potential at the common starting point: $i=0$, $R=0$.

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

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

$$\exists P, \;\; \forall Q, \;\; \forall 1\le i\le T, \;\; \forall R\in\mathbb{R}, \;\; \phi^{\downarrow}_P(i,R) \ge \phi_{P,Q}(i,R) \tag{9}$$

$$\exists Q, \;\; \forall P, \;\;\; \forall 1\le i\le T, \;\; \forall R\in\mathbb{R}, \;\; \phi^{\uparrow}_Q(i,R) \le \phi_{P,Q}(i,R) \tag{10}$$

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

Our goal is to find strategies such that $\phi^{\uparrow}_Q = \phi^{\downarrow}_P$, as that would mean that we have found min-max strategies for both players. We will not achieve this goal, instead, we will show sequences of strategies $P(1), P(2), \ldots$ and $Q(1), Q(2), \ldots$ such that

$$\forall (i,R) \;\; \lim_{j\to\infty} \phi^{\uparrow}_{Q(j)}(i,R) = \lim_{j\to\infty} \phi^{\downarrow}_{P(j)}(i,R) \tag{11}$$

Before we attempt this goal, we start by analyzing the integer time game.

7 Strategies for the integer time game

We go back to the integer time game and show the strategies for both sides and the corresponding upper and lower potentials.

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

We define the adversarial strategy

$$Q^I(i-1,R) = \begin{cases} +1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases} \tag{12}$$

and the learner strategy

$$P^I(i-1,R) = \frac{1}{Z}\frac{\phi(i,R+2)-\phi(i,R-2)}{2} \tag{13}$$

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.

Let $B(n,a)$ denote 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 $-a,+a$ with equal probabilities.

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 (12) starting from $\mathbf{\Psi}(i,R_0)$. Guarantees a final potential*

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

- *There learner strategy (13) guarantees*

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

The next Lemma is the main part of the proof ot Theorem (3). We use the backward induction from Theorem (2) To compute upper and lower potentials (Equations (9,10)) for Strategies (12) and (13)

For the first step in the backward induction we define

$$\phi_Q^{\uparrow}(T,R) = \phi_P^{\downarrow}(T,R) = \phi(T,R)$$

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

1. *The adversarial strategy Guarantees the lower potential*

$$\phi_Q^{\uparrow}(i-1,R) = \frac{\phi_Q^{\uparrow}(i,R+1)+\phi_Q^{\uparrow}(i,R-1)}{2} \tag{14}$$

2. *The learner strategy: guarantees the upper potential*

$$\phi_P^{\downarrow}(i-1,R) = \frac{\phi_P^{\downarrow}(i,R+2)+\phi_P^{\downarrow}(i,R-2)}{2} \tag{15}$$

Proof. 1. By symmetry adversarial strategy (12) 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:

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

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

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

As this adversaril 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 (13) into equation (5) we find that

$$\ell(i-1) = \frac{1}{Z_{i-1}} \mathbf{E}_{R\sim\mathbf{\Psi}(i-1)} \left[\left(\phi_P^{\downarrow}(i, R+2) - \phi_P^{\downarrow}(i, R-2) \right) B(i-1, R) \right] \tag{17}$$

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

$$\Phi_{P^*,Q}(i-1, R) = \mathbf{E}_{R\sim\mathbf{\Psi}(i-1)} \left[\mathbf{E}_{y\sim Q(i-1)(R)} \left[\phi(i, R+y-\ell(i-1)) \right] \right] \tag{18}$$

As $\phi(i, \cdot)$ is convex and as $(y - \ell(i-1)) \in [-2, 2]$,

$$\phi(i, R+y) \le \frac{\phi(i, R+2) + \phi(i, R-2)}{2} + (y - \ell(i)) \frac{\phi(i, R+2) - \phi(i, R-2)}{2} \tag{19}$$

Combining the equations (17) and (18) we find that

$$\begin{aligned}
\Phi_{P^*,Q}(i-1, R) &= \mathbf{E}_{R\sim\mathbf{\Psi}(i-1)} \left[\mathbf{E}_{y\sim Q(i-1)(R)} \left[\phi(i, R+y-\ell(i-1)) \right] \right] && (20)\\
&\le \mathbf{E}_{R\sim\mathbf{\Psi}(i-1)} \left[\frac{\phi(i, R+2) + \phi(i, R-2)}{2} \right] && (21)\\
&+ \mathbf{E}_{R\sim\mathbf{\Psi}(i-1)} \left[\mathbf{E}_{y\sim Q(i-1)(R)} \left[(y - \ell(i-1)) \frac{\phi(i, R+2) - \phi(i, R-2)}{2} \right] \right] && (22)
\end{aligned}$$

The final step is to show that the term (22) is equal to zero. As $\ell(i-1)$ is a constant with respect to R and y the term (22) can be written as:

$$\begin{aligned}
& \mathbf{E}_{R\sim\mathbf{\Psi}(i-1)} \left[\mathbf{E}_{y\sim Q(i-1)(R)} \left[(y - \ell(i-1)) \frac{\phi(i, R+2) - \phi(i, R-2)}{2} \right] \right] && (23)\\
= \; & \mathbf{E}_{R\sim\mathbf{\Psi}(i-1)} \left[B(i-1, R) \frac{\phi(i, R+2) - \phi(i, R-2)}{2} \right] && (24)\\
- \; & \ell(i) \mathbf{E}_{R\sim\mathbf{\Psi}(i-1)} \left[\frac{\phi(i, R+2) - \phi(i, R-2)}{2} \right] && (25)\\
= \; & 0 && (26)
\end{aligned}$$

□

Proof. of Theorem 3

□

7.1 A learner strategy with a variance-dependent bound

As shown in Lemma **??**, the adversary always prefers mixed strategies that assign zero probability for all steps other than ±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 5. *The learner strategy:*

$$P^2(i-1, R) = \frac{1}{Z} \left. \frac{\partial}{\partial r} \right|_{r=R} \phi(i, r) \tag{27}$$

Where Z is a normalization factor

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

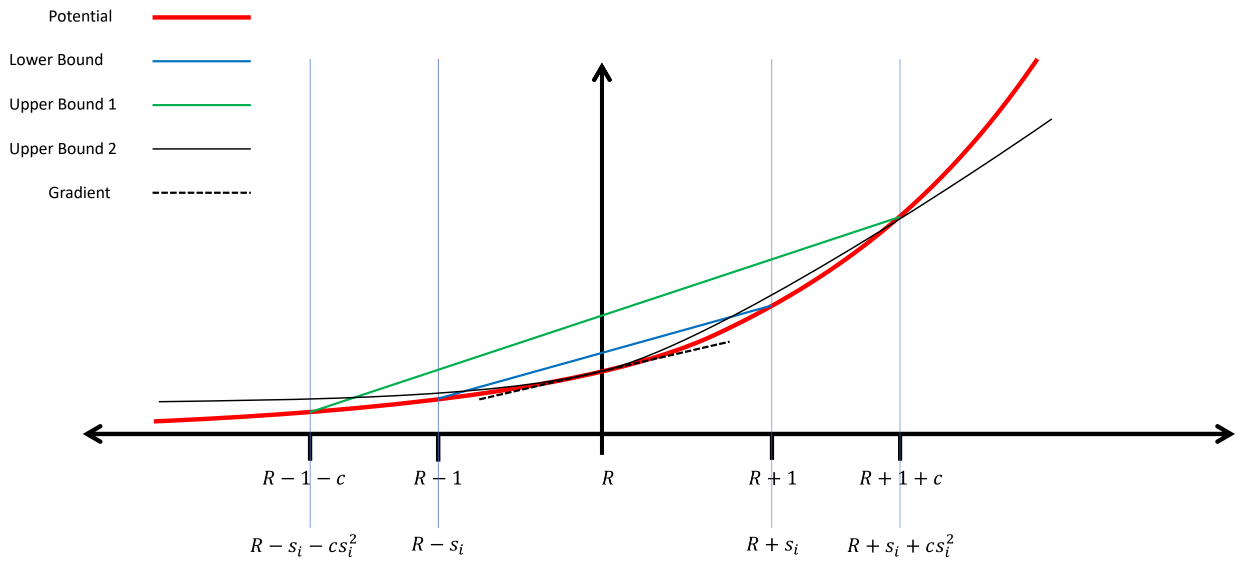

Figure 2: This figure depicts the relationship between the different upper and lower bounds used in the analysis. To aid understanding we describe the elements of the figure twice: once for the integer game, and once for the continuous time game.

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 $R-2 \leq r \leq R+2$. 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: $Q^{1/2}_{\pm s_i}(t_i, R)$. The first-order learner strategy: $P^{1c}(t_i, R)$ corresponds to the green line. The second-order learner strategy: $P^{2c}(i,R)$ corresponds to the black curve. Observe that when $s_i \to 0$ 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.

---

guarantees the following upper potential against any adversarial strategy Q

$$\phi_P^{\downarrow}(i-1,R) = \phi(i,R) + b(i,R)\mathbf{E}_{l\sim Q(i,R)}\left[l^2\right] \tag{28}$$

where $b(i,R) = \phi(i,R+2) - \phi(i,R) - 2 \left.\frac{\partial}{\partial r}\right|_{r=R} \phi(i,r)$

Proof. To Do □

We compare the bound for P^2 to the bound for P^1 given in Lemma 4. We find that when the adversary is optimal: $\mathbf{E}_{l\sim Q(t,R)}\left[l^2\right] = 1$ then the bound for P^1 is better than the bound for P^2, on the other hand, when $\mathbf{E}_{l\sim Q(t,R)}\left[l^2\right]$ 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 $i = 0, 1, \ldots$. 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 Δt_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: $\mathbf{\Psi}(i-1) = \mathbf{\Psi}(i)$, $\Delta t_i = 0$ and the instantanous regret is zero. On the other hand $\Delta t_i = 1$ corresponds to the adversarial strategy $Q^{1/2}(t-1, R)$ (Eqn. **??**) which maximizes the instantanous 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 instantanous expert lossesat 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 Δt_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 ± 1 with equal probability, we deonte 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 $\mathbf{\Psi}(0)$ which is a delta functions at $R = 0$. One iteration of $Q_{\pm 1}^{1/2}$ results in a distribution ± 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 $Q_{\pm 1/k}^{1/2}$ 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 $k \to \infty$ 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 $k \to \infty$. On the other hand, the variance increases like t.

Next lets consider effect of reducing the step size on a *biased* strategy $Q_{\pm 1}^{1/2+\gamma}$ as defined in Eqn (12) for some $0 \leq \gamma \leq 1/2$. We now figure out what γ' 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 2γ 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 $k \to \infty$, $\gamma' \to 0$. 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 $i = 1, 2, 3, \ldots$ 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 $i = 1, 2, \ldots$

1. If $t_{i-1} = T$ the game terminates.
2. The adversay 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 $[-s_i, +s_i]$: $Q(t) : \mathbb{R} \to \Delta^{[-s_i,+s_i]}$
5. The aggregate loss is calculated:
$$\ell(t_i) = \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[P(t_i,R)B(t_i,R)\right], \quad \text{where } B(t_i,R) \doteq \mathbf{E}_{y\sim Q(t_i,R)}\left[y\right] \tag{29}$$
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:
$$\mathbf{\Psi}(t_i) = \mathbf{E}_{R\sim\mathbf{\Psi}(t_{i-1})}\left[Q(t_i)(R) \oplus (R - \ell(t_i))\right]$$

When $t_i = T$ the game is terminated, and the final value is calculated:
$$\Phi(T) = \mathbf{E}_{R\sim\mathbf{\Psi}(T)}\left[\phi(T,R)(R)\right]$$

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 σ.

Theorem 6.

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

- *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 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 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 7. *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 it's consequence for the online learning problem. We can restrict Theorem 7 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. The backwards recursion that defines the value function is the celebrated Backwrds Kolmogorov Equation with zero dift and unit variance

$$\frac{\partial}{\partial t}\phi(t, R) + \frac{1}{2}\frac{\partial^2}{\partial R^2}\phi(t, R) = 0 \tag{30}$$

Given a final value function with a strictly positive fourth derivative we can use Equation (30) to compute the value function for all $0 \leq t \leq T$. We will do so in he next section.

We now go back to proving Theorem 7. 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 $0 \leq t \leq T$ 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 , $\phi(t', R)$ 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 8. *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 $\phi(T,R)$ the would garantee that

$$\phi_1(t,R) - \phi_0(t,R) = \frac{1}{16}(\phi(t',R+2) - 4\phi(t',R+1) + 6\phi(t',R) - 4\phi(t',R-1) + \phi(t',R-2)) > 0 \qquad (31)$$

Inequalities of this form have been studied extensively under the name "divided differences" [16, 3, 9]. A function $\phi(T,R)$ that satisfies inequality 31 is said to be *4'th order convex* (see details in in [3]).

n-convex functions have a very simple characterization:

Theorem 9. *Let f be a function with 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 therefor retain their continuity and convexity properties.

□

Proof. of Theorem 7
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 8.

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 $\mathbf{\Psi}(t_{i-1})$ 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.

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 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 13) is

$$P^{1d}(t_{i-1},R) = \frac{1}{Z^{1d}} \frac{\phi(t_i, R + s_{i-1} + cs_{i-1}^2) - \phi(t_i, R - s_{i-1} - cs_{i-1}^2)}{2} \qquad (32)$$
$$\text{where } Z^{1d} = \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\frac{\phi(t_i, R + s_{i-1} + cs_{i-1}^2) - \phi(t_i, R - s_{i-1} - cs_{i-1}^2)}{2}\right]$$

Next, we consider the discrete time version of P^2: (Eqn 27)

$$P^{2d}(t_{i-1},R) = \frac{1}{Z^{2d}} \left.\frac{\partial}{\partial r}\right|_{r=R} \phi(t_{i-1} + s_{i-1}^2, r) \qquad (33)$$
$$\text{where } Z^{2d} = \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\left.\frac{\partial}{\partial r}\right|_{r=R} \phi(t_{i-1} + s_{i-1}^2, r)\right]$$

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 $\boldsymbol{\Psi}(i) = \boldsymbol{\Psi}(i+1)$. 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 $s_i \to 0$.

9.0.1 The continuous time game and a learner strategy

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

1. If $t_i = T$ the game terminates.

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

$$P^{cc}(t, R) = \frac{1}{Z^{cc}} \left.\frac{\partial}{\partial r}\right|_{r=R} \phi(t, r) \text{ where } Z^{cc} = \mathbf{E}_{R\sim\boldsymbol{\Psi}(t_i)} \left[\left.\frac{\partial}{\partial r}\right|_{r=R} \phi(t, r) \right] \tag{34}$$

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:

$$\ell(t_i) = \mathbf{E}_{R\sim\boldsymbol{\Psi}(t_i)} \left[P^{cc}(t_i, R) B(t_i, R) \right], \text{ where } B(t_i, R) \doteq \mathbf{E}_{y\sim Q(t_i,R)} [y] \tag{35}$$

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

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

$$\Delta t_i = \mathbf{E}_{R\sim\boldsymbol{\Psi}(t_i)} \left[\mathbf{E}_{y\sim Q(t_i,R)} \left[\left\{ \left. \frac{\partial^2}{\partial r^2} \phi(\tau, \rho) \right|_{\substack{\tau,\,\rho\,= \\ t_i,\,R}} \right\} ((y - \ell(t_i))^2) \right] \right] \tag{36}$$

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:

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

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 10. *Let $\phi \in \mathcal{P}^{\infty}$ be a potential function that satisfies the Kolmogorov backward equation (30). Assume a sequence of (n_i, s_i, τ)-bounded games where τ is a constant $n \to \infty$ and $s = \sqrt{\frac{\tau}{n}}$*

Then

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

The proof is given in appendix B

If we define

$$V_n = t_n = \sum_{i=1}^{n} \Delta t_i = \sum_{i=1}^{n} \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)} \left[\mathbf{E}_{y\sim Q(t_i,R)} \left[\left\{ \frac{\partial^2}{\partial r^2}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \right\} (y - \ell(t_i))^2 \right] \right] \tag{37}$$

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

10 Two self-consistent potential functions

The potential functions, $\phi(t, R)$ is a solution of PDE (30):

$$\frac{\partial}{\partial t}\phi(t, R) + \frac{1}{2}\frac{\partial^2}{\partial r^2}\phi(t, R) = 0 \tag{38}$$

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 30, 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 $t > 0$

$$R_\epsilon \leq \sqrt{2t \ln\frac{1}{\epsilon}} \tag{39}$$

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

$$\phi_{\text{NH}}(R,t) = \begin{cases} \frac{1}{\sqrt{t+\nu}} \exp\left(\frac{R^2}{2(t+\nu)}\right) & \text{if } R \geq 0 \\ \frac{1}{\sqrt{t+\nu}} & \text{if } R < 0 \end{cases} \tag{40}$$

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:

$$R_\epsilon \leq \sqrt{(t+\nu)\left(\ln(t+\nu) + 2\ln\frac{1}{\epsilon}\right)} \tag{41}$$

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.

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 ϵ-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

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

then ϕ_1 *dominates* ϕ_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. 40). 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 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. 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.

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

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

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

  but $\mathbf{E}_{R\sim\boldsymbol{\Psi}}\left[\phi(R)\right] > 1$ 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 $\mathbf{\Psi}$:

$$\begin{aligned} \mathbf{E}_{R\sim\mathbf{\Psi}}\left[\phi(R)\right] &= \mathbf{E}_{R\sim\mathbf{\Psi}}\left[\phi(R)\mathbf{P}_{R'\sim\mathbf{\Psi}}\left[\phi(R') \geq \phi(R)\right]\right] && (42)\\ &\leq \mathbf{E}_{R\sim\mathbf{\Psi}}\left[\phi(R)\mathbf{P}_{R'\sim\mathbf{\Psi}}\left[R' \geq R\right]\right] && (43)\\ &< \mathbf{E}_{R\sim\mathbf{\Psi}}\left[\phi(R)B(R)\right] && (44)\\ &= \mathbf{E}_{R\sim\mathbf{\Psi}}\left[\frac{B(R)}{B(R)}\right] = \mathbf{E}_{R\sim\mathbf{\Psi}}\left[1\right] = 1 && (45) \end{aligned}$$

□

B Proof of Theorem 10

We start with two technical lemmas

Lemma 11. *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 $\mathbf{E}_{x\sim\mathbf{\Psi}}\left[f(x)\right]$ is well-defined (and finite) , then $\mathbf{E}_{x\sim\mathbf{\Psi}}\left[h(x)f'(x)\right]$ 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 $0 \leq h^+(x) \leq 1$ we get that $\mathbf{E}_{x\sim\mathbf{\Psi}}\left[f'(x)\right] = \infty$. As $f(x+1) \geq f'(x)$ we get that $\mathbf{E}_{x\sim\mathbf{\Psi}}\left[f(x)\right] = \infty$ which is a contradiction. □

Lemma 12. *Let $f(x, y)$ be a differentiable function with continuous derivatives up to degree three. Then*

$$\begin{aligned} f(x_0+\Delta x, y_0+\Delta y) = f(x_0, y_0) &+ \left\{ \frac{\partial}{\partial x} f(x,y)\Big|_{\substack{x,\,y\,=\\ x_0,\,y_0}} \right\}\Delta x + \left\{ \frac{\partial}{\partial y} f(x,y)\Big|_{\substack{x,\,y\,=\\ x_0,\,y_0}} \right\}\Delta y && (46)\\ + \ \frac{1}{2}\left\{ \frac{\partial^2}{\partial x^2} f(x,y)\Big|_{\substack{x,\,y\,=\\ x_0,\,y_0}} \right\}\Delta x^2 &+ \left\{ \frac{\partial^2}{\partial x \partial y} f(x,y)\Big|_{\substack{x,\,y\,=\\ x_0,\,y_0}} \right\}\Delta x\Delta y + \frac{1}{2}\left\{ \frac{\partial^2}{\partial y^2} f(x,y)\Big|_{\substack{x,\,y\,=\\ x_0,\,y_0}} \right\}\Delta y^2 && (47)\\ + \ \frac{1}{6}\left\{ \frac{\partial^3}{\partial x^3} f(x,y)\Big|_{\substack{x,\,y\,=\\ x_0\,+\,t\Delta x,\,y_0\,+\,t\Delta y}} \right\}\Delta x^3 &+ \frac{1}{2}\left\{ \frac{\partial^3}{\partial x^2 \partial y} f(x,y)\Big|_{\substack{x,\,y\,=\\ x_0\,+\,t\Delta x,\,y_0\,+\,t\Delta y}} \right\}\Delta x^2\Delta y && (48)\\ +\frac{1}{2}\left\{ \frac{\partial^3}{\partial x \partial y^2} f(x,y)\Big|_{\substack{x,\,y\,=\\ x_0\,+\,t\Delta x,\,y_0\,+\,t\Delta y}} \right\}\Delta x\Delta y^2 &+ \frac{1}{6}\left\{ \frac{\partial^3}{\partial y^3} f(x,y)\Big|_{\substack{x,\,y\,=\\ x_0\,+\,t\Delta x,\,y_0\,+\,t\Delta y}} \right\}\Delta y^3 && (49) \end{aligned}$$

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

Proof. of Lemma 12 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:

$$\frac{d^n}{dt^n}F(t) = \sum_{m=1}^{n} \binom{n}{m} \frac{\partial^n}{\partial x^m \partial y^{n-m}} f(x_0 + t\Delta x, y_0 + t\Delta y)\Delta x^m \Delta y^{n-m} \tag{50}$$

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

$$f(x_0 + \Delta x, y_0 + \Delta y) = F(1) = F(0) + \frac{d}{dt}F(0) + \frac{1}{2}\frac{d^2}{dt^2}F(0) + \frac{1}{6}\frac{d^3}{dt^3}F(t') \tag{51}$$

Where $0 \leq t' \leq 1$. Using Eqn (50) to expand each term in Eqn. (51) completes the proof. □

Proof. of Theorem 10
We prove the claim by an upper bound on the increase of potential that holds for any iteration $1 \leq i \leq n$:

$$\Phi(\mathbf{\Psi}(t_{i+1})) \leq \Phi(\mathbf{\Psi}(t_i)) + as_i^3 \text{ for some constant } a > 0 \tag{52}$$

Summing inequality (52) over all iterations we get that

$$\Phi(\mathbf{\Psi}(T)) \leq \Phi(\mathbf{\Psi}(0)) + c\sum_{i=1}^{n} s_i^3 \leq \Phi(\mathbf{\Psi}(0)) + as\sum_{i=1}^{n} s_i^2 = \Phi(\mathbf{\Psi}(0)) + asT \tag{53}$$

From which the statement of the theorem follows.

We now prove inequality (52). We use the notation $r = y - \ell(i)$ to denote the instantaneous regret at iteration i.

Applying Lemma 12 to $\phi(t_{i+1}, R_{i+1}) = \phi(t_i + \Delta t_i, R_i + r_i)$ we get

$$\begin{aligned}
\phi(t_i + \Delta t_i, R_i + r_i) \quad &= \quad \phi(t_i, R_i) \tag{54}\\
&+ \quad \left\{ \frac{\partial}{\partial \rho}\phi(\tau, \rho)\bigg|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \right\} r_i \tag{55}\\
&+ \quad \left\{ \frac{\partial}{\partial \tau}\phi(\tau, \rho)\bigg|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \right\} \Delta t_i \tag{56}\\
&+ \quad \frac{1}{2}\left\{ \frac{\partial^2}{\partial \rho^2}\phi(\tau, \rho)\bigg|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \right\} r_i^2 \tag{57}\\
&+ \quad \left\{ \frac{\partial^2}{\partial r \partial \tau}\phi(\tau, \rho)\bigg|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \right\} r_i \Delta t_i \tag{58}\\
&+ \quad \frac{1}{2}\left\{ \frac{\partial^2}{\partial \tau^2}\phi(\tau, \rho)\bigg|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \right\} \Delta t_i^2 \tag{59}\\
&+ \quad \frac{1}{6}\left\{ \frac{\partial^3}{\partial \rho^3}\phi(\tau, \rho)\bigg|_{\substack{\tau,\,\rho\,=\\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \right\} r_i^3 \tag{60}\\
&+ \quad \frac{1}{2}\left\{ \frac{\partial^3}{\partial \rho^2 \partial \tau}\phi(\tau, \rho)\bigg|_{\substack{\tau,\,\rho\,=\\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \right\} r_i^2 \Delta t_i \tag{61}\\
&+ \quad \frac{1}{2}\left\{ \frac{\partial^3}{\partial \rho \partial \tau^2}\phi(\tau, \rho)\bigg|_{\substack{\tau,\,\rho\,=\\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \right\} r_i \Delta t_i^2 \tag{62}\\
&+ \quad \frac{1}{6}\left\{ \frac{\partial^3}{\partial \tau^3}\phi(\tau, \rho)\bigg|_{\substack{\tau,\,\rho\,=\\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \right\} \Delta t_i^3 \tag{63}
\end{aligned}$$

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

$$\begin{align}
\phi(t_i+\Delta t_i, R_i+r_i) &= \phi(t_i, R_i) \tag{64}\\
&+ \left\{ \frac{\partial}{\partial \rho}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,= \\ t_i,\,R}} \right\} r_i \tag{65}\\
&+ \left\{ \frac{\partial^2}{\partial \rho^2}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,= \\ t_i,\,R}} \right\} \left(\frac{r_i^2}{2} - \Delta t_i\right) \tag{66}\\
&- \left\{ \frac{\partial^3}{\partial \rho^3}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,= \\ t_i,\,R}} \right\} r_i \Delta t_i \tag{67}\\
&+ \frac{1}{2}\left\{ \frac{\partial^4}{\partial \rho^4}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,= \\ t_i,\,R}} \right\} \Delta t_i^2 \tag{68}\\
&+ \frac{1}{6}\left\{ \frac{\partial^3}{\partial \rho^3}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,= \\ t_i+g\Delta t_i,\,R_i+gr_i}} \right\} r_i^3 \tag{69}\\
&- \frac{1}{2}\left\{ \frac{\partial^4}{\partial \rho^4}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,= \\ t_i+g\Delta t_i,\,R_i+gr_i}} \right\} r_i^2 \Delta t_i \tag{70}\\
&+ \frac{1}{2}\left\{ \frac{\partial^5}{\partial \rho^5}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,= \\ t_i+g\Delta t_i,\,R_i+gr_i}} \right\} r_i \Delta t_i^2 \tag{71}\\
&- \frac{1}{6}\left\{ \frac{\partial^6}{\partial \rho^6}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,= \\ t_i+g\Delta t_i,\,R_i+gr_i}} \right\} \Delta t_i^3 \tag{72}
\end{align}$$

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 (65):** $r_i = 2s_i \frac{r_i}{2s_i} = 2s_i h_1(r_i)$.
- **For (66):** $r_i^2 - \frac{1}{2}\Delta t_i = 4s_i^2 \frac{r_i^2 - \frac{1}{2}\Delta t_i}{4s_i^2} = 4s_i^2 h_2(r_i, \Delta t_i)$
- **For (67):** $r_i \Delta t_i = 2s_i^3 \frac{r_i \Delta t_i}{2s_i^3} = 2s_i^3 h_3(r_i, \Delta t_i)$

- **For (68):** $\Delta t_i^2 = s_i^4 \frac{\Delta t_i^2}{s_i^4} = s_i^3 h_4(\Delta t_i)$
- **For (69):** $r_i^3 = 8 s_i^3 \frac{r_i^3}{8 s_i^3} = 8 s_i^3 h_5(r_i, \Delta t_i)$
- **For (70):** $r_i^2 \Delta t_i = 4 s_i^4 \frac{r_i^2 \Delta t_i}{4 s_i^4} = 4 s_i^3 h_6(r_i, \Delta t_i)$
- **For (71):** $r_i \Delta t_i^2 = 2 s_i^5 \frac{r_i \Delta t_i^2}{2 s_i^5}$
- **For (72):** $\Delta t_i^3 = s_i^6 \frac{\Delta t_i^3}{s_i^6}$

We therefor get the simplified equation

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

and therefor

$$\begin{aligned}
\phi(t_i + \Delta t_i, R + r) &= \phi(t_i, R) + \left\{ \frac{\partial}{\partial r} \left. \phi(\tau, \rho) \right|_{\substack{\tau,\, \rho\, = \\ t_i,\, R}} \right\} r \\
&+ \left\{ \frac{\partial^2}{\partial r^2} \left. \phi(\tau, \rho) \right|_{\substack{\tau,\, \rho\, = \\ t_i,\, R}} \right\} (r^2 - \Delta t_i) + O(s^3) \qquad (73)
\end{aligned}$$

Our next step is to consider the expected value of (73) 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:

$$\mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\mathbf{E}_{y \sim Q(t_i, R)} \left[\phi(t_i + \Delta t_i, R + y - \ell(t_i)) \right] \right] \leq \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\phi(t_i, R) \right] \qquad (74)$$

Plugging Eqn (73) into the LHS of Eqn (74) we get

$$\begin{aligned}
&\mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\mathbf{E}_{y \sim Q(t_i, R)} \left[\phi(t_i + \Delta t_i, R + y - \ell(t_i)) \right] \right] && (75) \\
&= \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\phi(t_i, R) \right] && (76) \\
&+ \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\mathbf{E}_{y \sim Q(t_i, R)} \left[\left[\left\{ \frac{\partial}{\partial r} \left. \phi(\tau, \rho) \right|_{\substack{\tau,\, \rho\, = \\ t_i,\, R}} \right\} (y - \ell(t_i)) \right] \right] \right] && (77) \\
&+ \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\mathbf{E}_{y \sim Q(t_i, R)} \left[\left[\left\{ \frac{\partial^2}{\partial r^2} \left. \phi(\tau, \rho) \right|_{\substack{\tau,\, \rho\, = \\ t_i,\, R}} \right\} ((y - \ell(t_i))^2 - \Delta t_i) \right] \right] \right] && (78) \\
&+ O(s^3) && (79)
\end{aligned}$$

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 11 this implies that the expected value of higher derivatives of $\frac{\partial}{\partial R}\phi(R)$ are also finite.[2]

To prove inequality (52), we need to show that the terms 77 and 78 are smaller or equal to zero.

Term (77) is equal to zero:
As $\ell(t_i)$ is a constant relative to R and y, and $\left\{ \frac{\partial}{\partial r}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \right\}$ is a constant with respect to y we can rewrite (77) as

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

Combining the definitions of $\ell(t)$ (35) and and the learner's strategy P^{cc} (34) we get that

$$\ell(t_i) \quad = \quad \mathbf{E}_{R\sim\mathbf{\Psi}t_i}\left[\frac{1}{Z}\left\{ \frac{\partial}{\partial r}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \right\}\mathbf{E}_{y\sim Q(i,R)}\left[y\right]\right] \text{ where } Z = \mathbf{E}_{R\sim\mathbf{\Psi}t_i}\left[\frac{1}{Z}\left\{ \frac{\partial}{\partial r}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \right\}\right] \tag{81}$$

Plugging (81) into (80) and recalling the requirement that $\ell(t_i) < \infty$ we find that term (77) is equal to zero.

Term (78) 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 (36)

$$\mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\mathbf{E}_{y\sim Q(t_i,R)}\left[Q(t_i,R)\right]\left\{ \frac{\partial^2}{\partial r^2}\phi(\tau,\rho)\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \right\}(y-\ell(t_i))^2 - \Delta t_i\right] = \Delta t_i - \Delta t_i = 0 \tag{82}$$

Where $G(t_i, R)$ is defined in Equation (**??**) We find that (78) is zero.

Finally (79) is negligible relative to the other terms as $s \to 0$. □

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