An Optimal potential-based hedging algorithm

Yoav Freund

September 7, 2022

Abstract

We study a family of potential functions for online learning. We show that if the potential function has strictly positive derivatives of order 1-4 then the min-max optimal strategy for the adversary is Brownian motion. Using that fact we analyze different potential functions and show that the Normal-Hedge potential is optimal.

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 experts, or actions, while the learner controls a distribution over the actions.
Iteration $t = 1, \ldots, T$ of the game consists of the following steps:

1. The learner chooses a distribution P_j^t over the actions $j \in \{1, \ldots, N\}$.

2. The adversary chooses an *instantaneous loss* for each of the N actions:
   $l_j^t \in [-1, +1]$ for $j \in \{1, \ldots, N\}$.

3. The learner incurs an *instantanous expected loss* defined as $\ell^t = \sum_{j=1}^N P_j^t l_j^t$

Using standard notation we denote the *cumulative loss* of action j for times $1, \ldots, T$ by $L_j^T = \sum_{t=1}^T l_j^t$. Similarly, we denote the *cumulative loss* of the learner by $L_\ell^T = \sum_{t=1}^T \ell^T$. The *cumulative regret* of the learner with respect to action j is $R_j^T = \sum_{t=1}^T r_j^t = L_\ell^T - L_j^T$. Finally, for any $0 \le \epsilon \le 1$ we define the ϵ regret to R_ϵ^T to be value v such that the fraction of values of R_j^T that are smaller than v is at most ϵN.

The goal of the learner is to minimize the maximal regret at time T $R_\epsilon^T \doteq \max_j R_j^T = L_\ell^T - \min_j L_j^T$. The goal of the adversary is to maximize R_ϵ^T.

Our goal is to identify algorithms that improve over known bounds of the form $O(\sqrt{T \ln \frac{1}{\epsilon}})$. **Needs comparison to existing results, NormalHedge, Squint, square loss bounds**

We are interested in bounds the hold simultanously for all values of ϵ. Denote the distribution over regrets at some fixed iteration by μ. We say that the distribution μ satisfies the regret function B if

Definition 1 (Simultanous regret bound)**.** *Let $B : \mathbb{R} \to [0, 1]$ be a non-increasing function which maps regret bounds to probabilities.*

A distribution over regrets μ is simultanously bounded by B if

$$\forall r \in \mathbb{R} \quad \mathbf{P}_{R \sim \mu}[R \ge r] \le B(r)$$

Where $B : \mathbb{R} \to [0, 1]$ is a non-increasing function.

Our main tool for designing algorithms and proving bounds is potential functions.

We say that A distribution over regrets μ satisfies the potential function ϕ if

Definition 2 (Average potential bound). *A distribution over he reals μ satisfies the average potential function ϕ if*

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

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

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

Unlike uniform regret bounds, that are rather opaque, one can easily write a single step backward recursion for potentials. In other words, given a potential function at iteration t, one can construct potential functions for iteration $t - 1$. This povides a way for optimizing the potential function.

The original DTOL game does not seem to have min-max strategies. However, in an extended version of the game we can characterize min-max strategies. The extended game provides the adversary with additional possibilities, but not the learner. In other words, the standard game is more restrictive to the adversary, which means that upper bounds on the regret for the extended game also hold for the restricted game.

- min-max for known T

- Almost min/max for unknown T

- Simultanously for all ϵ

- Replacing T with V_T

- Interpretation of the limit game.

1.1 Some known Bounds

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

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

for some small constant C (typically smaller than 2). These bounds can be extended to infinite sets of experts by defining the 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

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

A few things are worth noting. First, as $|l_j^i| \le 1$, $\mathrm{Var}_j \le 1$ and therefor $V_T \le T$. 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.

1.2 Potential Functions

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

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

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

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

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

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

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

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

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

1.3 Variable time steps

The second order bounds of [7] show that, in some contexts, the time parameter t can be replaced by a variance-type parameter $V_T = \sum_{i=1}^{T} \mathrm{Var}_i$. This corresponds to a significant tightening of the bound. For example if instead of $l_j^i \in [-\epsilon, +\epsilon]$, then $V_T = T\epsilon^2$, and if we set $T = \epsilon^{-2}$ we get n upper bound on the regret that does not depend on T. The idea is to replace T with V_T in the definition of the upper and lower potentials. Which is what we do in Section4.

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

---

[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].

The rest of the paper is organized as follows In Section 2 we provide some basic definitions. In Section 3 we describe the potential game in the standard integer time case and prove some upper and lower bounds for this case. In Section 4 we define a the discrete time game, where time advances are real valued and finite. The core of our argument, which is based on the theory of "divided-differences" appears in Section 4.2. In Section 5 we define the continuous time limit of the game. In Section 6 we show that optimal any-time potential functions correspond to the solution of the kolmogorov Backwards differential equation, and show the Exponential weights and NormalHedge are solutions for that Equation. Finally, in Section 7 we argue that NormalHedge is the best possible potential function.

2   Preliminaries

The game takes place on the set $(t, R) \in [0, T] \times \mathbb{R}$, where t corresponds to time, T correspond to the time at the end of the game (i.e. this is a bounded horizon game where the horizon is known to both players. R corresponds to (cumulative) regret).

As in the DTOL setting, an strategy (or expert) correspond to a sequence of instantaneous losses. However, unlike the standard DTOL the number of strategies is not finite or known in advance. In DTOL the *state* of the learning process is defined by the by the regret of each of the N strategies: $\langle R_1(t), \ldots, R_N(t) \rangle$. In order to represent the regret of a potentially uncountable set of strategies we define the state as a distribution over possible regret values. Thus the *state* of the game at time t is a distribution (finite measure) over the real line, denoted $\mathbf{\Psi}(t)$. We denote by $R \sim \mathbf{\Psi}(t)$ a random regret R that is chosen according to the state at time t.

The transition from an N dimensional vector to a distribution is *central* to our analysis. The adversarial strategy we use splits each one of the strategies at time t into two: one half has instantaneous regret of $+1$ and the other instantaneous regret of -1. The adversary might not be able to use this strategy when the number of strategies is finite. If there is only one strategy with cumulative regret R then that strategy cannot, by definition, be split into two. Giving the adversary more possibilities allows us to narrow the gap between upper and lower bounds.

The initial state $\mathbf{\Psi}(0)$ is a point mass at $R = 0$. The state $\mathbf{\Psi}(t)$ is defined by $\mathbf{\Psi}(t_{i-1})$ and the choices made by the two players as described in the next section.

The score of the game is defined by $\mathbf{\Psi}(T)$ and a function called the *final potential function* $\phi(T, R)(R)$ that is fixed a-priori and is known to both players. The final score of the game is defined as

$$\Phi(T) = \mathbf{E}_{R \sim \mathbf{\Psi}(T)} \left[\phi(T, R)(R) \right] \tag{4}$$

The goal of the learner is to minimize $\Phi(T)$ and the goal of the adversary is to maximize it.

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

Definition 3 (Strict Positivity of degree k). *A function $f : \mathbb{R} \to \mathbb{R}$ 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.*

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

Lemma 2. *Let $\phi(T, R)(\cdot) \in \mathcal{P}^1$, $\Phi(T) \leq U$ and $\epsilon = \mathbf{P}_{R \sim \mathbf{\Psi}t} [R > R']$ be the probability of the set of actions with respect to which the regret is larger than R' Then*
then the regret of the algorithm relative to the top ϵ of the strategies is upper bounded by

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

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

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

3   Integer time game

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

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

The game is defined by three parameters:

- T : The number of iterations

- $\mathbf{\Psi}(0) = \delta(0)$ is the initial state of the game which is a point mass distribution at 0.

- $\phi(T, R)(R)$: The function that is in $\mathcal{P}^2$.

- $0 < c \leq 1$ - An upper bound on aggregate loss (loss of the master) in a single iteration. Note that the cumulative aggregate loss is at most cT. Note that $c = 1$ is always satisfied and eliminates the constraint.

The transition from $\mathbf{\Psi}(i)$ to $\mathbf{\Psi}(i+1)$ is defined by the choices made by the adversary and the learner.

1. The learner chooses weights. Formally, this is a distribution over $R \in \mathbb{R}$: $P(i)$.

2. The adversary chooses the losses of the strategies. Formally this is a mapping from $\mathbb{R}$ to distributions over $[-1, +1]$: $Q(i) : \mathbb{R} \to \Delta^{[-1, +1]}$. We use $l \sim Q(i, R)$ to denote the distribution over the instantaneous loss associated with iteration i and regret R.

3. The aggregate loss (also called "the loss of the master") is calculated:

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

   The adversary is constrained to Q such $|\ell(i)| \leq c$. We define the *bias* at (i, R) to be $B(i, R) \doteq \mathbf{E}_{l \sim Q(i, R)} [l]$ which allows us to rewrite Eqn (5) as

$$\ell(i) = \mathbf{E}_{R \sim \mathbf{\Psi}(i)} [P(i, R) B(i, R)] \tag{6}$$

   note that $B(i, R)$ is in $[-1, 1]$ and that $\ell(i)$ is the mean of $B(i, \cdot)$. Note also that $-1 - c \leq y - \ell(i) \leq 1 + c$ corresponds to the instantanous regret. In the integer game, setting $c \geq 1$ is equivalent to placing no restriction on $\ell(i)$.

4. The state is updated.
$$\mathbf{\Psi}(i+1) = \mathbf{E}_{R \sim \mathbf{\Psi}(i)} [R \oplus Q(i, R)] \ominus \ell(i) \tag{7}$$

   Where $Q(i, R)$ is the distribution of the losses of experts that are at location R after iteration $i-1$. $R \oplus Q(i, R)$ is the same distribution shiften right by the amount R. $\mathbf{E}_{R \sim \mathbf{\Psi}(i-1)} [\cdot]$ indicates the expectation over distribution, yealding a new distribution. Finally $\ominus \ell(i)$ is a shift left of the resulting distribtion according to the aggregate loss.

The final score is the mean of the potential according to the final state, as given in Equation (4)

$$\Phi(T) = \mathbf{E}_{R \sim \mathbf{\Psi}(T)} [\phi(T, R)(R)] \tag{8}$$

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

3.1 Results for integer time game

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

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

Theorem 3.

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

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

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

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

3.2 Analysis of the integer time game

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

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

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

The upper potential for learner strategy P is defined as

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

3.3 Strategies for integer time game

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

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

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

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

Lemma 4. *Suppose $\phi^{\uparrow}(i, R)$ is strictly convex with respect to R. Let $\mathcal{Q}(i-1, R, B)$ be the set of adversarial strategies $Q(i-1, R)$ that have bias $B = B(i-1, R) = \mathbf{E}_{y \sim Q(i-1, R)}[y]$ then the strategy in $\mathcal{Q}(i-1, R, B)$ that is best for the adversary is*

$$Q^p(i-1, R) = \begin{cases} +1 & w.p. \ \frac{1+B}{2} \\ -1 & w.p. \ \frac{1-B}{2} \end{cases} \tag{9}$$

6

$$\phi^\uparrow(i-1, R) = p\phi^\uparrow(i, R+1) + (1-p)\phi^\uparrow(i, R-1) \tag{10}$$

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

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

Lemma 5. *If $\phi(i, R)$ is CIC then*

1. *The adversarial strategy*

$$Q^{1/2}(i-1, R) = \begin{cases} -1 & w.p. \ 1/2 \\ +1 & w.p. \ 1/2 \end{cases} \tag{11}$$

   Guarantees the lower potential

$$\phi^\uparrow(i-1, R) = \frac{\phi(i, R+1) + \phi(i, R-1)}{2} \tag{12}$$

2. *The learner strategy:*

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

   Where Z is a normalization factor

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

   guarantees the upper potential

$$\phi^\downarrow(i-1, R) = \frac{\phi(i, R+1+c) + \phi(i, R-1-c)}{2} \tag{14}$$

Proof.   1. By symmetry adversarial strategy (11) guarantees that the aggregate loss (6) is zero regardless of the choice of the learner: $\ell(t) = 0$. Therefor the state update (7) is equivalent to the symmetric random walk:

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

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

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

   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 (6) we find that

$$\ell(i-1) = \frac{1}{Z_{i-1}} \mathbf{E}_{R \sim \boldsymbol{\Psi}(i-1)} \left[(\phi(i, R+1+c) - \phi(i, R-1-c)) B(i-1, R) \right] \tag{16}$$

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

$$\Phi_{P^*, Q}(i-1, R) = \mathbf{E}_{R \sim \boldsymbol{\Psi}(i-1)} \left[\mathbf{E}_{y \sim Q(i-1)(R)} \left[\phi(i, R+y - \ell(i-1)) \right] \right] \tag{17}$$

7

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

$$\phi(i, R + y) \leq \frac{\phi(i, R + 1 + c) + \phi(i, R - 1 - c)}{2} + (y - \ell(i))\frac{\phi(i, R + 1 + c) - \phi(i, R - 1 - c)}{2} \quad (18)$$

Combining the equations (16) and (17) we find that

$$
\begin{aligned}
\Phi_{P^*, Q}(i - 1, R) &= \mathbf{E}_{R \sim \boldsymbol{\Psi}(i-1)} \left[\mathbf{E}_{y \sim Q(i-1)(R)} \left[\phi(i, R + y - \ell(i-1)) \right] \right] &(19) \\
&\leq \mathbf{E}_{R \sim \boldsymbol{\Psi}(i-1)} \left[\frac{\phi(i, R + 1 + c) + \phi(i, R - 1 - c)}{2} \right] &(20) \\
&+ \mathbf{E}_{R \sim \boldsymbol{\Psi}(i-1)} \left[\mathbf{E}_{y \sim Q(i-1)(R)} \left[(y - \ell(i-1))\frac{\phi(i, R + 1 + c) - \phi(i, R - 1 - c)}{2} \right] \right] &(21)
\end{aligned}
$$

The final step is to show that the term (21) is equal to zero. As $\ell(i - 1)$ is a constant with respect to R and y the term (21) can be written as:

$$\mathbf{E}_{R \sim \boldsymbol{\Psi}(i-1)} \left[\mathbf{E}_{y \sim Q(i-1)(R)} \left[(y - \ell(i-1))\frac{\phi(i+1, R+1) - \phi(i+1, R-1)}{2} \right] \right] \quad (22)$$

$$= \mathbf{E}_{R \sim \boldsymbol{\Psi}(i-1)} \left[B(i - 1, R)\frac{\phi(i, R+1+c) - \phi(i, R-1-c)}{2} \right] \quad (23)$$

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

$$= 0 \quad (25)$$

$\square$

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

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

3.4 A learner strategy with a variance-dependent bound

As shown in Lemma 4, 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 6. *The learner strategy:*

$$P^2(i - 1, R) = \frac{1}{Z} \left. \frac{\partial}{\partial r} \right|_{r=R} \phi(i, r) \quad (26)$$

Where Z is a normalization factor

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

guarantees the following upper potential against any adversarial strategy Q

$$\phi^{\downarrow}(i - 1, R) = \phi(i, R) + b(i, R)\mathbf{E}_{l \sim Q(i, R)} \left[l^2 \right] \quad (27)$$

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

$R-1-c$ $R-1$ R $R+1$ $R+1+c$
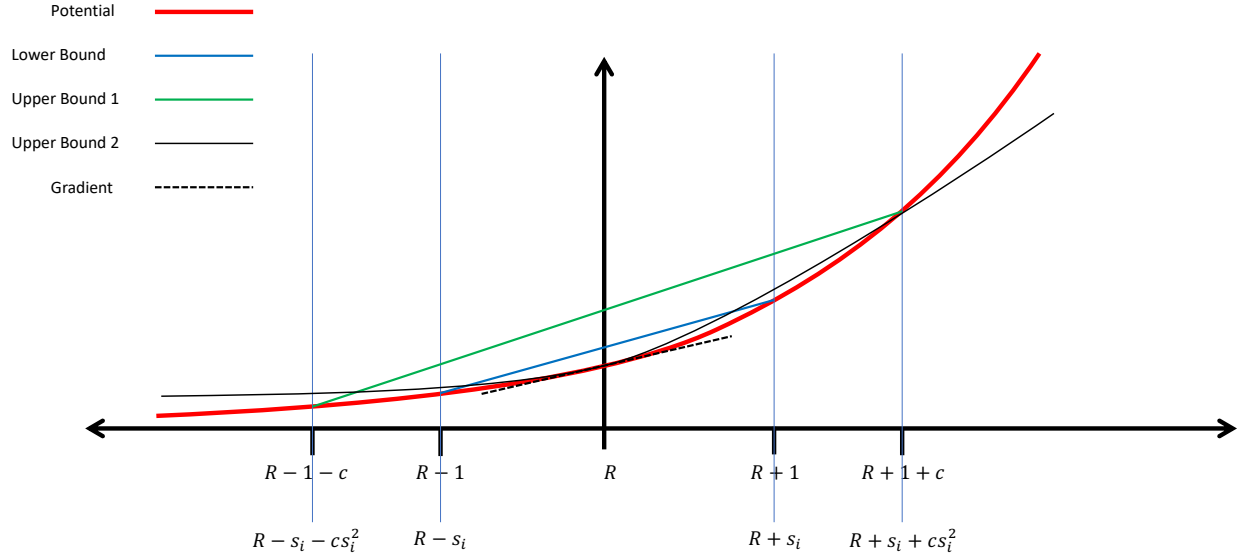
$R-s_i-cs_i^2$ $R-s_i$ $R+s_i$ $R+s_i+cs_i^2$

Figure 1: This figure depicts the relationship between the different upper and lower bounds used in the analysis. To aid understanding we describe the elements of the figure twice: once for the integer game, and once for the continuous time game.

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

Continuous time game: let the current iteration be i, the current time be t_i and the current regret be R. Let $0 < s_i \leq 1$ be the step size chosen by the adversary, so that the next time is $t_{i+1} = t_i + s_i^2$. Let r be the regret at iteration $i+1$, we have that $R - s_i - cs_i^2 \leq r \leq R + s_i + cs_i^2$. The potential at iteration $i+1$ is $\phi(t_{i+1}, r)$ (the red curve). The lower bound (blue line) corresponds to the adversarial strategy: $Q^{1/2}_{\pm s_i}(t_i, R)$. 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.

We compare the bound for P^2 to the bound for P^1 given in Lemma 5. 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.

4   Discrete time game

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: $\Psi(i-1) = \Psi(i)$, $\Delta t_i = 0$ 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. 11) 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 $Q^{1/2}_{\pm 1}$. The other strategy, denoted $Q^{1/2}_{\pm 1/k}$, generates losses of $\pm 1/k$ with equal probabilities.

From the adversarial point of view $Q^{1/2}_{\pm 1/k}$ is worse than $Q^{1/2}_{\pm 1}$. 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 $Q^{1/2}_{\pm 1}$ 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 $Q^{1/2}_{\pm 1/k}$. Equivalently, we associate j iterations of $Q^{1/2}_{\pm 1/k}$ 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 $Q^{1/2}_{\pm 1/k}$ is zero for any k. As for the variances, a single iteration of $Q^{1/2}_{\pm 1}$ is 1 and a single iteration of $Q^{1/2}_{\pm 1/k}$ is $1/k^2$. It follows that the variance after j iterations of $Q^{1/2}_{\pm 1/k}$ j/k^2. Equating this variance with that of a single step of $Q^{1/2}_{\pm 1}$ 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 $Q^{1/2+\gamma}_{\pm 1}$ as defined in Eqn (9) 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 $Q^{1/2+\gamma'}_{\pm 1/k}$ has the same mean as a single iteration of $Q^{1/2+\gamma}_{\pm 1}$. The mean of a single iteration of $Q^{1/2+\gamma}_{\pm 1}$ is 2γ while the mean of a single iteration of $Q^{1/2+\gamma'}_{\pm 1/k}$ 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.

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:

$$\ell(t_i) = \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[P(t_i, R) B(t_i, R) \right], \quad \text{where } B(t_i, R) \doteq \mathbf{E}_{y \sim Q(t_i, R)} [y] \tag{28}$$

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

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:

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

4.1 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 7.

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

- *For any $\epsilon > 0$ there exists a strategy for the adversary such that for any strategy of the learner $\Phi(T) \geq A - \epsilon$*

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

4.2 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 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 $Q_{\pm 2^{-k}}^{1/2}$ for $T2^{2k}$ iterations. This corresponds to the distribution created by a random walk with $T2^{2k}$ time steps, each step equal to $+2^{-k}$ 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. The backwards recursion that defines the value function is the celebrated Backwrds Kolmogorov Equation with zero dift and unit variance

$$\frac{\partial}{\partial t}\phi(t, R) + \frac{1}{2}\frac{\partial^2}{\partial R^2}\phi(t, R) = 0 \tag{29}$$

Given a final value function with a strictly positive fourth derivative we can use Equation (29) to compute the value function for all $0 \le t \le T$. We will do so in he next 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 $\phi(t', R)$ 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

$$\phi_1(t, R) - \phi_0(t, R) = \frac{1}{16}(\phi(t', R+2) - 4\phi(t', R+1) + 6\phi(t', R) - 4\phi(t', R-1) + \phi(t', R-2)) > 0 \tag{30}$$

Inequalities of this form have been studied extensively under the name "divided differences" [16, 3, 9]. A function $\phi(T, R)$ that satisfies inequality 30 is said to be *4'th order convex* (see details in in [3]).

n-convex functions have a very simple characterization:

Theorem 10. *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.

$\square$

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.

$\square$

4.3 Strategies for the Learner in the discrete time game

The strategies we propose for the learner in the continuous time game are an adaptation of the strategies P^1, P^2 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

$$P^{1d}(t_{i-1}, R) = \frac{1}{Z^{1d}} \frac{\phi(t_i, R + s_{i-1} + cs_{i-1}^2) - \phi(t_i, R - s_{i-1} - cs_{i-1}^2)}{2} \tag{31}$$

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

Next, we consider the discrete time version of P^2: (Eqn 26)

$$P^{2d}(t_{i-1}, R) = \frac{1}{Z^{2d}} \left. \frac{\partial}{\partial r} \right|_{r=R} \phi(t_{i-1} + s_{i-1}^2, r) \tag{32}$$

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

5 Continuous time game

In Section 4 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 already have a worst case bound on this problem, which corresponds to the adversary performing Brownian motion. Our goal here is to refine this bound for the case where the adversary's strategies are not worst case.

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 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 $t_{i+1} = t_i + s_i^2$. The analysis below shows that these are two extremes on a spectrum and that intermediate cases can be characterized using a variance-like quantity.

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

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

We let $s \to 0$ keeping $\sum_i s_i^2 = T$. In this limit the strategy P^{2c} (Equation (32)) simplifies to

$$P^{cc}(t, R) = \frac{1}{Z^{cc}} \left. \frac{\partial}{\partial r} \right|_{r=R} \phi(t, r) \text{ where } Z^{cc} = \mathbf{E}_{R \sim \boldsymbol{\Psi}(t_i)} \left[\left. \frac{\partial}{\partial r} \right|_{r=R} \phi(t, r) \right] \tag{33}$$

Theorem 11. *Let $\phi \in \mathcal{P}^\infty$ be a potential function that satisfies the Kolmogorov backward equation (29). Let T be fixed and let G be an (n, s, T)-bounded instance of the discrete time game using ϕ. Let the time increment associated with iteration i: $\Delta t_i = t_{i+1} - t_i$ be set so that*

$$\Delta t_i = \mathbf{E}_{R \sim \boldsymbol{\Psi}(t_i)} \left[\mathbf{E}_{y \sim Q(t_i, R)} \left[G(t_i, R)(y - \ell(t_i))^2 \right] \right] \tag{34}$$

Where

$$G(t, R) = \frac{1}{Z^G} \left. \frac{\partial^2}{\partial \rho^2} \right|_{\rho=R} \phi(t, \rho), \ \ Z^G = \mathbf{E}_{R \sim \boldsymbol{\Psi}(t_i)} \left[\left. \frac{\partial^2}{\partial \rho^2} \right|_{\rho=R} \phi(t, \rho) \right]$$

then

$$\Phi(\boldsymbol{\Psi}(0)) \leq \Phi(\boldsymbol{\Psi}(T)) + O(sT)$$

The proof of the theorem is given in appendix A

Theorem 11 shows that the average potential does not stay constant, instead, it increases by at most $O(s^3)$ at each iteration. As T increases by at most s^2 at each iteration we find that the total increasein the potential is at most Ts.

6   Two self-consistent potential functions

The potential functions, $\phi(t, R)$ is a solution of PDE (29):

$$\frac{\partial}{\partial t} \phi(t, R) + \frac{1}{2} \frac{\partial^2}{\partial r^2} \phi(t, R) = 0 \tag{35}$$

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 29, and derive the regret bound corresponding to it.

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

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

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 $t > 0$

$$R_\epsilon \leq \sqrt{2t \ln \frac{1}{\epsilon}} \tag{36}$$

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

$$\phi_{\mathrm{NH}}(R, t) = \begin{cases} \frac{1}{\sqrt{t+\nu}} \exp\left(\frac{R^2}{2(t+\nu)}\right) & \text{if } R \geq 0 \\ \frac{1}{\sqrt{t+\nu}} & \text{if } R < 0 \end{cases} \tag{37}$$

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

The regret bound we get is:

$$R_\epsilon \leq \sqrt{(t + \nu)\left(\ln(t + \nu) + 2 \ln \frac{1}{\epsilon}\right)} \tag{38}$$

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.

7 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. 37). 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 11

We start with two technical lemmas

Lemma 12. *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 $\mathbf{E}_{x \sim \mathbf{\Psi}}[h(x)f'(x)]$ is undefined. Define $h^+(x) = \max(0, h(x))$. As $f'(x) > 0$, this implies that either $\mathbf{E}_{x \sim \mathbf{\Psi}}[h^+(x)f'(x)] = \infty$ or $\mathbf{E}_{x \sim \mathbf{\Psi}}[(-h)^+(x)f'(x)] = \infty$ (or both).

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

Lemma 13. *Let $f(x, y)$ be a differentiable function with continuous derivatives up to degree three. Then*

$$f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0) + \left\{ \frac{\partial}{\partial x} f(x, y) \Big|_{\substack{x, y = \\ x_0, y_0}} \right\} \Delta x + \left\{ \frac{\partial}{\partial y} f(x, y) \Big|_{\substack{x, y = \\ x_0, y_0}} \right\} \Delta y \tag{39}$$

$$+ \; \frac{1}{2} \left\{ \frac{\partial^2}{\partial x^2} f(x, y) \Big|_{\substack{x, y = \\ x_0, y_0}} \right\} \Delta x^2 + \left\{ \frac{\partial^2}{\partial x \partial y} f(x, y) \Big|_{\substack{x, y = \\ x_0, y_0}} \right\} \Delta x \Delta y + \frac{1}{2} \left\{ \frac{\partial^2}{\partial y^2} f(x, y) \Big|_{\substack{x, y = \\ x_0, y_0}} \right\} \Delta y^2 \tag{40}$$

$$+ \; \frac{1}{6} \left\{ \frac{\partial^3}{\partial x^3} f(x, y) \Big|_{\substack{x, y = \\ x_0 + t\Delta x, \, y_0 + t\Delta y}} \right\} \Delta x^3 + \frac{1}{2} \left\{ \frac{\partial^3}{\partial x^2 \partial y} f(x, y) \Big|_{\substack{x, y = \\ x_0 + t\Delta x, \, y_0 + t\Delta y}} \right\} \Delta x^2 \Delta y \tag{41}$$

$$+ \frac{1}{2} \left\{ \frac{\partial^3}{\partial x \partial y^2} f(x, y) \Big|_{\substack{x, y = \\ x_0 + t\Delta x, \, y_0 + t\Delta y}} \right\} \Delta x \Delta y^2 + \frac{1}{6} \left\{ \frac{\partial^3}{\partial y^3} f(x, y) \Big|_{\substack{x, y = \\ x_0 + t\Delta x, \, y_0 + t\Delta y}} \right\} \Delta y^3 \tag{42}$$

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

Proof. of Lemma 13 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:

$$\frac{d^n}{dt^n} F(t) = \sum_{m=1}^{n} \binom{n}{m} \frac{\partial^n}{\partial x^m \partial y^{n-m}} f(x_0 + t\Delta x, y_0 + t\Delta y) \Delta x^m \Delta y^{n-m} \tag{43}$$

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

$$f(x_0 + \Delta x, y_0 + \Delta y) = F(1) = F(0) + \frac{d}{dt} F(0) + \frac{1}{2} \frac{d^2}{dt^2} F(0) + \frac{1}{6} \frac{d^3}{dt^3} F(t') \tag{44}$$

Where $0 \leq t' \leq 1$. Using Eqn (43) to expand each term in Eqn. (44) completes the proof. $\qquad\square$

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

$$\Phi(\mathbf{\Psi}(t_{i+1})) \leq \Phi(\mathbf{\Psi}(t_i)) + a s_i^3 \text{ for some constant } a > 0 \tag{45}$$

Summing inequality (45) over all iterations we get that

$$\Phi(\mathbf{\Psi}(T)) \le \Phi(\mathbf{\Psi}(0)) + c\sum_{i=1}^{n} s_i^3 \le \Phi(\mathbf{\Psi}(0)) + as\sum_{i=1}^{n} s_i^2 = \Phi(\mathbf{\Psi}(0)) + asT \tag{46}$$

From which the statement of the theorem follows.

We now prove inequality (45). We use the notation $r = y - \ell(i)$ to denote the instantaneous regret at iteration i.

Applying Lemma 13 to $\phi(t_{i+1}, R_{i+1}) = \phi(t_i + \Delta t_i, R_i + r_i)$ we get

$$\phi(t_i + \Delta t_i, R_i + r_i) \quad = \quad \phi(t_i, R_i) \tag{47}$$

$$+ \quad \left\{ \left. \frac{\partial}{\partial \rho}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} r_i \tag{48}$$

$$+ \quad \left\{ \left. \frac{\partial}{\partial \tau}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} \Delta t_i \tag{49}$$

$$+ \quad \frac{1}{2}\left\{ \left. \frac{\partial^2}{\partial \rho^2}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} r_i^2 \tag{50}$$

$$+ \quad \left\{ \left. \frac{\partial^2}{\partial r \partial \tau}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} r_i \Delta t_i \tag{51}$$

$$+ \quad \frac{1}{2}\left\{ \left. \frac{\partial^2}{\partial \tau^2}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} \Delta t_i^2 \tag{52}$$

$$+ \quad \frac{1}{6}\left\{ \left. \frac{\partial^3}{\partial \rho^3}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i + g\Delta t_i, R_i + gr_i}} \right\} r_i^3 \tag{53}$$

$$+ \quad \frac{1}{2}\left\{ \left. \frac{\partial^3}{\partial \rho^2 \partial \tau}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i + g\Delta t_i, R_i + gr_i}} \right\} r_i^2 \Delta t_i \tag{54}$$

$$+ \quad \frac{1}{2}\left\{ \left. \frac{\partial^3}{\partial \rho \partial \tau^2}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i + g\Delta t_i, R_i + gr_i}} \right\} r_i \Delta t_i^2 \tag{55}$$

$$+ \quad \frac{1}{6}\left\{ \left. \frac{\partial^3}{\partial \tau^3}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i + g\Delta t_i, R_i + gr_i}} \right\} \Delta t_i^3 \tag{56}$$

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

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.

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

18

Which yields

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

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

$$
+ \quad \left\{ \frac{\partial^2}{\partial \rho^2} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} \left(\frac{r_i^2}{2} - \Delta t_i \right) \tag{59}
$$

$$
- \quad \left\{ \frac{\partial^3}{\partial \rho^3} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} r_i \Delta t_i \tag{60}
$$

$$
+ \quad \frac{1}{2} \left\{ \frac{\partial^4}{\partial \rho^4} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} \Delta t_i^2 \tag{61}
$$

$$
+ \quad \frac{1}{6} \left\{ \frac{\partial^3}{\partial \rho^3} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i + g\Delta t_i, R_i + g r_i}} \right\} r_i^3 \tag{62}
$$

$$
- \quad \frac{1}{2} \left\{ \frac{\partial^4}{\partial \rho^4} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i + g\Delta t_i, R_i + g r_i}} \right\} r_i^2 \Delta t_i \tag{63}
$$

$$
+ \quad \frac{1}{2} \left\{ \frac{\partial^5}{\partial \rho^5} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i + g\Delta t_i, R_i + g r_i}} \right\} r_i \Delta t_i^2 \tag{64}
$$

$$
- \quad \frac{1}{6} \left\{ \frac{\partial^6}{\partial \rho^6} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i + g\Delta t_i, R_i + g r_i}} \right\} \Delta t_i^3 \tag{65}
$$

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

1. $|r_i| \leq s_i + c s_i^2 \leq 2 s_i$

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 (58):** $r_i = 2 s_i \frac{r_i}{2 s_i} = 2 s_i h_1(r_i)$.

- **For (59):** $r_i^2 - \frac{1}{2} \Delta t_i = 4 s_i^2 \frac{r_i^2 - \frac{1}{2} \Delta t_i}{4 s_i^2} = 4 s_i^2 h_2(r_i, \Delta t_i)$

- **For (60):** $r_i \Delta t_i = 2 s_i^3 \frac{r_i \Delta t_i}{2 s_i^3} = 2 s_i^3 h_3(r_i, \Delta t_i)$

- **For (61):** $\Delta t_i^2 = s_i^4 \frac{\Delta t_i^2}{s_i^4} = s_i^3 h_4(\Delta t_i)$

- **For (62):** $r_i^3 = 8 s_i^3 \frac{r_i^3}{8 s_i^3} = 8 s_i^3 h_5(r_i, \Delta t_i)$

- **For (63):** $r_i^2 \Delta t_i = 4 s_i^4 \frac{r_i^2 \Delta t_i}{4 s_i^4} = 4 s_i^3 h_6(r_i, \Delta t_i)$

- **For (64):** $r_i \Delta t_i^2 = 2 s_i^5 \frac{r_i \Delta t_i^2}{2 s_i^5}$

- **For (65):** $\Delta t_i^3 = s_i^6 \frac{\Delta t_i^3}{s_i^6}$

We therefor get the simplified equation

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

and therefor

$$
\begin{aligned}
\phi(t_i + \Delta t_i, R + r) \;=\;& \phi(t_i, R) + \left\{ \left. \frac{\partial}{\partial r}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} r \\
+\;& \left\{ \left. \frac{\partial^2}{\partial r^2}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} (r^2 - \Delta t_i) + O(s^3)
\end{aligned}
\tag{66}
$$

Our next step is to consider the expected value of (66) 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:

$$
\mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\mathbf{E}_{y \sim Q(t_i, R)} \left[\phi(t_i + \Delta t_i, R + y - \ell(t_i)) \right] \right] \leq \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\phi(t_i, R) \right]
\tag{67}
$$

Plugging Eqn (66) into the LHS of Eqn (67) we get

$$
\mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\mathbf{E}_{y \sim Q(t_i, R)} \left[\phi(t_i + \Delta t_i, R + y - \ell(t_i)) \right] \right]
\tag{68}
$$
$$
= \;\; \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\phi(t_i, R) \right]
\tag{69}
$$
$$
+ \;\; \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\mathbf{E}_{y \sim Q(t_i, R)} \left[\left\{ \left. \frac{\partial}{\partial r}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} (y - \ell(t_i)) \right] \right]
\tag{70}
$$
$$
+ \;\; \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\mathbf{E}_{y \sim Q(t_i, R)} \left[\left\{ \left. \frac{\partial^2}{\partial r^2}\phi(\tau, \rho) \right|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} ((y - \ell(t_i))^2 - \Delta t_i) \right] \right]
\tag{71}
$$
$$
+ \;\; O(s^3)
\tag{72}
$$

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

To prove inequality (45), we need to show that the terms 70 and 71 are smaller or equal to zero.

---

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

Term (70) is equal to zero:

As $\ell(t_i)$ is a constant relative to R and y, and $\left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \big|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\}$ is a constant with respect to y we can rewrite (70) as

$$\mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} \mathbf{E}_{y \sim Q(t_i, R)} [y] \right] - \ell(t_i) \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} \right] \tag{73}$$

Combining the definitions of $\ell(t)$ (28) and and the learner's strategy P^{cc} (33) we get that

$$\ell(t_i) \;=\; \mathbf{E}_{R \sim \mathbf{\Psi} t_i} \left[\frac{1}{Z} \left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} \mathbf{E}_{y \sim Q(i, R)} [y] \right] \text{ where } Z = \mathbf{E}_{R \sim \mathbf{\Psi} t_i} \left[\frac{1}{Z} \left\{ \frac{\partial}{\partial r} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} \right] \tag{74}$$

Plugging (74) into (73) and recalling the requirement that $\ell(t_i) < \infty$ we find that term (70) is equal to zero.

Term (71) is equal to zero:

As Δt_i is a constant relative to y, we can take it outside the expectaation to give

$$\mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\mathbf{E}_{y \sim Q(t_i, R)} \left[\left\{ \frac{\partial^2}{\partial r^2} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} (y - \ell(t_i))^2 \right] \right] - \Delta t_i \, \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\left\{ \frac{\partial^2}{\partial r^2} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} \right] \tag{75}$$

On the other hand, from the choice of Δt_i given in Eqn (34) we get

$$\Delta t_i \, \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\left\{ \frac{\partial^2}{\partial r^2} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} \right] = \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[\mathbf{E}_{y \sim Q(t_i, R)} \left[\left\{ \frac{\partial^2}{\partial r^2} \phi(\tau, \rho) \Big|_{\substack{\tau, \rho = \\ t_i, R_i}} \right\} (y - \ell(t_i))^2 \right] \right] \tag{76}$$

Combining () and (76) we find that (71) is zero.

Finally (72) is negligible relative to the other terms as $s \to 0$. $\qquad \square$