Optimal online Learning using potential functions

Yoav Freund

January 30, 2023

Abstract

We study regret-minimizing online algorithms based on potential functions. First, we show that any algorithm with a regret bound that holds for any ϵ is equivalent to a potential minimizing algorithm and vice versa. Second we should a min-max learning algorithm for known horizon. We show a regret bound that is close to optimal when the horizon is not known. Finally we give an algorithm with second order bounds that characterize easy sequences.

1 Introduction

In this paper we study a popular variant of online learning called *the decision-theoretic online learning game* (DTOL) [12]. DTOL (Figure 1) is a repeated zero sum game between a *learner* and an *adversary*. The adversary controls the losses of N actions, while the learner controls a distribution over the actions.

For $i = 1, \ldots, T$

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

Figure 1: Decision theoretic online learning

[12] presents the DTOL framework and applies multiplicative weights algorithm [] to give *zero-order* regret bounds of the form

$$\max_{j=1}^{N} R(t, j) \prec \sqrt{t \ln N} \tag{1}$$

Many generalizations and refinements have been made over the years. We describe some of them below

- **percentile-based bounds** [14] refines this bound by replacing the comparison to the best single action to a comparison to the regret to the top ϵ-percentile of the actions. And give bounds of the form

$$R(t, \epsilon) \prec \sqrt{t \ln \frac{1}{\epsilon}} \tag{2}$$

Using quantiles allows actions sets that are uncountably infinite. For example consider an action set that corresponds to linear functions with real valued parameters.

- **Parameter-free algorithms** The multiplicative weights algorithm has a learning rate parameter $\eta > 0$. The bounds above require a-priori knowledge of the length of the game, the number of actions, or te percentile ϵ in order to tune the learning rate. "Parameter free" algorithms ...... remove the need for a-priori knowledge of the sequence.

- **Second order bounds** take advantage of so-called "easy sequences" where $|l_j^i| < 1$ Roughly speaking, second order bounds replace the length of the sequence t with a a sum of squares that is smaller than t for easy sequences. Several definitions for the sum of squares have been analyzed. Cesa-Binachi et al. [7] propose two such quantities. The *Quadratic Variation* $Q_j^t = \sum_{i=1}^t (l_j^i)^2$ measures the cumulative variation of the action j at. While the *cumulative variance* $V^t = \sum_{i=1}^t \sum_{j=1}^N P_j^i (l_j^i - \ell^i)^2$ is a measure of the variance of all of the actions.

  A large number variants of the quadratic variation bounds have been proposed [9, 16, 15, 18]

  We know of no prior work other than [7] and the work presented here that is based on cumulative variance.

- **Potential based Online learning algorithms** There are two main paradigms for designing online learning algorithms: the *follow the leader* paradigm and the *potential function* paradigm. Our work here is in the potential function paradigm. A potential function is a positive increasing function $\phi : \mathbb{R} \to \mathbb{R}$. The main quantity that is analyzed is the score, the average potential the actions, $\Phi_j = \frac{1}{N}\sum_{i=1}^N \phi(R_j^i)$. The regret buonds are proven by combining an upper and lower bound on Φ_t.

Using these definitions, we describe our main results

1. **Upper bound** We present a parameter-free potential-based algorithm such that for any t, ϵ the following regret bound holds.

$$R(t,\epsilon) \leq \sqrt{(V_t + 1)\left(\ln(V_t+1) + 2\ln\frac{1}{\epsilon}\right)} \text{ where } V_t = \frac{1}{N}\sum_{i=1}^t \sum_{j=1}^N H_j^i (l_j^i - \ell^i)^2 \tag{3}$$

   Where for each iteration i, $(H_1^i, \ldots, H_N^i)$ is a distribution over the actions $1, \ldots, N$ (Different from P_i^j)

2. **Lower bound** For the sake of comparing with the lower bound we express the upper bound using ϵ as a function of R

$$\forall t, R, \;\; \epsilon(t,R) \leq \sqrt{V_t + 1}\exp\left(\frac{-R^2}{2(V_t+1)}\right) \tag{4}$$

   Let $S_t = \sum_{i=1}^t \sigma_i^2$ for $0 \geq \sigma_i \leq 1$ be a sequence such that $\lim_{t\to\infty} S_t = \infty$. Then there is an adversarial strategy such that $V_t = S_t$ and for any learner strategy (whether or no it is potential based):

$$\forall t, R, \;\; \epsilon(t,R) \geq \left(\frac{\sqrt{V_t}}{R} - \left(\frac{\sqrt{V_t}}{R}\right)^3\right)\exp\left(\frac{-R^2}{2V_t}\right) \tag{5}$$

3. **Characterization of the min/max solutions of the potential game**

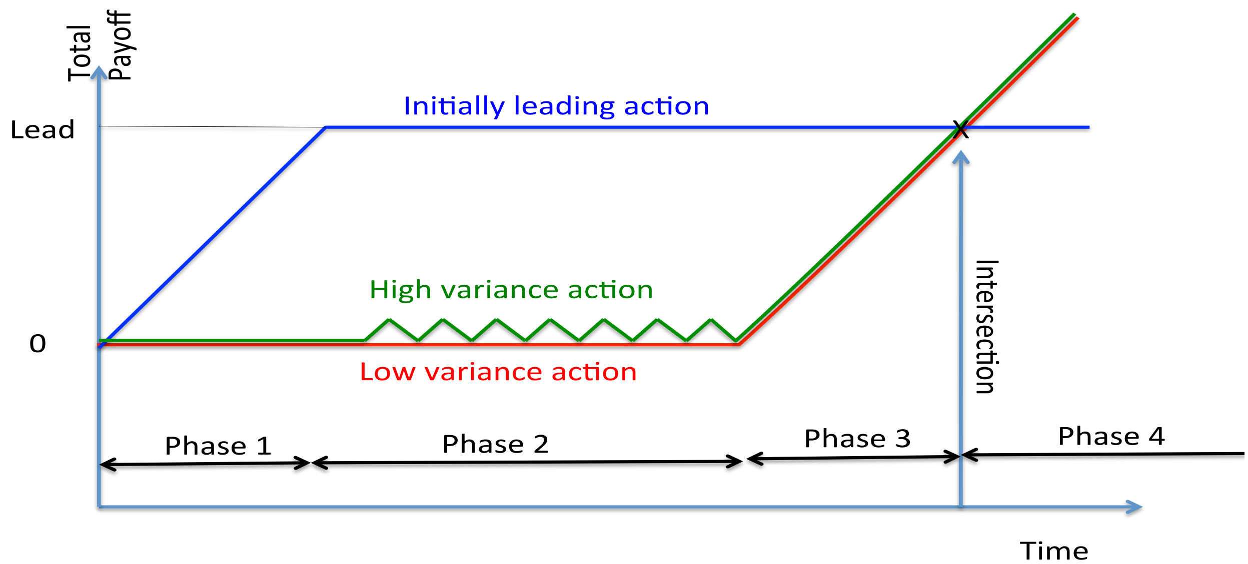
   We formalize the notion of potential based learning by defining an online game, similar to DTOL, where the goal of the learner is to minimize the average potential. We derive the min-max solution of this game for any potential function that has four strictly positive derivatives.

2 The concept of state and a critique of quadratic variation

Central to our analysis is the concept of the *state* of the DTOL game described in figure 1. A state is a function of history that is minimally sufficient to compute the future of the game. Clearly, a sufficient state of the game at time t is the vector of N cumulative regrets: $(R_1^t, \ldots, R_N^t)$. As the indexing of the actions is immaterial we can replace the vector with a distribution of regrets at time t which we denote $\mathbf{\Psi}(t)$.

This definition of state is sufficient in two senses. First, the regret bounds at time t depend only on $\mathbf{\Psi}(t)$. Second, given $\mathbf{\Psi}(t)$ and the actions of the learner and the adversary at time t uniquely defines $\mathbf{\Psi}(t)$. Other information about individual actions, such as the quadratic variation of individual actions, do not effect the progression of the states.

To sharpen this argument consider an example depicted in Figure 2 [1] We compare two loss histories, each involving two actions. The first consists of the the blue and the red lines, the second consists of the blue and the green line. In both cases the total loss of both action at the end of phase 3 are equal. In other words, the state at the end of phase 3 is identical for the two examples. An algorithm that depends on quadratic variation would give different weights to the high and the low variation action even though the states are identical. [2]



To analyze these potential functions we define a different game, which we call the "potential game". In this game the primary goal of the learner is not to minimize regret, rather, it is to minimize the final score Φ^T. To do so we define potential functions for intermediate steps: $0 \le t < T$.[3]

3 Upper bound

4 Lower bound

We prove the lower bound using Lyapunov CLT [3]

[1] This example appeared in an open problem in COLT2016

[2] One can argue for a different game in which the state includes the quadratic variation. However, we can't see a natural Way to define a game that requires the expanded state.

[3] 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 [21, 11].

Set $V_0 = 1$
For $t = 0, 1, 2, \ldots$

1. The learner chooses a distribution

$$P^{NH2}(t,j) = \frac{1}{Z^t}[R_j^t]^{+2}\exp\left(\frac{[R_j^t]^{+2}}{2V_t}\right) \text{ where } Z^t = \sum_{j=1}^{N}[R_j^t]^{+2}\exp\left(\frac{[R_j^t]^{+2}}{2V_t}\right) \tag{6}$$

2. The aggregate loss is calculated:

$$\ell^t = \sum_{j=1}^{N} P^{NH2}(t,j) l_j^t, \tag{7}$$

3. Increment $V_{t+1} = V_t + Var_t$ where

$$Var_t = \sum_{j=1}^{N} H(V_t + 1, R_j^t + 1)(l_j^t - \ell_t)^2 \tag{8}$$

$$W(V,R) = \exp\left(\frac{[R]^{+2}}{2V}\right)\left(\frac{1}{2V^{3/2}} + \frac{[R]^{+2}}{V^{5/2}}\right), \quad H(V,R) = \frac{W(V,R)}{\sum_{j=1}^{N} W(V,R_j)} \tag{9}$$

4. Update regrets $\forall j = 1, \ldots N, \quad R_j^{t+1} = R_j^t + l_j^t - \ell^t$

Figure 2: NormalHedge.2

5 Preliminaries

We define some terms and notationthat will be used in the rest of the paper.

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

Definition 1 (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.*

The following useful lemma states that $\mathcal{P}^k$ is closed under positive conbinations.

Lemma 1. *Suppose that for $i = 1, \ldots, n$, $f_i \in \mathcal{P}^k$ and $\alpha_i > 0$, Then $\sum_{i=1}^{k} \alpha_i f_i \in \mathcal{P}^k$*

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

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

$$S_\omega = (R_\omega^1, R_\omega^2, \ldots, R_\omega^N) \tag{10}$$

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

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

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

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

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

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

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

6 Integer time game

The integer time game is described in Figure 3. The integer time game generalizes the decision theoretic online learning problem [13] in the following ways:

1. The goal of the learner in DTOL is to guarantee an upper bounds on the regret. The learner's goal in the integer time game is to minimize the final score. From theorem **??** we know that if we set the final potential as $\boldsymbol{\phi}_T(R) = \frac{1}{G(R)}$ then the two conditions are equivalent, allowing us to focus on the score.
2. The number of iterations T is given as input, as is the potential function at the end: $\boldsymbol{\phi}_T(R)$.
3. The actions are assumed to be *divisible*. For our purposes it is enough to assume that any action can be split into two equal weight parts.

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

6.1 Defining potential Functions for all iterations

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

A single action defines a path S_ω (as defined in (10)). Fixing the strategies of the learner and the adversary determines a distribution $\mathcal{D}$ over paths. We describe two equivalent ways to define $\phi_{P,Q}(i, R)$ for $i < T$

1. **Using conditional expectation** We can define the potential on iteration i based on the fixed potential at iteration T.
$$\forall i = 1, \ldots, T, \; \forall R \quad \phi_{P,Q}(i, R) = \mathbf{E}_{\omega \sim \mathcal{D} | R^i_\omega = R} \left[\phi(T, R^T_\omega) \right] \tag{14}$$
2. **Using backward induction** It is sometimes convenient to compute the the potential for time i from the potential at time $i + 1$:
$$\forall i = 1, \ldots, T-1, R \quad \phi_{P,Q}(i, R) = \mathbf{E}_{\omega \sim \mathcal{D} | M^i_\omega = R} \left[\phi_{P,Q}(i+1, R^{i+1}_\omega) \right] \tag{15}$$
by using backwards induction: $i = T-1, T-2, \ldots, 1$ we can compute the potential for all iterations.

   We use Equations (11,12) and marginalizing over R to express Equation (15) in terms of the single step strategies:
$$\forall i = 1, \ldots, T-1, R \quad \phi_{P,Q}(i, R) \doteq \mathbf{E}_{r \sim [(R - \ell(i)) \oplus Q(i,R)]} [\phi_{P,Q}(i+1, r)] \tag{16}$$

Initialization:

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

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

1. The learner chooses a non-negative random variable over Ω that is the *weight function* $P(i, R)$ such that $\mathbf{\Psi}(i) \odot P(i) = 1$
2. The adversary chooses a function $Q(i, R)$ that maps i, R to a distribution over $[-1, +1]$. This random variable corresponds to the instantaneous loss of each action at time t.
3. We define the *bias* at (i, R) to be
$$B(i, R) \doteq \mathbf{E}_{l \sim Q(i,R)}[l] \tag{11}$$
4. the average loss is
$$\ell(i) = \mathbf{\Psi}(i, R) \odot (P(i, R) B(i, R)) \tag{12}$$
5. The state is updated.
$$\mathbf{\Psi}(i+1) = \mathbf{E}_{R \sim \mathbf{\Psi}(i)}[R \oplus Q(i, R)] \oplus -\ell(i) \tag{13}$$
Where $Q(i, R)$ is the distribution of the losses of actions with respect to which the regret is R after iteration $i - 1$. $\oplus$ denotes the convolution as defined above.

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

Figure 3: The integer time game

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

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

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

A few things worth noting:

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

6.2 Upper and Lower potentials

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

$$\exists P, \quad \forall Q, \quad \forall 1 \leq i \leq T, \quad \forall R \in \mathbb{R}, \quad \phi_P^{\downarrow}(i, R) \geq \phi_{P,Q}(i, R) \tag{17}$$

$$\exists Q, \quad \forall P, \quad \forall 1 \leq i \leq T, \ \forall R \in \mathbb{R}, \ \phi^{\uparrow}_{Q}(i,R) \leq \phi_{P,Q}(i,R) \tag{18}$$

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

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

$$\mathbf{\Psi}_P(T) \odot \boldsymbol{\phi}_T = \Phi^{\downarrow}_{P}(T) = \Phi^{\downarrow}_{P}(T-1) = \cdots = \Phi^{\downarrow}_{P}(0) = \phi^{\downarrow}_{P}(0,0) \tag{19}$$

and

$$\mathbf{\Psi}_Q(T) \odot \boldsymbol{\phi}_T = \Phi^{\uparrow}_{Q}(T) = \Phi^{\uparrow}_{Q}(T-1) = \cdots = \Phi^{\uparrow}_{Q}(0) = \phi^{\uparrow}_{Q}(0,0) \tag{20}$$

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

$$\forall i, R, \quad \phi^{\uparrow}_{Q}(i,R) = \phi^{\downarrow}_{P}(i,R) \tag{21}$$

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

We do not achieve this for the integer game described in the next section. To achieve min/max optimality we extend the integer time game to the discrete time game (section 7) and to the continuous time game (8).

6.3 Strategies for the integer time game

We assume that $\boldsymbol{\phi}_T \in \mathcal{P}^2$, in other words, the final potential is positive, increasing and convex. $\boldsymbol{\phi}_T$ defines the upper and lower potentials at time T:

$$\phi^{\uparrow}_{Q_I}(T,R) = \phi^{\downarrow}_{P_I}(T,R) = \boldsymbol{\phi}_T(R)$$

We define a backwards recursion for the lower potential:

$$\phi^{\uparrow}_{Q_I}(i-1,R) = \frac{\phi^{\uparrow}_{Q_I}(i,R+1) + \phi^{\uparrow}_{Q_I}(i,R-1)}{2} \tag{22}$$

and a backwards recursion for the upper potential:

$$\phi^{\downarrow}_{P_I}(i-1,R) = \frac{\phi^{\downarrow}_{P_I}(i,R+2) + \phi^{\downarrow}_{P_I}(i,R-2)}{2} \tag{23}$$

We define strategies that correspond to these potentials. A strategy for the adversary:

$$Q_I(i,R) = \begin{cases} +1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases} \tag{24}$$

and a strategy for the learner:

$$P_I(i,R) = \frac{1}{Z} \frac{\phi(i,R+2) - \phi(i,R-2)}{2} \tag{25}$$

Where Z is a normalization factor

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

The following lemma states that these strategies guarantee the corresponding potentials.

Lemma 3.
Let i be an integer between 1 and T
If $\phi^{\uparrow}_{Q_I}(i,R) \in \mathcal{P}^2$

1. **Positivity:** $\phi^{\uparrow}_{Q_I}(i-1,R) \in \mathcal{P}^2$
2. **Adversary:** *The adversarial strategy (24) guarantees the recursion given in Eq. (22)*

If $\phi^{\downarrow}_{P_I}(i,R) \in \mathcal{P}^2$

1. **Positivity:** $\phi^{\downarrow}_{P_I}(i-1,R) \in \mathcal{P}^2$
2. **Learner:** *The learner strategy (25) guarantees the recursion given in Eq. (23)*

Proof. We prove each claim in turn

1. **Positivity:** Follows from Lemma 1.
2. **Adversary:** By symmetry adversarial strategy (24) guarantees that the aggregate loss (12) is zero regardless of the choice of the learner: $\ell(i) = 0$. Therefor the state update (13) is equivalent to the symmetric random walk:
$$\mathbf{\Psi}(i) = \frac{1}{2}((\mathbf{\Psi}(i) \oplus 1) + (\mathbf{\Psi}(i) \ominus 1))$$
Which in turn implies that if the adversary plays Q^* and the learner plays an arbitrary strategy P
$$\phi^{\uparrow}_{Q_I}(i-1,R) = \frac{\phi^{\uparrow}_{Q_I}(i,R-1) + \phi^{\uparrow}_{Q_I}(i,R+1)}{2} \tag{26}$$
As this adversarial strategy is oblivious to the learner's strategy, it guarantees that the average value at iteration i is *equal* to the average of the lower value at iteration i.
3. **Learner:** Plugging learner's strategy (25) into equation (12) we find that
$$\ell(i) = \frac{1}{Z_i}\mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\left(\phi^{\downarrow}_{P_I}(i,R+2) - \phi^{\downarrow}_{P_I}(i,R-2)\right)B(i,R)\right] \tag{27}$$
Consider the score at iteration i when the learner's strategy is P^* and the adversarial strategy Q is arbitrary
$$\Phi_{P^*,Q}(i,R) = \mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\mathbf{E}_{y\sim Q(i)(R)}\left[\phi(i,R+y-\ell(i))\right]\right] \tag{28}$$
As $\phi(i,\cdot)$ is convex and as $y-\ell(i) \in [-2,2]$,
$$\phi^{\downarrow}_{P_I}(i-1,R+y) \le \frac{\phi^{\downarrow}_{P_I}(i,R+2) + \phi^{\downarrow}_{P_I}(i,R-2)}{2} + (y-\ell(i))\frac{\phi^{\downarrow}_{P_I}(i,R+2) - \phi^{\downarrow}_{P_I}(i,R-2)}{2} \tag{29}$$
Combining the equations (27) and (28) we find that
$$\begin{aligned}
\Phi_{P^*,Q}(i,R) &= \mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\mathbf{E}_{y\sim Q(i)(R)}\left[\phi^{\downarrow}_{P_I}(i,R+y-\ell(i))\right]\right] && (30)\\
&\le \mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\frac{\phi^{\downarrow}_{P_I}(i,R+2) + \phi^{\downarrow}_{P_I}(i,R-2)}{2}\right] && (31)\\
&+ \mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\mathbf{E}_{y\sim Q(i)(R)}\left[(y-\ell(i))\frac{\phi^{\downarrow}_{P_I}(i,R+2) - \phi^{\downarrow}_{P_I}(i,R-2)}{2}\right]\right] && (32)
\end{aligned}$$

The final step is to show that the term (32) is equal to zero. As $\ell(i)$ is a constant with respect to R and y the term (32) can be written as:

$$\begin{aligned} & \mathbf{E}_{R\sim\boldsymbol{\Psi}(i)}\left[\mathbf{E}_{y\sim Q(i)(R)}\left[(y-\ell(i))\frac{\phi^{\downarrow}_{P_I}(i,R+2)-\phi^{\downarrow}_{P_I}(i,R-2)}{2}\right]\right] && (33)\\ = \ & \mathbf{E}_{R\sim\boldsymbol{\Psi}(i)}\left[B(i,R)\frac{\phi^{\downarrow}_{P_I}(i,R+2)-\phi^{\downarrow}_{P_I}(i,R-2)}{2}\right] && (34)\\ - \ & \ell(i)\mathbf{E}_{R\sim\boldsymbol{\Psi}(i)}\left[\frac{\phi^{\downarrow}_{P_I}(i,R+2)-\phi^{\downarrow}_{P_I}(i,R-2)}{2}\right] && (35)\\ = \ & 0 && (36)\end{aligned}$$

□

Repeating the induction steps of Lemma 3 from $i=T$ to $i=1$ yields the following theorem.

Theorem 4. *Let $\boldsymbol{\phi}_T \in \mathcal{P}^2$, for any iteration $0 \le i \le T$ and regret $R_0 \in \mathbb{R}$*

- *The lower potential guaranteed by Q_I is*

$$\phi^{\uparrow}_{Q_I}(i,R_0) = \mathbf{E}_{R\sim R_0\oplus\mathbb{B}(T-i,1)}\left[\boldsymbol{\phi}_T(R)\right]$$

- *The upper potential guaranteed by P_I is*

$$\phi^{\downarrow}_{P_I}(i,R_0) = \mathbf{E}_{R\sim R_0\oplus\mathbb{B}(T-i,2)}\left[\boldsymbol{\phi}_T(R)\right]$$

Plugging in $i=0$,$R=0$ we get the following Corrolary:

Corollary 5. *if the learner plays P_I on every iteration it guarantees that the final score satisfies*

$$\boldsymbol{\Psi}(T)\odot\boldsymbol{\phi}_T \le \mathbb{B}(T,2)\odot\boldsymbol{\phi}_T$$

If the Adversary plays Q_I on every iteration it guarantees that:

$$\boldsymbol{\Psi}(T)\odot\boldsymbol{\phi}_T = \mathbb{B}(T,1)\odot\boldsymbol{\phi}_T$$

7 From integer to discrete time

The upper and lower bound on the final score given in Theorem 4 do not match. If $\boldsymbol{\phi}_T \in \mathcal{P}^2$ then $\mathbb{B}(T,1)\odot\boldsymbol{\phi}_T < \mathbb{B}(T,2)\odot\boldsymbol{\phi}_T$ In other words, the strategies (24,25) are not a min/max pair.[4]

To close this gap we extend the integer time game into a new game we call the discrete time game (Fig. 4). The discrete time game increases the options available to the adversary, but not to the learner. As the integer step game is a special case of the new game, any upper potential that can be guaranteed by the learner in the discrete time game is also an upper potential for the discrete time game.

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

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

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

Initialization: $t_0 = 0$

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

1. If $t_i = \mathcal{T}$ the game terminates.
2. The adversary chooses a *step size* $0 < s_i \leq \min(\sqrt{1 - t_i}, 1)$, which advances time by $t_i = t_{i-1} + s_i^2$
3. Given s_i, the learner chooses a distribution $P(i)$ over $\mathbb{R}$.
4. The adversary chooses a mapping from $\mathbb{R}$ to distributions over $[-s_i, +s_i]$: $Q(t, \cdot) : \mathbb{R} \to \Delta^{[-s_i, +s_i]}$
5. The aggregate loss is calculated:
$$\ell(t_i) = \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[P(t_i, R) B(t_i, R) \right] \text{ where } B(t_i, R) \doteq \mathbf{E}_{y \sim Q(t_i, R)} [y] \tag{37}$$
Such that $|\ell(t_i)| \leq s_i^2$
6. The state is updated.
$$\mathbf{\Psi}(t_i) = \mathbf{E}_{R \sim \mathbf{\Psi}(t_i)} \left[Q(t_i, R) \oplus (R - \ell(t_i)) \right]$$
Where $\oplus$ is a convolution as defined in the preliminaries.

Upon termination, the final value is calculated:

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

Figure 4: The discrete time game

1. **real-valued time** In the integer time game we use an integer to indicate the iteration number: $i = 1, 2, \ldots, T$. In the discrete time game we use an positive real value, which we call "time" and use the update rule $t_{i+1} = t_i + s_i^2$, and define the final time, which is used in the regret bound, to be $\mathcal{T} = \sum_{i=0}^{T} s_i^2$
2. **Bounded average loss** We restrict the average loss to a range much smaller than $[-s_i, +s_i]$, specifically: $|\ell(i)| \leq s_i^2$

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

1. **Justification of real-valued time** To justify these choices we consider the following adversarial strategy for the discrete time game:
$$Q_D[s, p](t, R) = \begin{cases} +s & \text{w.p. } p \\ -s & \text{w.p. } 1 - p \end{cases} \tag{38}$$
From Equation (20) we get that the initial score,
$$\Phi^{\uparrow}_{Q_D}(0) = \Phi^{\uparrow}_{Q_D}(T) = \mathbf{\Psi}_{Q_D}(T) \odot \phi(T)$$
On the other hand, we know that $\mathbf{\Psi}_{Q_D}(T) = \mathbb{B}(T, s)$. Suppose T is large enough that the normal approximation for the binomial can be used. Let $\mathcal{N}(\mu, \sigma^2)$ be the normal distribution with mean μ and variance σ^2.
$$\lim_{T \to \infty} \Phi^{\uparrow}_{Q_D}(0) = \mathcal{N}(0, Ts^2) \odot \phi(T) \tag{39}$$
Recall that $\phi(T)$ is a fixed strictly convex function. It is not hard to see that if $Ts^2 \to 0$ minimizes $\Phi^{\uparrow}_{Q_D}(0)$ and makes it equal to to $\phi(T, 0)$, which means that the learner wins, while if $Ts^2 \to \infty$,

$\Phi^{\uparrow}_{Q_D}(0) \to \infty$ which means that the adversary wins. In order to keep the game balanced keep Ts^2 constant as we let $s \to 0$. We achieve that by defining the real-valued discrete time as $t_j = \sum_{i=0}^{j-1} s_i^2$.

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

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

.

If ϵ is kept constant as $s \to 0$ then $\lim_{T\to\infty} L_\ell^T = \infty$, biasing the game towards the adversary. On the other hand, if $\epsilon = s^\alpha$ for $\alpha < 1$ then $L_\ell^T \to 0$, biasing the game towards the learner. To keep the game balanced we have to set $\epsilon = cs$ for some constant c. Without loss of generality we set $c = 1$.

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

7.1 Strategies for discrete time

We fix a real number $\mathcal{T}$ as the real length of the game.

We define a sequence of adversarial strategies, indexed by k, where the step size of $Q_{D(k)}$ is $s_k = 2^{-2k}\sqrt{\mathcal{T}}$.

We define a sequence of adversarial strategies $Q_{D(k)}$ and matching learner strategies $P_{D(k)}$ for $k = 0, 1, 2, \ldots$. The adversarial strategies are designed so that the upper and lower potentials converge to a limit as $k \to \infty$.

We set the time points $t_i = is_k^2$ for $i = 0, 1, \ldots, 2^{2k}$. We call the resulting games k-discrete and denote them as $D(k)$.

For a given k we define upper and lower potentials for each t_i. This is done by induction starting with the final potential function $\boldsymbol{\phi}_{\mathcal{T}}(R) = \phi^{\uparrow}_{Q_{D(k)}}(\mathcal{T}, R) = \phi^{\downarrow}_{P_{D(k)}}(\mathcal{T}, R)$ and iterating backwards for $i = T, T-1, \ldots, 0$, $t_i = is_k^2$

$$\phi^{\uparrow}_{Q_{D(k)}}(t_{i-1}, R) = \frac{\phi^{\uparrow}_{Q_{D(k)}}(t_i, R + s_k) + \phi^{\uparrow}_{Q_{D(k)}}(t_i, R - s_k)}{2} \tag{40}$$

$$\phi^{\downarrow}_{P_{D(k)}}(t_{i-1}, R) = \frac{\phi^{\downarrow}_{P_{D(k)}}(t_i, R + s_k(1 + s_k)) + \phi^{\downarrow}_{P_{D(k)}}(t_i, R - s_k(1 + s_k))}{2} \tag{41}$$

These upper and lower potentials correspond to strategies for the adversary and the learner. The adversarial strategy is

$$Q_{D(k)} = \begin{cases} +s_k & \text{w.p. } \frac{1}{2} \\ -s_k & \text{w.p. } \frac{1}{2} \end{cases} \tag{42}$$

The learner's strategy is:

$$P_{D(k)}(t_i, R) = \frac{1}{Z} \frac{\phi^{\downarrow}_{P_{D(k)}}(t_{i+1}, R + s_k(1 + s_k)) - \phi^{\downarrow}_{P_{D(k)}}(t_{i+1}, R - s_k(1 + s_k))}{2} \tag{43}$$

$$\text{where } Z = \mathbf{E}_{R\sim\boldsymbol{\Psi}(t_{i+1})} \left[\frac{\phi^{\downarrow}_{P_{D(k)}}(t_{i+1}, R + s_k(1 + s_k)) - \phi^{\downarrow}_{P_{D(k)}}(t_{i+1}, R - s_k(1 + s_k))}{2} \right]$$

The potentials and strategies defined above are scaled versions of the integer time potential recursions defined in Equations (22,23) and the strategies defined in Equations (24,25). Specifically, the games operate on lattices that we will now describe.

The adversarial strategy Q_I defines the following lattice over i and R:

$$\boldsymbol{I}_T = \{(i, 2j-i)\,|0 \leq i \leq T, 0 \leq j \leq i\}$$

The k'th adversarial strategy $Q_{D(k)}$ uses step size $s_k = \sqrt{\mathcal{T}}2^{-k}$ and time increments $s_k^2 = \mathcal{T}2^{-2k}$. We define the *game lattice* for k as the set of (t, R) pairs that are reached by $Q_{D(k)}$.

$$\boldsymbol{K}_{\mathcal{T},k} = \left\{(t, R)\,\middle|t = is_k^2, 0 \leq i \leq 2^{2k}, R = (2j-i)s_k, 0 \leq j \leq i\right\} \tag{44}$$

$\boldsymbol{I}_T$ is a special case of $\boldsymbol{K}_{\mathcal{T},k}$ because setting $\mathcal{T} = T = 2^{2k}$ we get that $s_k = s_k^2 = 1$ and $\boldsymbol{K}_{\mathcal{T},k} = \boldsymbol{I}_T$.

It is not hard to show that the lattices get finer with k, i.e. if $j \leq k$, $\boldsymbol{K}_{\mathcal{T},j} \subseteq \boldsymbol{K}_{\mathcal{T},k}$.

The following Lemma parallels Lemma 3 for the integer time game.

Lemma 6.
Let i be an integer between 1 and T

If $\phi^{\uparrow}_{Q_{D(k)}}(t_i, R) \in \mathcal{P}^2$

1. $\phi^{\uparrow}_{Q_{D(k)}}(t_{i-1}, R) \in \mathcal{P}^2$
2. *The adversarial strategy (42) guarantees the recursion given in Eq. (40)*

If $\phi^{\downarrow}_{P_{D(k)}}(t_i, R) \in \mathcal{P}^2$

1. $\phi^{\downarrow}_{P_{D(k)}}(t_{i-1}, R) \in \mathcal{P}^2$
2. *The learner strategy (43) guarantees the recursion given in Eq. (41)*

Proof. The statement of the Lemma and the proof are scaled versions of Lemma 3 and its proof. The iteration step is s_k^2 instead of 1 while the loss/gain of an action in a single step is $[-s_k, s_k]$ instead of $[-1, +1]$.

One change worth noting is at the step from Equation (28) and Equation (29), where the bound $y - \ell(i) \in [-2, 2]$ is replace by $y - \ell(i) \in [-s_k - s_k^2, s_k + s_k^2]$. This follows from the bound $|\ell(i)| \leq s_k^2$ which is discussed in Section 7. □

Theorem 7. *Let $\boldsymbol{\phi}_{\mathcal{T}} \in \mathcal{P}^2$ be the final potential in the discrete time game. Fix k and the step size $s_k = \sqrt{\mathcal{T}}2^{-k}$, and let $t_i = is_k^2$ for $i = 0, 1, \ldots, 2^{2k}$ and let R_0 be a real value, then*

- *The lower potential guaranteed by $Q_{D(k)}$ is*

$$\phi^{\uparrow}_{Q_{D(k)}}(t_i, R_0) = \mathbf{E}_{R \sim R_0 \oplus \mathbb{B}(2^{2k}-i, s_k)}\left[\boldsymbol{\phi}_{\mathcal{T}}(R)\right] \tag{45}$$

- *The upper potential guaranteed by $P_{D(k)}$ is*

$$\phi^{\downarrow}_{P_{D(k)}}(t_i, R_0) = \mathbf{E}_{R \sim R_0 \oplus \mathbb{B}(2^{2k}-i, s_k(1+s_k))}\left[\boldsymbol{\phi}_{\mathcal{T}}(R)\right] \tag{46}$$

Using Theorem 7 we can show that, as $k \to \infty$, the upper lower potential converge to the same limit.

Theorem 8.
Fix $\mathcal{T}$ and assume $\boldsymbol{\phi}_{\mathcal{T}} \in \mathcal{P}^2$. Consider the sequence of upper and lower potentials $\phi^{\downarrow}_{P_{D(k)}}, \phi^{\uparrow}_{Q_{D(k)}}$ for $k = 0, 1, 2, \ldots$.

Then for any $0 < t \leq \mathcal{T}$ and any R_0:

$$\lim_{k \to \infty} \phi^{\downarrow}_{P_{D(k)}}(t, R_0) = \lim_{k \to \infty} \phi^{\uparrow}_{Q_{D(k)}}(t, R_0) = \mathcal{N}(R_0, \mathcal{T} - t) \odot \boldsymbol{\phi}_{\mathcal{T}} \tag{47}$$

Proof. We first assume that $(t, R_0) \in \boldsymbol{K}_{j,\mathcal{T}}$ and that $k \geq j$. We later expand to any $0 < t \leq \mathcal{T}$ and any $R_0 \in \mathbb{R}$. Consider Equation 46 for $P_{D(k)}$ and $P_{D(j)}$, keeping t and j constant and letting $k \to \infty$.

$$\phi^{\downarrow}_{P_{D(j)}}(t, R_0) = \mathbf{E}_{R \sim R_0 \oplus \mathbb{B}(2^{2j} - i_j, s_j(1+s_j))} \left[\boldsymbol{\phi}_{\mathcal{T}}(R)\right] \tag{48}$$

$$\phi^{\downarrow}_{P_{D(k)}}(t, R_0) = \mathbf{E}_{R \sim R_0 \oplus \mathbb{B}(2^{2k} - i_k, s_k(1+s_k))} \left[\boldsymbol{\phi}_{\mathcal{T}}(R)\right] \tag{49}$$

We rewrite the binomial factor in Eq (49)

$$\mathbb{B}\big(2^{2k} - i_k, s_k(1+s_k)\big) = \mathbb{B}\Big(2^{2(k-j)}\big(2^{2j} - i_j\big), 2^{j-k} s_j(1 + 2^{j-k} s_j)\Big)$$

As j is constant, s_j is constant and so is $a_j \doteq 2^{2j} - i_j$. Multiplying the number of steps by the variance per step we get

$$Var(\mathbb{B}_k) = 2^{2(k-j)} a_j \big(2^{j-k} s_j (1 + (2^{j-k} s_j)\big)^2 = a_j s_j^2 (1 + (2^{j-k} s_j))^2$$

As s_j, a_j are constants we get that $\lim_{k\to\infty} Var(\mathbb{B}_k) = a_j s_j$. From the central limit theorem we get that for any $(t, R_0) \in \boldsymbol{K}_{j,\mathcal{T}}$

$$\lim_{k\to\infty} \phi^{\downarrow}_{P_{D(k)}}(t, R_0) \odot \boldsymbol{\phi}_{\mathcal{T}} = \mathcal{N}(R_0, \mathcal{T} - t) \odot \boldsymbol{\phi}_{\mathcal{T}} \tag{50}$$

Our argument hold for all $(t, R_0) \in \bigcup_{k=0}^{\infty} \boldsymbol{K}_{k,\mathcal{T}}$ which is dense in the set $0 < t \leq \mathcal{T}, R_0 \in \mathbb{R}$. On the other hand, $\phi^{\downarrow}_{P_{D(k)}}(t, R) \odot \boldsymbol{\phi}_{\mathcal{T}}$ is continuous in both t and R, therefor Equation (50) holds for all t and R.

As similar (slightly simpler) argsument holds for the lower potential limit $\lim_{k\to\infty} \phi^{\uparrow}_{Q_{D(k)}}(t, R_0)$

□

We have shown that in the limit $s \to 0$ the learner and the adversary converge to the same potential function. In the next section we show that this limit is the min/max solution by describing conditions under which the adversary prefers using ever smaller steps size.

7.2 The adversary prefers smaller steps

Theorem 50 characterizes the limits of the upper and lower potentials, as $k \to \infty$ are equal to each other and to $\mathcal{N}(R_0, \mathcal{T} - t) \odot \boldsymbol{\phi}_{\mathcal{T}}$ To show that this limit corresponds to the min/max slolution of the game we need to show that the adversary perfers smaller steps. In other words, that for any t, R, $\phi^{\uparrow}_{Q_{D(k)}}(t, R)$ increases with k.

To prove this claim we strengther the condition $\boldsymbol{\phi}_{\mathcal{T}} \in \mathcal{P}^2$ used above to $\boldsymbol{\phi}_{\mathcal{T}} \in^4$. In words, we assume that the function $\boldsymbol{\phi}_{\mathcal{T}}(R)$ is continuous and strictly positive and it's first four derivatives are continuous and strictly positive.

We use the sequence of discrete adversarial strategies $Q_{D(k)}, k = 1, 2, \ldots$ defined in Section 7.1.

Theorem 9. *If $\boldsymbol{\phi}_{\mathcal{T}} \in \mathcal{P}^4$, and $\mathcal{T} > 0$ then for any $k \geq 0$, any $t \in [0, \mathcal{T}]$ and any R*

$$\phi^{\uparrow}_{Q_{D(k+1)}}(t, R) > \phi^{\uparrow}_{Q_{D(k)}}(t, R)$$

The proof of the theorem relies on a reduction to a simpler case: dividing a single time step of duration τ into four time steps of duration $\tau/4$

Lemma 10. *If $\boldsymbol{\phi}_{\tau} \in \mathcal{P}^4$, and $\tau > 0$ then for any R*

$$\phi^{\uparrow}_{Q_{D(1)}}(0, R) > \phi^{\uparrow}_{Q_{D(0)}}(0, R)$$

Proof. The step sizes for $k = 0, 1$ are $s_0 = \sqrt{\tau}, s_1 = \frac{\sqrt{\tau}}{2}$ and the time icrements are $\Delta t_0 = \tau, \Delta t_1 = \frac{\tau}{4}$. In other words $k = 0$ corresponds to a single step of size $\sqrt{\tau}$, while $k = 1$ corresponds to *four* steps of size $\frac{\sqrt{\tau}}{2}$.

By definition $\boldsymbol{\phi}_{\tau}(R) = \phi^{\uparrow}_{Q_{D(0)}}(\tau, R) = \phi^{\uparrow}_{Q_{D(1)}}(\tau, R)$

For $k=0$ we we get the recursion

$$\phi^{\uparrow}_{Q_{D(0)}}(0,R) = \frac{\phi^{\uparrow}_{Q_{D(0)}}(\tau, R-\sqrt{\tau}) + \phi^{\uparrow}_{Q_{D(0)}}(\tau, R+\sqrt{\tau})}{2} = \frac{\boldsymbol{\phi}_\tau(R-\sqrt{\tau}) + \boldsymbol{\phi}_\tau(R+\sqrt{\tau})}{2} \tag{51}$$

For $k=1$ we we have for $i=0,1,2,3$:

$$\phi^{\uparrow}_{Q_{D(0)}}\left(\frac{i}{4}\tau, R\right) = \frac{\phi^{\uparrow}_{Q_{D(0)}}\left(\frac{i+1}{4}\tau, R-\frac{1}{2}\sqrt{\tau}\right) + \phi^{\uparrow}_{Q_{D(0)}}\left(\frac{i+1}{4}\tau, R+\frac{1}{2}\sqrt{\tau}\right)}{2} \tag{52}$$

Combining Equation (52) for $k=0,1,2,3$ we get

$$\begin{aligned}\phi^{\uparrow}_{Q_{D(1)}}(0,R) &= \frac{1}{16}\Big[\phi^{\uparrow}_{Q_{D(1)}}(\tau, R{-}2\sqrt{\tau}) + 4\phi^{\uparrow}_{Q_{D(1)}}(\tau, R{-}\sqrt{\tau}) \\ &\quad + 6\phi^{\uparrow}_{Q_{D(1)}}(\tau, R) + 4\phi^{\uparrow}_{Q_{D(1)}}(\tau, R{+}\sqrt{\tau}) + \phi^{\uparrow}_{Q_{D(1)}}(\tau, R{+}2\sqrt{\tau})\Big] \\ &= \frac{1}{16}\left[\boldsymbol{\phi}_\tau(R{-}2\sqrt{\tau}) + 4\boldsymbol{\phi}_\tau(R{-}\sqrt{\tau}) + 6\boldsymbol{\phi}_\tau(R) + 4\boldsymbol{\phi}_\tau(R{+}\sqrt{\tau}) + \boldsymbol{\phi}_\tau(R{+}2\sqrt{\tau})\right]\end{aligned} \tag{53}$$

the difference between Equations (53) and (51) is

$$\begin{aligned}&\phi^{\uparrow}_{Q_{D(1)}}(0,R) - \phi^{\uparrow}_{Q_{D(0)}}(0,R) \\ &= \frac{1}{16}\left[\boldsymbol{\phi}_\tau(R{-}2\sqrt{\tau}) - 4\boldsymbol{\phi}_\tau(R{-}\sqrt{\tau}) + 6\boldsymbol{\phi}_\tau(R) - 4\boldsymbol{\phi}_\tau(R{+}\sqrt{\tau}) + \boldsymbol{\phi}_\tau(R{+}2\sqrt{\tau})\right]\end{aligned} \tag{54}$$

Our goal is to show that the LHS of Eqn. 54 is positive. This is equivalent to proving positivity of

$$\begin{aligned}g_a(R) &= \frac{2}{3a^2}\left(\phi^{\uparrow}_{Q_{D(1)}}(0,R) - \phi^{\uparrow}_{Q_{D(0)}}(0,R)\right) \\ &= \frac{1}{24a^4}\left[\boldsymbol{\phi}_\tau(R{-}2a) - 4\boldsymbol{\phi}_\tau(R{-}a) + 6\boldsymbol{\phi}_\tau(R) - 4\boldsymbol{\phi}_\tau(R{+}a) + \boldsymbol{\phi}_\tau(R{+}2a)\right]\end{aligned} \tag{55}$$

where $a = 2s_1 = \sqrt{\tau}$

The function $g_a(R)$ has a special form called "divided differences". The proof of the following lemma uses this fact to show that Eqn (55) is strictly positive.

Lemma 11. *If $\boldsymbol{\phi}_\tau \in \mathcal{P}^4$ and $\tau > 0$, then $\forall R, g_a(R) > 0$*

The proof of Lemma 11 is given in appendix B

Proof. of Theorem 9

Let $\boldsymbol{K}_{\mathcal{T},k}$ be the game lattices defined in Equation 44. The statement of Lemma 10 can be trivially generalized as follows. Fix $k \geq 0$ and $t = is_k^2$ then

$$\text{if } \phi((i+1)s_k^2, R) \in \mathcal{P}^4 \text{ then } \phi^{\uparrow}_{Q_{D(k+1)}}(is_k^2, R) > \phi^{\uparrow}_{Q_{D(k)}}(is_k^2, R) \tag{56}$$

The proof of the theorem is by double induction, first on $k = 0,1,2,\ldots$ and then on $t = \tau - s_k^2, \tau - 2s_k^2, \ldots, 0$.

The base case $k=0, t=\tau$ follows directly from Lemma 10.

For $k > 0$ we use a second induction $t = \tau - s_{k-1}^2, \tau - 2s_{k-1}^2, \ldots, 0$, which, combined with 56 holds for $k-1$, shows that for any $(t,R) \in \boldsymbol{K}_{\mathcal{T},k-1}$ we have $\phi^{\uparrow}_{Q_{D(k)}}(t,R) > \phi^{\uparrow}_{Q_{D(k-1)}}(t,R)$

The theorem follows by forward induction on k. □

Combining Theorems 8 and 9 we get our characterization of the min/max solution for the potential game.

Theorem 12. *Let $\mathcal{T} > 0$ and $\boldsymbol{\phi}_{\mathcal{T}} \in \mathcal{P}^4$, define $\phi^*(t,R) = \mathcal{N}(R_0, \mathcal{T} - t) \odot \boldsymbol{\phi}_{\mathcal{T}}$ Fix a state at time $t < \mathcal{T}$ $\boldsymbol{\Psi}(t)$ and define the score of this state to be $\Phi(t) = \boldsymbol{\Psi}(t) \odot \phi(t)$. Then for any $\epsilon > 0$ there exists*

- *A learner strategy that guarantees* $\Phi(T) \odot \boldsymbol{\phi}_{\mathcal{T}} \leq \Phi(t) + \epsilon$ *against any adversarial strategy and*
- *an adversarial strategy that guarantees* $\Phi(T) \odot \boldsymbol{\phi}_{\mathcal{T}} \geq \Phi(t) - \epsilon$

Proof. From Theorem 8 we know that the $k \to \infty$. From Theorem 9 we know that the small step limit maximizes the

Together these theorems show that the the min/max optimal potential function is $\mathcal{N}(R_0, \mathcal{T} - t) \odot \boldsymbol{\phi}_{\mathcal{T}}$.

8 Brownian motion

There seems to be a paradox: the adversary prefers to set $s_i > 0$ as small as possible. On the other hand, there is no minimal strictly positive number, so whatever the adversary chooses has to be suboptimal. In other words, time is not continuous, it increases in discrete steps. As that is the case, why is brownian motion still the correct way to compute the potential?

One can use the following argument: the learner knows the range $[-s_i, +s_i]$ for the next instantanous losses before it has to choose the weights he places on the actions. On the other hand, it does not know the range of the following losses, but he knows that the adversary always prefers small ranges. The safe thing for the learner to do is to assume that the following steps will be infinitesmaly small, i.e. that the future losses form a brownian process

It is well known that the limit of random walks where $s \to 0$ and $\Delta t = s^2$ is the the Brownian or Wiener process (see [17]).

An alternative characterization of Brownian Process is

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

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

$$\frac{\partial}{\partial t}\phi(t, R) + \frac{1}{2}\frac{\partial^2}{\partial R^2}\phi(t, R) = 0 \tag{57}$$

Given a final value function with a strictly positive fourth derivative we can use Equation (57) to compute the value function for all $0 \leq t \leq T$. We will do so in he next section.

9 The continuous time game and bounds for easy sequences

In Section 7 we have shown that the integer time game has a natural extension to a setting where $\Delta t_i = s_i^2$. We also demonstrated sequences of adversarial strategies $S_1, S_2, \ldots$ such that $\sup_{k \to \infty} \phi^{\uparrow}_{Q_k}(0, R) =$

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

To see that such an improvement is possible, consider the following *constant* adversary. This adversary associates the same loss to all actions 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 behavior. The constant adversary described above for which $t_{i+1} = t_i$, and the random walk adversary described earlier, in which each action is split into two, one half with loss $-s_i$ and the other with loss $+s_i$. In which case $t_{i+1} = t_i + s_i^2$ which is the maximal increase in t that the adversary can guarantee. The analysis below shows that these are two extremes on a spectrum and that intermediate cases can be characterized using a variance-like quantity.

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

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

1. If $t_i = T$ the game terminates.
2. Given t_i, the learner chooses a distribution $P(i)$ over $\mathbb{R}$:
$$P^{cc}(t, R) = \frac{1}{Z^{cc}} \left.\frac{\partial}{\partial r}\right|_{r=R} \phi(t, r) \text{ where } Z^{cc} = \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)} \left[\left.\frac{\partial}{\partial r}\right|_{r=R} \phi(t, r) \right] \tag{58}$$
3. The adversary chooses a *step size* $0 < s_i \le s$ and a mapping from $\mathbb{R}$ to distributions over $[-s_i, +s_i]$: $Q(t) : \mathbb{R} \to \Delta^{[-s_i,+s_i]}$
4. The aggregate loss is calculated:
$$\ell(t_i) = \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)} \left[P^{cc}(t_i, R)B(t_i, R)\right], \quad \text{where } B(t_i, R) \doteq \mathbf{E}_{y\sim Q(t_i,R)} [y] \tag{59}$$
the aggregate loss is restricted to $|\ell(t_i)| \le cs_i^2$.
5. Increment $t_{i+1} = t_i + \Delta t_i$ where
$$\Delta t_i = \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)} \left[H(t_i, R) \;\; \mathbf{E}_{y\sim Q(t_i,R)} \left[(y - \ell(t_i))^2\right]\right] \tag{60}$$
Where
$$H(t_i, R) = \frac{1}{Z^H} \left.\frac{\partial^2}{\partial r^2}\right|_{r=R} \phi(t_i, r) \text{ and } Z^H = \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)} \left[\left.\frac{\partial^2}{\partial r^2}\right|_{r=R} \phi(t_i, r) \right] \tag{61}$$
6. The state is updated.
$$\mathbf{\Psi}(t_{i+1}) = \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)} \left[Q(t_i)(R) \oplus (R - \ell(t_i))\right]$$

Figure 5: The continuous time game and learner strategy

Theorem 13. *Let $\phi \in \mathcal{P}^\infty$ be a potential function that satisfies the Kolmogorov backward equation (57) Then*
$$\Phi(\mathbf{\Psi}(\tau)) \le \Phi(\mathbf{\Psi}(0))$$

The proof is given in appendix C
If we define
$$V_n = t_n = \sum_{i=1}^{n} \Delta t_i = \sum_{i=1}^{n} \mathbf{E}_{R\sim\mathbf{\Psi}(i)} \left[\mathbf{E}_{y\sim Q(t_i,R)} \left[H(t_i, R)((y - \ell_i)^2)\right]\right] \tag{62}$$
We can use V_n instead of T giving us a variance based bound.

10 Anytime potential functions

The results up to this point hold for any potential function in $\mathcal{P}^4$. Given a final potential function $\boldsymbol{\phi}_{\mathcal{T}} \in \mathcal{P}^4$ we can compute the potential for any $0 \le t \le \mathcal{T}$ and any R using the equation
$$\phi(t, R) = \mathcal{N}(R_0, \mathcal{T} - t) \odot \boldsymbol{\phi}_{\mathcal{T}} \tag{63}$$
By using uses the R-derivative of this potential function to define the weights the learner guarantees that the final average score is at most $\phi(0, 0)$.

A major limition of this result is that the horizon $\mathcal{T}$ is set in advance. It is desirable that the potential is defined without knowledge of the horizon. In what follows we show that Hedge and NormalHedge can both be used in such "anytime" algorithms.

Our solution is based on the observation that a potential function satisfies Eqn (63) if and only if it satisfies the Kolmogorov backwards PDE (57):

$$\frac{\partial}{\partial t}\phi(t,R) + \frac{1}{2}\frac{\partial^2}{\partial r^2}\phi(t,R) = 0 \tag{64}$$

The potential $\boldsymbol{\phi}_{\mathcal{T}} \in \mathcal{P}^4$ defines a boundary condition of the PDE.

We derive our anytime algorithm by finding solutions to the Kolmogorov PDE that are not restricted in time, and that have a fixed parametric form. In other words, the evolution of the potential with time is defined by changing the parameter values, without changing the form.

We describe two potential functions that are solutions of PDE. In the following section we use our general results to prove simultanous regret bounds

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

$$\phi_{\text{exp}}(R,t) = e^{\sqrt{2}\eta R - \eta^2 t} \tag{65}$$

Equation 65.

For the standard (non simultanous) bound we fix ϵ and t, choose $\eta = \sqrt{\frac{\ln(1/\epsilon)}{t}}$ and get a bound of the form

$$R_\epsilon \leq \sqrt{2t \ln \frac{1}{\epsilon}} \tag{66}$$

To derive a simultanous regret bound we fix the learning rate η and take the reciprocal of the potential:

$$G(R,t) \leq e^{\eta^2 t - \sqrt{2}\eta R}$$

Which holds for any t and R. The bound $G(R,t)$ depends on η.

The NormalHedge potential function parametrized by $\nu > 0$ is:

$$\phi_{\text{NH}(\nu)}(R,t) = \begin{cases} \frac{1}{\sqrt{t+\nu}} \exp\left(\frac{R^2}{2(t+\nu)}\right) & \text{if } R \geq 0 \\ \frac{1}{\sqrt{t+\nu}} & \text{if } R < 0 \end{cases} \tag{67}$$

The function $\phi_{\text{NH}}(R,t)$ is not in $\mathcal{P}^4$, however, the positive part $R \geq 0$ is in $\mathcal{P}^4$ while the negative part $R \leq 0$ is a constant. A constant potential corresponds to zero weight which means that actions whose regret is negative are ignored by the learner. In this case the optimal adversarial is not unconstrained brownian motion, instead it is brownian motion with a reflective boundary at $R = 0$.

11 Upper and lower bounds on the simultanous regret

Plan: (1) first order bound with Variable ν to show optimality. (2) Second order bound with $\nu = 1$

We now combine Theorems **??** and 13 with the Normal-Hedge potential (67) to derive a second order bound on the regret of NormalHedge.

$$R_{\text{NH}(\nu)}(\epsilon) \leq \sqrt{(t_i+\nu)\left(2\ln\frac{1}{2\epsilon} + \ln(t_i+\nu)\right)} \tag{68}$$

Where $t_i = \nu + \sum_{j=1}^{i} \Delta t_j$

$$\Delta t_i = \mathbf{E}_{R\sim\Psi(i)}\left[H(t_i,R)\ \mathbf{E}_{y\sim Q(t_i,R)}\left[(y-\ell_i)^2\right]\right] \tag{69}$$

and

$$W(t_i, R) = e^{\frac{R^2}{2t_i}} \left(\frac{1}{t^{3/2}} + \frac{R^2}{t^{5/2}} \right) \tag{70}$$

$$H(t_i, R) = \frac{W(t_i, R)}{\mathbf{E}_{\rho \sim \mathbf{\Psi}(i)} [W(t_i, \rho)]} \tag{71}$$

The initial potential is $1/\sqrt{\nu}$ and it remains this way in the continuous case. In the discrete case it is $1/\nu + O(1/n)$ where n is the number of steps We can upper bound the potential by $1/\nu + 1$

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] P Billingsley. The central limit theorem. *Probability and Measure, 3rd ed.; John Wiley & Sons, Inc.: Toronto, ON, Canada*, pages 357–363, 1995.

[4] 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.

[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] Alexey Chernov and Vladimir Vovk. Prediction with expert evaluators' advice. In *International Conference on Algorithmic Learning Theory*, pages 8–22. Springer, 2009.

[10] Carl de Boor. Divided differences. *arXiv preprint math/0502036*, 2005.

[11] Yoav Freund and Manfred Opper. Drifting games and brownian motion. *Journal of Computer and System Sciences*, 64(1):113–132, 2002.

[12] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[13] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.

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

[15] Pierre Gaillard, Gilles Stoltz, and Tim Van Erven. A second-order bound with excess losses. In *Conference on Learning Theory*, pages 176–196. PMLR, 2014.

[16] Elad Hazan and Satyen Kale. Extracting certainty from uncertainty: Regret bounded by variation in costs. *Machine learning*, 80(2):165–188, 2010.

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

[18] Wouter M Koolen and Tim Van Erven. Second-order quantile methods for experts and combinatorial games. In *Conference on Learning Theory*, pages 1155–1175. PMLR, 2015.

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

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

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

A Proof of Theorem ??

Proof.

- **A distribution $\boldsymbol{\Psi}$ satisfies SRB for G if it satisfies APB for $\phi(R) = G(R)^{-1}$**
  Assume by contradiction that $\boldsymbol{\Psi}$ does not satisfy the simultaneous 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 $\phi(R) = B(R)^{-1}$
- **A distribution $\boldsymbol{\Psi}$ satisfies an APB for ϕ if it satisfies SRB for G and $\int_{-\infty}^{\infty} \phi(R)G(R)dR \leq 1$**
  From the SRB condition we have
  $$\forall R, \mathbf{P}_{\rho\sim\boldsymbol{\Psi}}\left[\rho \geq R\right] \leq G(R)$$
  From the condition $\int_{-\infty}^{\infty} \phi(R)G(R)dR \leq 1$ we get
  $$1 \geq \int_{-\infty}^{\infty} \phi(R)G(R)dR \geq \int_{-\infty}^{\infty} \phi(R)\mathbf{P}_{\rho\sim\boldsymbol{\Psi}}\left[\rho \geq R\right] dR = \mathbf{E}_{R\sim\boldsymbol{\Psi}}\left[\phi(R)\right]$$

□

B Divided differences of a function

The function $g_s(R)$ has a special form called "divided difference" that has been extensively studied [20, 4, 10]. and is closely related to to derivatives of different orders. Using this connection and the fact that $\phi(\cdot, R) \in \mathcal{P}^4$ we prove the following lemma:

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.

A function $\boldsymbol{\phi}$ that satisfies inequality **??** is said to be *4'th order convex* (see details in in [4]).

Following[4] we give a brief review of divided differences and of n-convexity.

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

Definition 2 (n'th order divided difference of a function)**.** *The n'th order divided different of a function $f : [a, b] \to \mathbb{R}$ at mutually distinct and ordered points $a \le x_0 < x_1 < \cdots < x_n \le b$ defined recursively by*

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

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

Definition 3 (n-convexity)**.** *A function $f : [a, b] \to \mathbb{R}$ is said to be n-convex $n \ge 0$ if and only if for all choices of $n+1$ distinct points: $a \le x_0 < x_1 < \cdots < x_n \le b$, $[x_0, \ldots, x_n; f] \ge 0$ holds.*

n-convexity is has a close connection to the sign of $f^{(n)}$ - the n'th derivative of f, this connection was proved in 1965 by popoviciu [20].

Theorem 14. *If $f^{(n)}$ exists then f is n-convex if and only if $f^{(n)} \ge 0$.*

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

Proof. **of Lemma (11)**

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

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

$$h(x_0, x_1, x_2, x_3, x_4) = \frac{1}{24s^4}(f(x_4) - 4f(x_3) + 6f(x_2) - 4f(x_1) + f(x_0)) \tag{72}$$

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

1.

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

2.

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

3.

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

4.

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

5.

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

$\square$

C Proof of Theorem 13

We start with two technical lemmas

Lemma 15. *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).

Assume 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. □

Lemma 16. *Let $f(x,y)$ be a differentiable function with continuous derivatives up to degree three. Then*

$$f(x_0+\Delta x, y_0+\Delta y) = f(x_0,y_0) + \left\{ \frac{\partial}{\partial x}\bigg|_{\substack{x,\,y\,=\\ x_0,\,y_0}} f(x,y) \right\}\Delta x + \left\{ \frac{\partial}{\partial y}\bigg|_{\substack{x,\,y\,=\\ x_0,\,y_0}} f(x,y) \right\}\Delta y \tag{73}$$

$$+ \quad \frac{1}{2}\left\{ \frac{\partial^2}{\partial x^2}\bigg|_{\substack{x,\,y\,=\\ x_0,\,y_0}} f(x,y) \right\}\Delta x^2 + \left\{ \frac{\partial^2}{\partial x \partial y}\bigg|_{\substack{x,\,y\,=\\ x_0,\,y_0}} f(x,y) \right\}\Delta x \Delta y + \frac{1}{2}\left\{ \frac{\partial^2}{\partial y^2}\bigg|_{\substack{x,\,y\,=\\ x_0,\,y_0}} f(x,y) \right\}\Delta y^2 \tag{74}$$

$$+ \quad \frac{1}{6}\left\{ \frac{\partial^3}{\partial x^3}\bigg|_{\substack{x,\,y\,=\\ x_0+t\Delta x,\,y_0+t\Delta y}} f(x,y) \right\}\Delta x^3 + \frac{1}{2}\left\{ \frac{\partial^3}{\partial x^2 \partial y}\bigg|_{\substack{x,\,y\,=\\ x_0+t\Delta x,\,y_0+t\Delta y}} f(x,y) \right\}\Delta x^2 \Delta y \tag{75}$$

$$+\frac{1}{2}\left\{ \frac{\partial^3}{\partial x \partial y^2}\bigg|_{\substack{x,\,y\,=\\ x_0+t\Delta x,\,y_0+t\Delta y}} f(x,y) \right\}\Delta x \Delta y^2 + \frac{1}{6}\left\{ \frac{\partial^3}{\partial y^3}\bigg|_{\substack{x,\,y\,=\\ x_0+t\Delta x,\,y_0+t\Delta y}} f(x,y) \right\}\Delta y^3 \tag{76}$$

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

Proof. of Lemma 16 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:

$$\frac{d^n}{dt^n}F(t) = \sum_{m=1}^{n} \binom{n}{m} \frac{\partial^n}{\partial x^m \partial y^{n-m}} f(x_0+t\Delta x, y_0+t\Delta y)\Delta x^m \Delta y^{n-m} \tag{77}$$

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

$$f(x_0+\Delta x, y_0+\Delta y) = F(1) = F(0) + \frac{d}{dt}F(0) + \frac{1}{2}\frac{d^2}{dt^2}F(0) + \frac{1}{6}\frac{d^3}{dt^3}F(t') \tag{78}$$

Where $0 \leq t' \leq 1$. Using Eq (77) to expand each term in Eq. (78) completes the proof. □

Proof. of Theorem 13
We prove the claim by an upper bound on the increase of potential that holds for any iteration $1 \leq i \leq n$:

$$\Phi(\mathbf{\Psi}(t_{i+1})) \leq \Phi(\mathbf{\Psi}(i)) + a s_i^3 \text{ for some constant } a > 0 \tag{79}$$

Summing inequality (79) over all iterations we get that

$$\Phi(\mathbf{\Psi}(T)) \leq \Phi(\mathbf{\Psi}(0)) + c\sum_{i=1}^{n} s_i^3 \leq \Phi(\mathbf{\Psi}(0)) + as\sum_{i=1}^{n} s_i^2 = \Phi(\mathbf{\Psi}(0)) + asT \tag{80}$$

From which the statement of the theorem follows.

We now prove inequality (79). We use the notation $r = y - \ell(i)$ to denote the instantaneous regret at iteration i.

Applying Lemma 16 to $\phi(t_{i+1}, R_{i+1}) = \phi(t_i + \Delta t_i, R_i + r_i)$ we get

$$\begin{aligned}
\phi(t_i + \Delta t_i, R_i + r_i) &= \phi(t_i, R_i) \qquad (81)\\
&+ \left\{ \frac{\partial}{\partial \rho}\bigg|_{\substack{\tau,\,\rho\,= \\ t_i,\,R}} \phi(\tau, \rho) \right\} r_i \qquad (82)\\
&+ \left\{ \frac{\partial}{\partial \tau}\bigg|_{\substack{\tau,\,\rho\,= \\ t_i,\,R}} \phi(\tau, \rho) \right\} \Delta t_i \qquad (83)\\
&+ \frac{1}{2}\left\{ \frac{\partial^2}{\partial \rho^2}\bigg|_{\substack{\tau,\,\rho\,= \\ t_i,\,R}} \phi(\tau, \rho) \right\} r_i^2 \qquad (84)\\
&+ \left\{ \frac{\partial^2}{\partial r \partial \tau}\bigg|_{\substack{\tau,\,\rho\,= \\ t_i,\,R}} \phi(\tau, \rho) \right\} r_i \Delta t_i \qquad (85)\\
&+ \frac{1}{2}\left\{ \frac{\partial^2}{\partial \tau^2}\bigg|_{\substack{\tau,\,\rho\,= \\ t_i,\,R}} \phi(\tau, \rho) \right\} \Delta t_i^2 \qquad (86)\\
&+ \frac{1}{6}\left\{ \frac{\partial^3}{\partial \rho^3}\bigg|_{\substack{\tau,\,\rho\,= \\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \phi(\tau, \rho) \right\} r_i^3 \qquad (87)\\
&+ \frac{1}{2}\left\{ \frac{\partial^3}{\partial \rho^2 \partial \tau}\bigg|_{\substack{\tau,\,\rho\,= \\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \phi(\tau, \rho) \right\} r_i^2 \Delta t_i \qquad (88)\\
&+ \frac{1}{2}\left\{ \frac{\partial^3}{\partial \rho \partial \tau^2}\bigg|_{\substack{\tau,\,\rho\,= \\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \phi(\tau, \rho) \right\} r_i \Delta t_i^2 \qquad (89)\\
&+ \frac{1}{6}\left\{ \frac{\partial^3}{\partial \tau^3}\bigg|_{\substack{\tau,\,\rho\,= \\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \phi(\tau, \rho) \right\} \Delta t_i^3 \qquad (90)
\end{aligned}$$

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 exchangeability 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

$$
\begin{aligned}
\phi(t_i+\Delta t_i, R_i+r_i) &= \phi(t_i, R_i) && (91)\\
&+ \left\{ \frac{\partial}{\partial\rho}\bigg|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} r_i && (92)\\
&+ \left\{ \frac{\partial^2}{\partial\rho^2}\bigg|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} \left(\frac{r_i^2}{2} - \Delta t_i\right) && (93)\\
&- \left\{ \frac{\partial^3}{\partial\rho^3}\bigg|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} r_i\Delta t_i && (94)\\
&+ \frac{1}{2}\left\{ \frac{\partial^4}{\partial\rho^4}\bigg|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} \Delta t_i^2 && (95)\\
&+ \frac{1}{6}\left\{ \frac{\partial^3}{\partial\rho^3}\bigg|_{\substack{\tau,\,\rho\,=\\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \phi(\tau,\rho) \right\} r_i^3 && (96)\\
&- \frac{1}{2}\left\{ \frac{\partial^4}{\partial\rho^4}\bigg|_{\substack{\tau,\,\rho\,=\\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \phi(\tau,\rho) \right\} r_i^2\Delta t_i && (97)\\
&+ \frac{1}{2}\left\{ \frac{\partial^5}{\partial\rho^5}\bigg|_{\substack{\tau,\,\rho\,=\\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \phi(\tau,\rho) \right\} r_i\Delta t_i^2 && (98)\\
&- \frac{1}{6}\left\{ \frac{\partial^6}{\partial\rho^6}\bigg|_{\substack{\tau,\,\rho\,=\\ t_i\,+\,g\Delta t_i,\,R_i\,+\,gr_i}} \phi(\tau,\rho) \right\} \Delta t_i^3 && (99)
\end{aligned}
$$

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

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

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

- **For (92):** $r_i = 2s_i \frac{r_i}{2s_i} = 2s_i h_1(r_i)$.
- **For (93):** $r_i^2 - \frac{1}{2}\Delta t_i = 4s_i^2 \frac{r_i^2 - \frac{1}{2}\Delta t_i}{4s_i^2} = 4s_i^2 h_2(r_i, \Delta t_i)$
- **For (94):** $r_i\Delta t_i = 2s_i^3 \frac{r_i\Delta t_i}{2s_i^3} = 2s_i^3 h_3(r_i, \Delta t_i)$
- **For (95):** $\Delta t_i^2 = s_i^4 \frac{\Delta t_i^2}{s_i^4} = s_i^3 h_4(\Delta t_i)$
- **For (96):** $r_i^3 = 8s_i^3 \frac{r_i^3}{8s_i^3} = 8s_i^3 h_5(r_i, \Delta t_i)$
- **For (97):** $r_i^2\Delta t_i = 4s_i^4 \frac{r_i^2\Delta t_i}{4s_i^4} = 4s_i^3 h_6(r_i, \Delta t_i)$
- **For (98):** $r_i\Delta t_i^2 = 2s_i^5 \frac{r_i\Delta t_i^2}{2s_i^5}$

- **For (99):** $\Delta t_i^3 = s_i^6 \frac{\Delta t_i^3}{s_i^6}$

We therefor get the simplified equation

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

and therefor

$$
\begin{aligned}
\phi(t_i+\Delta t_i, R+r) &= \phi(t_i,R) + \left\{ \left.\frac{\partial}{\partial r}\right|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} r \\
&+ \left\{ \left.\frac{\partial^2}{\partial r^2}\right|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} (r^2 - \Delta t_i) + O(s^3) \qquad (100)
\end{aligned}
$$

Our next step is to consider the expected value of (100) wrt $R \sim \mathbf{\Psi}(i)$, $y \sim Q(t_i, R)$ for an arbitrary adversarial strategy Q.

We will show that the expected potential does not increase:

$$\mathbf{E}_{R\sim\mathbf{\Psi}(i)}\left[\mathbf{E}_{y\sim Q(t_i,R)}\left[\phi(t_i+\Delta t_i, R+y-\ell_i)\right]\right] \le \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\phi(t_i,R)\right] \qquad (101)$$

Plugging Eq (100) into the LHS of Eq (101) we get

$$
\begin{aligned}
&\mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\mathbf{E}_{y\sim Q(t_i,R)}\left[\phi(t_i+\Delta t_i, R+y-\ell_i)\right]\right] && (102)\\
&= \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\phi(t_i,R)\right] && (103)\\
&+ \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\mathbf{E}_{y\sim Q(t_i,R)}\left[\left\{ \left.\frac{\partial}{\partial r}\right|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\}(y-\ell_i)\right]\right] && (104)\\
&+ \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\mathbf{E}_{y\sim Q(t_i,R)}\left[\left\{ \left.\frac{\partial^2}{\partial r^2}\right|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\}((y-\ell_i)^2-\Delta t_i)\right]\right] && (105)\\
&+ O(s^3) && (106)
\end{aligned}
$$

Some care is needed here. we need to show that the expected value are all finite. We assume that the expected potential (Eq (103) is finite. Using Lemma 15 this implies that the expected value of higher derivatives of $\frac{\partial}{\partial R}\phi(R)$ are also finite.[5]

To prove inequality (79), we need to show that the terms 104 and 105 are smaller or equal to zero.

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

Term (104) is equal to zero:
As ℓ_i is a constant relative to R and y, and $\left\{ \frac{\partial}{\partial r}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\}$ is a constant with respect to y we can rewrite (104) as

$$\mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\left\{ \frac{\partial}{\partial r}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} \mathbf{E}_{y\sim Q(t_i,R)}\left[y\right]\right] - \ell_i \mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\left\{ \frac{\partial}{\partial r}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\}\right] \tag{107}$$

Combining the definitions of $\ell(t)$ (37) and and the learner's strategy P^{cc} (58) we get that

$$\ell(t_i) \quad = \quad \mathbf{E}_{R\sim\mathbf{\Psi}t_i}\left[\frac{1}{Z}\left\{ \frac{\partial}{\partial r}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\} \mathbf{E}_{y\sim Q(i,R)}\left[y\right]\right] \text{ where } Z = \mathbf{E}_{R\sim\mathbf{\Psi}t_i}\left[\frac{1}{Z}\left\{ \frac{\partial}{\partial r}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\}\right] \tag{108}$$

Plugging (108) into (107) and recalling the requirement that $\ell(t_i) < \infty$ we find that term (104) is equal to zero.

Term (105) 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 (60)

$$\mathbf{E}_{R\sim\mathbf{\Psi}(t_i)}\left[\mathbf{E}_{y\sim Q(t_i,R)}\left[Q(t_i,R)\right]\left\{ \frac{\partial^2}{\partial r^2}\Big|_{\substack{\tau,\,\rho\,=\\ t_i,\,R}} \phi(\tau,\rho) \right\}(y-\ell(t_i))^2 - \Delta t_i\right] = \Delta t_i - \Delta t_i = 0 \tag{109}$$

Where $G(t_i,R)$ is defined in Equation (**??**) We find that (105) is zero.
Finally (106) is negligible relative to the other terms as $s \to 0$. □