Minimax Optimal Algorithms for Unconstrained Linear Optimization

H. Brendan McMahan
Google Reasearch
Seattle, WA
mcmahan@google.com

Jacob Abernethy[*]
Computer Science and Engineering
University of Michigan
jabernet@umich.edu

Abstract

We design and analyze minimax-optimal algorithms for online linear optimization games where the player's choice is unconstrained. The player strives to minimize regret, the difference between his loss and the loss of a post-hoc benchmark strategy. While the standard benchmark is the loss of the best strategy chosen from a bounded comparator set, we consider a very broad range of benchmark functions. The problem is cast as a sequential multi-stage zero-sum game, and we give a thorough analysis of the minimax behavior of the game, providing characterizations for the value of the game, as well as both the player's and the adversary's optimal strategy. We show how these objects can be computed efficiently under certain circumstances, and by selecting an appropriate benchmark, we construct a novel hedging strategy for an unconstrained betting game.

1 Introduction

Minimax analysis has recently been shown to be a powerful tool for the construction of online learning algorithms [Rakhlin et al., 2012]. Generally, these results use bounds on the value of the game (often based on the sequential Rademacher complexity) in order to construct efficient algorithms. In this work, we show that when the learner is unconstrained, it is often possible to efficiently compute an exact minimax strategy for both the player and nature. Moreover, with our tools we can analyze a much broader range of problems than have been previously considered.

We consider a game where on each round $t = 1, \ldots, T$, first the learner selects $x_t \in \mathbb{R}^n$, and then an adversary chooses $g_t \in \mathcal{G} \subset \mathbb{R}^n$, and the learner suffers loss $g_t \cdot x_t$. The goal of the learner is to minimize regret, that is, loss in excess of that achieved by a post-hoc benchmark strategy. We define

$$\text{Regret} = \text{Loss} - (\text{Benchmark Loss}) = \sum_{t=1}^{T} g_t \cdot x_t - L(g_1, \ldots, g_T) \tag{1}$$

as the regret with respect to benchmark performance L (the L intended will be clear from context). The standard definition of regret arises from the choice

$$L(g_1, \ldots, g_T) = \inf_{x \in \mathcal{X}} g_{1:T} \cdot x = \inf_{x \in \mathbb{R}^n} g_{1:T} \cdot x + I(x \in \mathcal{X}), \tag{2}$$

where $I(\texttt{condition})$ is the indicator function: it returns 0 when $\texttt{condition}$ holds, and returns ∞ otherwise. The above choice of L represents the loss of the best fixed point x in the bounded convex set $\mathcal{X}$. Throughout we shall write $g_{1:t} = \sum_{s=1}^{t} g_s$ for a sum of scalars or vectors. When L depends only on the sum $G \equiv g_{1:T}$ we write $L(G)$.

In the present work we shall consider a broad notion of regret in which, for example, L is defined not in terms of a "best in hindsight" comparator but instead in terms of a "penalized best in hindsight" objective. Let Ψ be some penalty function, and consider

$$L(G) = \min_x G \cdot x + \Psi(x). \tag{3}$$

This is a direct generalization of the usual comparator notion which takes $\Psi(x) = I(x \in \mathcal{X})$.

We view this interaction as a sequential zero-sum game played over T rounds, where the player strives to minimize Eq. (1), and the adversary attempts to maximize it. We study the value of this game, defined as

$$V^T \equiv \inf_{x_1 \in \mathbb{R}^n} \sup_{g_1 \in \mathcal{G}} \ldots \inf_{x_T \in \mathbb{R}^n} \sup_{g_T \in \mathcal{G}} \left(\sum_{t=1}^T g_t \cdot x_t - L(g_1, \ldots, g_T) \right). \tag{4}$$

With this in mind, we can describe the primary contributions of the present paper:

1. We provide a characterization of the value of the game Eq. (4) in terms of the supremum over the expected value of a function of a martingale difference sequence. This will be made more explicit in Section 2.

2. We provide a method for computing the player's minimax optimal (deterministic) strategy in terms of a "discrete derivative." Similarly, we show how to describe the adversary's optimal randomized strategy in terms of martingale differences.

3. For "coordinate-decomposable" games we give a natural and efficiently computable description of the value of the game and the player's optimal strategy.

4. In Section 3, we consider several benchmark functions L, defined in Eq. (3) via a penalty function Ψ, which lead to interesting and surprising optimal algorithms; we also exactly compute the values of these games. Figure 1 summarizes these applications. In particular, we show that constant-step-size gradient descent is minimax optimal for a quadratic Ψ, and an exponential L leads to a bounded-loss hedging algorithm that can still yield exponential reward on "easy" sequences.

Applications   The primary contributions of this paper are to the theory. Nevertheless, it is worth pausing to emphasize that the framework of "unconstrained online optimization" is a fundamental template for (and strongly motivated by) several online learning settings, and the results we develop are applicable to a wide range of commonly studied algorithmic problems. The classic algorithm for linear pattern recognition, the Perceptron, can be seen as an algorithm for unconstrained linear optimization. Methods for training a linear SVM or a logistic regression model, such as stochastic gradient descent or the Pegasos algorithm [Shalev-Shwartz et al., 2011], are unconstrained optimization algorithms. Finally, there has been recent work in the pricing of options and other financial derivatives [DeMarzo et al., 2006, Abernethy et al., 2012] that can be described exactly in terms of a repeated game which fits nicely into our framework.

We also wish to emphasize that the algorithm of Section 3.2 is both practical and easily implementable: for a multi-dimensional problem one needs to only track the sum of gradients for each coordinate (similar to Dual Averaging), and compute Eq. (12) for each coordinate to derive the appropriate strategy. The algorithm provides us with a tool for making potentially unconstrained bets/investments, but as we discuss it also leads to interesting regret bounds.

Related Work   Regret-based analysis has received extensive attention in recent years; see Shalev-Shwartz [2012] and Cesa-Bianchi and Lugosi [2006] for an introduction. The analysis of alternative notions of regret is also not new. Vovk [2001] gives bounds relative to benchmarks similar to Eq. (3), though for different problems and not in the minimax setting. In the expert setting, there has been much work on tracking a shifting sequence of experts rather than the single best expert; see Koolen et al. [2012] and references therein. Zinkevich [2003] considers drifting comparators in an online convex optimization framework. This notion can be expressed by an appropriate $L(g_1, \ldots, g_T)$, but now the order of the gradients matters. Merhav et al. [2006] and Dekel et al. [2012] consider the stronger notion of policy regret in the online experts and bandit settings, respectively. Stoltz [2011] also considers some alternative notions of regret. For investing scenarios, Agarwal et al. [2006]

setting	$L(G)$	$\Psi(x)$	value	update
soft feasible set	$-\frac{G^2}{2\sigma}$	$\frac{\sigma}{2}x^2$	$\frac{T}{2\sigma}$	$x_{t+1} = -\frac{1}{\sigma}g_{1:t}$
bounded-loss betting	$-\exp(G/\sqrt{T})$	$-\sqrt{T}x\log(-\sqrt{T}x) + \sqrt{T}x$	$\to \sqrt{e}$	Eq. (12)
standard regret	$-\|G\|$	$I(\|x\| \le 1)$	$\to \sqrt{\frac{2}{\pi}T}$	Eq. (14)

Figure 1: Summary of specific online linear games considered in Section 3. Results are stated for the one-dimensional problem where $g_t \in [-1,1]$; Corollary 5 gives an extension to n dimensions. The benchmark L is given as a function of $G = g_{1:T}$. The standard notion of regret corresponds to the $L(G) = \min_{x \in [-1,1]} g_{1:t} \cdot x = -|G|$. The benchmark functions can alternatively be derived from a suitable penalty Ψ on comparator points x, so $L(G) = \min_x Gx + \Psi(x)$.

and Hazan and Kale [2009] consider regret with respect to the best constant-rebalanced portfolio. Our algorithm in Section 3.2 applies to similar problems, but does not require a "no junk bonds" assumption, and is in fact minimax optimal for a natural benchmark.

Existing algorithms do offer bounds for unconstrained problems, generally of the form $\|x^*\|/\eta + \eta \sum_t g_t x_t$. However, such bounds can only guarantee no-regret when an upper bound R on $\|x^*\|$ is known in advance and used to tune the parameter η. If one knows such a R, however, the problem is no longer truly unconstrained. The only algorithms we know that avoid this problem are those of Streeter and McMahan [2012], and the minimax-optimal algorithm we introduce in Sec 3.2; these algorithms guarantee guarantee $\text{Regret} \le \mathcal{O}\big(R\sqrt{T}\log((1+R)T)\big)$ for any $R > 0$.

The field has seen a number of minimax approaches to online learning. Abernethy and Warmuth [2010] and Abernethy et al. [2008b] give the optimal behavior for several zero-sum games against a budgeted adversary. Section 3.3 studies the online linear game of Abernethy et al. [2008a] under different assumptions, and we adapt some techniques from Abernethy et al. [2009, 2012]; the latter work also involves analyzing an unconstrained player. Rakhlin et al. [2012] utilizes powerful tools for non-constructive analysis of online learning as a technique to design algorithms; our work differs in that we focus on cases where the exact minimax strategy can be computed.

Notions of Regret The standard notion of regret corresponds to a hard penalty $\Psi(x) = I(x \in \mathcal{X})$. Such a definition makes sense when the player by definition must select a strategy from some bounded set, for example a probability from the n-dimensional simplex, or a distribution on paths in a graph. However, in contexts such as machine learning where any $x \in \mathbb{R}^n$ corresponds to a valid model, such a hard constraint is difficult to justify; while any $x \in \mathbb{R}^n$ is technically feasible, in order to prove regret bounds we compare to a much more restrictive set. As an alternative, in Sections 3.1 and 3.2 we propose soft penalty functions that encode the belief that points near the origin are more likely to be optimal (we can always re-center the problem to match our beliefs in this regard), but do not rule out any $x \in \mathbb{R}^n$ a priori.

Thus, one of our contributions is showing that interesting results can be obtained by choosing L differently than in Eq. (2). The player cannot do well in terms of the absolute loss $\sum_t g_t \cdot x_t$ for all sequences $g_1, \dots, g_T$, but she can do better on some sequences at the expense of doing worse on others. The benchmark L makes this notion precise: sequences for which $L(g_1, \dots, g_T)$ is large and negative are those on which the player desires good performance, at the expense of allowing more loss (in absolute terms) on sequences where $L(g_1, \dots, g_T)$ is large and positive. The value of the game V^T tells us to what extent any online algorithm can hope to match the benchmark L.

2 General Unconstrained Linear Optimization

In this section we develop general results on the unconstrained linear optimization problem. We start by analyzing (4) in greater detail, and give tools for computing the regret value V^T in such games. We show that in certain cases the computation of the minimax value can be greatly simplified.

Throughout we will assume that the function L is concave in each of its arguments (thought not necessarily jointly concave) and bounded on $\mathcal{G}^T$. We also include the following assumptions on the set $\mathcal{G}$. First, we assume that either $\mathcal{G}$ is a polytope or, more generally, that $\text{ConvexHull}(\mathcal{G})$ is a full-

rank polytope in $\mathbb{R}^n$. This is not strictly necessary but is convenient for the analysis; any bounded convex set in $\mathbb{R}^n$ can be approximated to arbitrary precision with a polytope. We also make the necessary assumption that the ConvexHull($\mathcal{G}$) contains the origin in its interior. We let $\mathcal{G}'$ be the set of "corners" of $\mathcal{G}$, that is $\mathcal{G}' = \{g^1, \ldots, g^m\}$ and hence ConvexHull(G) = ConvexHull($\mathcal{G}'$).

We are also concerned with the conditional value of the game, V_t, given $x_1, \ldots x_t$ and $g_1, \ldots g_t$ have already been played. That is, the Regret when we fix the plays on the first t rounds, and then assume minimax optimal play for rounds $t+1$ through T. However, following the approach of Rakhlin et al. [2012], we omit the terms $\sum_{s=1}^{t} x_s \cdot g_s$ from Eq. (4). We can view this as cost that the learner has already payed, and neither that cost nor the specific previous plays of the learner impact the value of the remaining terms in Eq. (1). Thus, we define

$$V_t(g_1, \ldots, g_t) = \inf_{x_{t+1} \in \mathbb{R}^n} \sup_{g_{t+1} \in \mathcal{G}} \ldots \inf_{x_T \in \mathbb{R}^n} \sup_{g_T \in \mathcal{G}} \left(\sum_{s=t+1}^{T} g_s \cdot x_s - L(g_1, \ldots, g_T) \right). \quad (5)$$

Note the conditional value of the game before anything has been played, $V_0()$, is exactly V^T.

The martingale characterization of the game   The fundamental tool used in the rest of the paper is the following characterization of the conditional value of the game:

Theorem 1. *For every t and every sequence $g_1, \ldots, g_t \in \mathcal{G}$, we can write the conditional value of the game as*

$$V_t(g_1, \ldots, g_t) = \max_{G \in \Delta(\mathcal{G}'), \mathbb{E}[G]=0} \mathbb{E}[V_{t+1}(g_1, \ldots, g_t, G)],$$

where $\Delta(\mathcal{G}')$ is the set of random variables on $\mathcal{G}'$. Moreover, for all t the function V_t is convex in each of its coordinates and bounded.

All proofs omitted from the body of the paper can be found in the appendix or the extended version of this paper.

Let $M_T(\mathcal{G})$ be the set of T-length *martingale difference sequences* on $\mathcal{G}'$, that is the set of all sequences of random variables $(G_1, \ldots, G_T)$, with G_t taking values in $\mathcal{G}'$, which satisfy $\mathbb{E}[G_t | G_1, \ldots, G_{t-1}] = \mathbf{0}$ for all $t = 1, \ldots, T$. Then, we immediately have the following:

Corollary 2. *We can write*

$$V^T = \max_{(G_1, \ldots, G_T) \in M_T(\mathcal{G}')} \mathbb{E}[-L(G_1, \ldots, G_T)],$$

with the analogous expression holding for the conditional value of the game.

Characterization of optimal strategies   The result above gives a nice expression for the value of the game V^T but unfortunately it does not lead directly to a strategy for the player. We now dig a bit deeper and produce a characterization of the optimal player behavior. This is achieved by analyzing a simple one-round zero-sum game. As before, we assume $\mathcal{G}$ is a bounded subset of $\mathbb{R}^n$ whose convex hull is a polytope whose interior contains the the origin $\mathbf{0}$. Assume we are given some convex function f defined and bounded on all of ConvexHull($\mathcal{G}$). We consider the following:

$$V = \inf_{x \in \mathbb{R}^n} \sup_{g \in \mathcal{G}} x \cdot g + f(g). \quad (6)$$

Theorem 3. *There exists a set of $n+1$ distinct points $\{g^1, \ldots, g^{n+1}\} \subset \mathcal{G}$ whose convex hull is of full rank, and a distribution $\vec{\alpha} \in \Delta_{n+1}$ satisfying $\sum_{i=1}^{n+1} \alpha_i g^i = \mathbf{0}$, such that $V = \sum_{i=1}^{n+1} \alpha_i f(g^i)$. Moreover, an optimal choice for the infimum in (6) is the gradient of the unique linear interpolation of the pairs $\{(g^1, -f(g^1)), \ldots, (g^{n+1}, -f(g^{n+1}))\}$.*

The theorem makes a useful point about determining the player's optimal strategy for games of this form. If the player can determine a full-rank set of "best responses" $\{g^1, \ldots, g^{n+1}\}$ to his optimal x^*, each of which should be a corner of the polytope $\mathcal{G}$, then we know that x^* must be a "discrete gradient" of the function $-f$ around $\mathbf{0}$. That is, if the size of $\mathcal{G}$ is small relative to the curvature of f, then an approximation to $-\nabla f(\mathbf{0})$ is the linear interpolation of $-f$ at a set of points around $\mathbf{0}$. An optimal x^* will be exactly this interpolation.

4

This result also tells us how to analyze the general T-round game. We can express (5), the conditional value of the game V_{t-1}, in recursive form as

$$V_{t-1}(g_1, \ldots, g_{t-1}) = \inf_{x_t \in \mathbb{R}^n} \sup_{g_t \in \mathcal{G}} g_t \cdot x_t + V_t(g_1, \ldots, g_{t-1}, g_t). \tag{7}$$

Hence by setting $f(g_t) = V_t(g_1, \ldots, g_{t-1}, g_t)$, noting that the latter is convex in g_t by Theorem 1, we see we have an immediate use of Theorem 3.

3   Minimax Optimal Algorithms for Coordinate-Decomposable Games

In this section, we consider games where $\mathcal{G}$ consists of axis-aligned constraints, and L decomposes so $L(g) = \sum_{i=1}^n L_i(g_i)$. In order to solve such games, it is generally sufficient to consider n independent one-dimensional problems. We study such games first:

Theorem 4. *Consider the one-dimensional unconstrained game where the player selects $x_t \in \mathbb{R}$ and the adversary chooses $g_t \in \mathcal{G} = [-1, 1]$, and L is concave in each of its arguments and bounded on $\mathcal{G}^T$. Then, $V^T = \mathbb{E}_{g_t \sim \{-1,1\}} \big[-L(g_1, \ldots, g_T) \big]$ where the expectation is over each g_t chosen independently and uniformly from $\{-1, 1\}$ (that is, the g_t are Rademacher random variables). Further, the conditional value of the game is*

$$V_t(g_1, \ldots, g_t) = \mathbb{E}_{g_{t+1}, \ldots, g_T \sim \{-1,1\}} \big[-L(g_1, \ldots, g_t, g_{t+1}, \ldots g_T) \big]. \tag{8}$$

The proof is immediate from Corollary 2, since the *only* possible martingale that both plays from the corners of $\mathcal{G}$ and has expectation 0 on each round is the sequence of independent Rademacher random variables.[1] Given Theorem 4, and the fact that the functions L of interest will generally depend only on $g_{1:T}$, it will be useful to define $\mathcal{B}_T$ to be the distribution of $g_{1:T}$ when each g_t is drawn independently and uniformly from $\{-1, 1\}$.

Theorem 4 can immediately be extended to coordinate-decomposable games as follows:

Corollary 5. *Consider the game where the player chooses $x_t \in \mathbb{R}^n$, the adversary chooses $g_t \in [-1, 1]^n$, and the payoff is $\sum_{t=1}^T g_t \cdot x_t - \sum_{i=1}^n L(g_{1:T,i})$ for concave L. Then the value V^T and the conditional value $V_t(\cdot)$ can be written as*

$$V^T = n \mathbb{E}_{G \sim \mathcal{B}_T} \big[-L(G) \big] \quad \text{and} \quad V_t(g_1, \ldots, g_t) = \sum_{i=1}^n \mathbb{E}_{G_i \sim \mathcal{B}_{T-t}} \big[-L(g_{1:t,i} + G_i) \big].$$

The proof follows by noting the constraints on both players' strategies and the value of the game fully decompose on a per-coordinate basis.

A recipe for minimax optimal algorithms in one dimension   Since Eq. (5) gives the minimax value of the game if both players play optimally from round $t + 1$ forward, a minimax strategy for the learner on round $t + 1$ must be $x_{t+1} = \arg\min_{x \in \mathbb{R}} \max_{g \in \{-1,1\}} g \cdot x + V_{t+1}(g_1, \ldots, g_t, g)$. Now, we can apply Theorem 3, and note that unique strategy for the adversary is to play $g = 1$ or $g = -1$ with equal probability. Thus, the player strategy is just the interpolation of the points $(-1, -f(-1))$ and $(1, -f(1))$, where we take $f = V_{t+1}$, giving us

$$x_{t+1} = \frac{1}{2} \big(V_{t+1}(g_1, \ldots, g_t, -1) - V_{t+1}(g_1, \ldots, g_t, +1) \big). \tag{9}$$

Thus, if we can derive a closed form for $V_t(g_1, \ldots, g_t)$, we will have an efficient minimax-optimal algorithm. Note that for any function L,

$$\mathbb{E}_{G \sim \mathcal{B}_T}[L(G)] = \frac{1}{2^T} \sum_{i=0}^T \binom{T}{i} L(2i - T), \tag{10}$$

since $2^{-T} \binom{T}{i}$ is the binomial probability of getting exactly i gradients of $+1$ over T rounds, which implies $T-i$ gradients of -1, so $G = i - (T-i) = 2i - T$. Using Theorem 4, and Eqs (9) and (10), in

---

[1] However, is easy to extend this to the case where $\mathcal{G} = [a, b]$, which leads to different random variables.

the following sections we exactly compute the game values and unique minimax optimal strategies for a variety of interesting coordinate-decomposable games. Even when such exact computations are not possible, *any* coordinate-decomposable game where L depends only on $G = g_{1:T}$ can be solved numerically in polynomial time. If $\tau = T - t$, the number of rounds remaining, then we can compute V_t exactly by using the appropriate binomial probabilities (following Eq. (8) and Eq. (10)), requiring only a sum over $\mathcal{O}(\tau)$ values. If τ is large enough, then using an approximation to the binomial (e.g., the Gaussian approximation) may be sufficient.

We can also immediately provide a characterization of the potentially optimal player strategies in terms of the subgradients of $-L$. For simplicity, we write $-\partial L(g)$ instead of $\partial(-L(g))$.

Theorem 6. *Let $\mathcal{G} = [a, b]$, with $a < 0 < b$, and $L : \mathbb{R} \to \mathbb{R}$ is bounded and concave. Then, on every round, the unique minimax optimal x_t^* satisfies $-x_t^* \in \mathcal{L}$ where $\mathcal{L} = \cup_{w \in \mathbb{R}} - \partial L(w)$.*

Proof. Following Theorem 3, we know the minimax x_{t+1} interpolates $(a, -f(a))$ and $(b, -f(b))$, where we take $f(g) = V_{t+1}(g_1, \ldots, g_t, g)$. In one dimension, this implies $-x_{t+1} \in \partial f(g)$ for some $g \in \mathcal{G}$. It remains to show $\partial f(g) \subseteq \mathcal{L}$. From Theorem 1 we have $f(g) = \mathbb{E}[-L(g_{1:t} + g + B)]$, where the $\mathbb{E}$ is with respect to mean-zero random variable $B \sim \mathcal{B}_\tau$, $\tau = T - t$. For each possible value b that B can take on, $-\partial_g L(g_{1:t} + g + b_i) \subseteq \mathcal{L}$ by definition, so $\partial f(g)$ is a convex combination of these sets (e.g., Rockafellar [1997, Thm. 23.8]). The result follows as $\mathcal{L}$ is convex. $\square$

Note that for standard regret, $L(g) = \inf_{x \in \mathcal{X}} gx$, we have $\partial L(g) \subseteq \mathcal{X}$, indicating that (in 1 dimension at least), the player never needs to play outside the comparator set $\mathcal{X}$. We will see additional consequences of this theorem in the following sections.

3.1 Constant step-size gradient descent can be minimax optimal

Suppose we use a "soft" feasible set for the benchmark via a quadratic penalty,

$$L(G) = \min_x \ Gx + \frac{\sigma}{2}x^2 = -\frac{1}{2\sigma}G^2, \tag{11}$$

for a constant $\sigma > 0$. Does a no-regret algorithm against this comparison class exist? Unfortunately, the general answer is no, as shown in the next theorem. Recalling $g_t \in [-1, 1]$,

Theorem 7. *The value of this game is $V^T = \mathbb{E}_{G \sim \mathcal{B}_T} \left[\frac{1}{2\sigma} G^2 \right] = \frac{T}{2\sigma}$.*

Thus, for a fixed σ, we cannot have a no regret algorithm with respect to this L. But this does not mean the minimax algorithm will be uninteresting. To derive the minimax optimal algorithm, we compute conditional values (using similar techniques to Theorem 7),

$$V_t(g_1, \ldots, g_t) = \mathbb{E}_{G \sim \mathcal{B}_{T-t}} \left[\frac{1}{2\sigma}(g_{1:t} + G)^2 \right] = \frac{1}{2\sigma} \left((g_{1:t})^2 + (T - t) \right),$$

and so following Eq. (9) the minimax-optimal algorithm must use

$$x_{t+1} = \frac{1}{4\sigma} \left(((g_{1:t} - 1)^2 + (T - t - 1)) - ((g_{1:t} + 1)^2 + (T - t - 1)) \right) = \frac{1}{4\sigma}(-4g_{1:t}) = -\frac{1}{\sigma}g_{1:t}$$

Thus, a minimax-optimal algorithm is simply constant-learning-rate gradient descent with learning rate $\frac{1}{\sigma}$. Note that for a fixed σ, this is the optimal algorithm independent of T; this is atypical, as usually the minimax optimal algorithm depends on the horizon (as we will see in the next two cases). Note that the set $\mathcal{L} = \mathbb{R}$ (from Theorem 6), and indeed the player could eventually play an arbitrary point in $\mathbb{R}$ (given large enough T).

3.2 Non-stochastic betting with exponential upside and bounded worst-case loss

A major advantage of the regret minimization framework is that the guarantees we can achieve are typically robust to arbitrary input sequences. But on the downside the model is very pessimistic: we measure performance *in the worst case*. One might aim to perform not too badly in the worst case yet extremely well under certain conditions.

We now show how the results in the present paper can lead to a very optimistic guarantee, particularly in the case of a sequential betting game. On each round t, the world offers the player a betting opportunity on a coin toss, i.e. a binary outcome $g_t \in \{-1, 1\}$. The player may take either side of the bet, and selects a wager amount x_t, where $x_t > 0$ implies a bet on tails ($g_t = -1$) and $x_t < 0$ a bet on heads ($g_t = 1$). The world then announces whether the bet was won or lost, revealing g_t. The player's wealth changes (additively) by $-g_t x_t$ (that is, the player strives to minimize loss $g_t x_t$). We assume that the player begins with some initial capital $\alpha > 0$, and at any time period the wager $|x_t|$ must not exceed $\alpha - \sum_{s=1}^{t-1} g_s x_s$, the initial capital plus the money earned thus far.

With the benefit of hindsight, the gambler can see $G = \sum_{t=1}^{T} g_t$, the total number of heads minus the total number of heads. Let us imagine that the number of heads significantly exceeded the number of tails, or vice versa; that is, $|G|$ was much larger than 0. Without loss of generality let us assume that G is positive. Let us imagine that the gambler, with the benefit of hindsight, considers what could have happened had he always bet a constant fraction β of his wealth on heads. A simple exercise shows that his wealth would become

$$\prod_{t=1}^{T} (1 + \beta g_t) = (1 + \beta)^{\frac{T+G}{2}} (1 - \beta)^{\frac{T-G}{2}}.$$

This is optimized at $\beta = \frac{G}{T}$, which gives a simple expression in terms of KL-divergence for the maximum wealth in hindsight, $\exp\left(T \cdot \text{KL}\left(\frac{1+G/T}{2} \mid\mid \frac{1}{2}\right)\right)$, and the former is well-approximated by $\exp(O(G^2/T))$ when G is not too large relative to T. In other words, with knowledge of the final G, a naïve betting strategy could have earned the gambler *exponentially large* winnings starting with constant capital. Note that this is essentially a Kelly betting scheme [Kelly Jr, 1956], expressed in terms of G. We ask: does there exist an adaptive betting strategy that can compete with this hindsight benchmark, even if the g_t are chosen fully adversarially?

Indeed we show we can get reasonably close. Our aim will be to compete with a slightly weaker benchmark $L(G) = -\exp(|G|/\sqrt{T})$. We present a solution for the one-sided game, without the absolute value, so the player only aims for exponential wealth growth for large positive G. It is not hard to develop a two-sided algorithm as a result, which we soon discuss.

Theorem 8. *Consider the game where $\mathcal{G} = [-1, 1]$ with benchmark $L(G) = -\exp(G/\sqrt{T})$. Then*

$$V^T = \left(\cosh \tfrac{1}{\sqrt{T}}\right)^T \leq \sqrt{e}$$

with the bound tight as $T \to \infty$. Let $\tau = T - t$ and $G_t = g_{1:t}$, then the conditional value of the game is $V_t(G_t) = \left(\cosh \tfrac{1}{\sqrt{T}}\right)^\tau \exp\left(\tfrac{G_t}{\sqrt{T}}\right)$ and the player's minimax optimal strategy is:

$$x_{t+1} = -\exp\left(\frac{G_t}{\sqrt{T}}\right) \sinh \tfrac{1}{\sqrt{T}} \left(\cosh \tfrac{1}{\sqrt{T}}\right)^{\tau-1} \tag{12}$$

Recall that the value of the game can be thought of as the largest possible difference between the payoff of the benchmark function $\exp(G/\sqrt{T})$ and the winnings of the player $-\sum g_t x_t$, when the player uses an optimal betting strategy. That the value of the game here is of constant order is critical, since it says that we can always achieve a payoff that is exponential in $\frac{G}{\sqrt{T}}$ at a cost of no more than $\sqrt{e} = O(1)$. Notice we have said nothing thus far regarding the nature of our betting strategy; in particular we have not proved that the strategy satisfies the required condition that the gambler cannot bet more than α plus the earnings thus far. We now give a general result showing that this condition can be satisfied:

Theorem 9. *Consider a one dimensional game with $\mathcal{G} = [-1, 1]$ with benchmark function L non-positive on $\mathcal{G}^T$. Then for the optimal betting strategy we have that $|x_t| \leq -\sum_{s=1}^{t} g_s x_s + V^T$, and further $V^T \geq \sum_{s=1}^{t} g_s x_s$ for any t and any sequence $g_1, \ldots, g_t$.*

In other words, the player's cumulative loss at any time is always bounded from below by V^T. This implies that the starting capital α required to "replicate" the payoff function is *exactly* the value[2] of the game V^T. Indeed, to replicate $\exp(G/\sqrt{T})$ we would require no more than $\alpha = \$1.65$.

---

[2]This idea has a long history in finance and was a key tool in Abernethy et al. [2012], DeMarzo et al. [2006], and other works.

It is worth noting an alternative characterization of the benchmark function L used here. For $a \geq 0$, $\min_{x \in \mathbb{R}^-} (Gx - ax \log(-ax) + ax) = -\exp\left(\frac{G}{a}\right)$. Thus, if we take $\Psi(x) = -ax \log(ax) + ax + I(x \leq 0)$, we have $\min_{x \in \mathbb{R}^-} g_{1:T}x + \Psi(x) = -\exp\left(\frac{G}{a}\right)$. Since this algorithm needs large Reward when G is large and positive, we might expect that the minimax optimal algorithm only plays $x_t \leq 0$. Another intuition for this is that the algorithm should not need to play any point x to which Ψ assigns an infinite penalty. This intuition can be confirmed immediately via Theorem 6.

We now sketch how to derive an algorithm for the "two-sided" game. To do this, we let $L_C(G) \equiv L(G) + L(-G) \leq -\exp(|G|/\sqrt{T})$. We can construct a minimax optimal algorithm for $L_C(G)$ by running two copies of the one-sided minimax algorithm simultaneously, switching the signs of the gradients and plays of the second copy. We formalize this in Appendix B.

This same benchmark and algorithm can be used in the setting introduced by Streeter and McMahan [2012]. In that work, the goal was to prove bounds on standard regret like Regret $\leq \mathcal{O}(R\sqrt{T} \log((1+R)T))$ simultaneously for any comparator x^* with $|x^*| = R$. Stating their Theorem 1 in terms of losses, this traditional regret bound is achieved by any algorithm that guarantees

$$\text{Loss} = \sum_{t=1}^{T} g_t x_t \leq -\exp\left(\frac{|G|}{\sqrt{T}}\right) + \mathcal{O}(1). \tag{13}$$

The symmetric algorithm (Appendix B) satisfies

$$\text{Loss} \leq -\exp\left(\frac{G}{\sqrt{T}}\right) - \exp\left(\frac{-G}{\sqrt{T}}\right) + 2\sqrt{e} \leq -\exp\left(\frac{|G|}{\sqrt{T}}\right) + 2\sqrt{e},$$

and so we also achieve a standard regret bound of the form given above.

3.3 Optimal regret against hypercube adversaries

Perhaps the simplest and best studied learning games are those that restrict both the player and adversary to a norm ball, and use the standard notion of regret. We can derive results for the game where the adversary has an L_∞ constraint, the comparator set is also the L_∞ ball, and the player is unconstrained. Corollary 5 implies it is sufficient to study the one-dimensional case.

Theorem 10. *Consider the game between an adversary who chooses losses $g_t \in [-1, 1]$, and a player who chooses $x_t \in \mathbb{R}$. For a given sequence of plays, $x_1, g_1, x_2, g_2, \ldots, x_T, g_T$, the value to the adversary is $\sum_{t=1}^{T} g_t x_t - |g_{1:T}|$. Then, when T is even with $T = 2M$, the minimax value of this game is given by*

$$V_T = 2^{-T} \frac{2M\,T!}{(T-M)!M!} \leq \sqrt{\frac{2T}{\pi}}.$$

Further, as $T \to \infty$, $V_T \to \sqrt{\frac{2T}{\pi}}$. Let B be a random variable drawn from $\mathcal{B}_{T-t}$. Then the minimax optimal strategy for the player given the adversary has played $G_t = g_{1:t}$ is given by

$$x_{t+1} = \Pr(B < -G_t) - \Pr(B > -G_t) = 1 - 2\Pr(B > -G_t) \in [-1, 1]. \tag{14}$$

The fact that the limiting value of this game is $\sqrt{2T/\pi}$ was previously known, e.g., see a mention in Abernethy et al. [2009]; however, we believe this explicit form for the optimal player strategy is new. This strategy can be efficiently computed numerically, e.g, by using the regularized incomplete beta function for the CDF of the binomial distribution. It also follows from this expression that even though we allow the player to select $x_{t+1} \in \mathbb{R}$, the minimax optimal algorithm always selects points from $[-1, 1]$, so our result applies to the case where the player is constrained to play from $\mathcal{X}$.

Abernethy et al. [2008a] shows that for the linear game with $n \geq 3$ where both the learner and adversary select vectors from the unit sphere, the minimax value is exactly $\sqrt{T}$. Interestingly, in the $n = 1$ case (where L_2 and L_∞ coincide), the value of the game is lower, about $0.8\sqrt{T}$ rather than $\sqrt{T}$. This indicates a fundamental difference in the geometry of the $n = 1$ space and $n \geq 3$. We conjecture the minimax value for the L_2 game with $n = 2$ lies somewhere in between.

References

Jacob Abernethy and Manfred K. Warmuth. Repeated games against budgeted adversaries. In *NIPS*, 2010.

Jacob Abernethy, Peter L. Bartlett, Alexander Rakhlin, and Ambuj Tewari. Optimal strategies and minimax lower bounds for online convex games. In *COLT*, 2008a.

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, 2008b.

Jacob Abernethy, Alekh Agarwal, Peter Bartlett, and Alexander Rakhlin. A stochastic view of optimal regret through minimax duality. In *COLT*, 2009.

Jacob Abernethy, Rafael M. Frongillo, and Andre Wibisono. Minimax option pricing meets black-scholes in the limit. In *STOC*, 2012.

Amit Agarwal, Elad Hazan, Satyen Kale, and Robert E. Schapire. Algorithms for portfolio management based on the Newton method. In *ICML*, 2006.

Nicolò Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

A. de Moivre. *The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play*. 1718.

Ofer Dekel, Ambuj Tewari, and Raman Arora. Online bandit learning against an adaptive adversary: from regret to policy regret. In *ICML*, 2012.

Peter DeMarzo, Ilan Kremer, and Yishay Mansour. Online trading algorithms and robust option pricing. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 477–486. ACM, 2006.

Persi Diaconis and Sandy Zabell. Closed form summation for classical distributions: Variations on a theme of de Moivre. *Statistical Science*, 6(3), 1991.

Elad Hazan and Satyen Kale. On stochastic and worst-case models for investing. In *NIPS*. 2009.

J. L. Kelly Jr. A new interpretation of information rate. *Bell System Technical Journal*, 1956.

Wouter Koolen, Dmitry Adamskiy, and Manfred Warmuth. Putting bayes to sleep. In *NIPS*. 2012.

N. Merhav, E. Ordentlich, G. Seroussi, and M. J. Weinberger. On sequential strategies for loss functions with memory. *IEEE Trans. Inf. Theor.*, 48(7), September 2006.

Alexander Rakhlin, Ohad Shamir, and Karthik Sridharan. Relax and randomize: From value to algorithms. In *NIPS*, 2012.

Ralph T. Rockafellar. *Convex Analysis (Princeton Landmarks in Mathematics and Physics)*. Princeton University Press, 1997.

Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 2012.

Shai Shalev-Shwartz, Yoram Singer, Nathan Srebro, and Andrew Cotter. Pegasos: Primal estimated sub-gradient solver for svm. *Mathematical Programming*, 127(1):3–30, 2011.

Gilles Stoltz. *Contributions to the sequential prediction of arbitrary sequences: applications to the theory of repeated games and empirical studies of the performance of the aggregation of experts*. Habilitation à diriger des recherches, Université Paris-Sud, 2011.

Matthew Streeter and H. Brendan McMahan. No-regret algorithms for unconstrained online convex optimization. In *NIPS*, 2012.

Volodya Vovk. Competitive on-line statistics. *International Statistical Review*, 69, 2001.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *ICML*, 2003.

A Proofs

We restate the results proved here for convenience.

A.1 Proof of Theorem 3

Theorem 1. *For every t and every sequence $g_1, \ldots, g_t \in \mathcal{G}$, we can write the conditional value of the game as*

$$V_t(g_1, \ldots, g_t) = \max_{G \in \Delta(\mathcal{G}'), \mathbb{E}[G]=0} \mathbb{E}[V_{t+1}(g_1, \ldots, g_t, G)],$$

where $\Delta(\mathcal{G}')$ is the set of random variables on $\mathcal{G}'$. Moreover, for all t the function V_t is convex in each of its coordinates and bounded.

Proof. We prove both statements simultaneously via induction on t. For the base case, $t = T - 1$, we have

$$V_{T-1}(g_1, \ldots, g_{T-1}) = \inf_{x_T} \sup_{g_T} g_T \cdot x_T - L(g_1, \ldots, g_{T-1}, g_T).$$

Because the supremum is taken over $\mathcal{G}$ whose convex hull is assumed to be a polytope, we can replace the $\sup_{g_T \in \mathcal{G}}$ with $\max_{g_T \in \mathcal{G}'}$. Furthermore, we can replace the maximization over points from $\mathcal{G}'$ with the maximization of distributions over $\mathcal{G}' = \{g^i\}_{i=1,\ldots,m}$. That is, we can write

$$V_{T-1}(g_1, \ldots, g_{T-1}) = \inf_{x_T} \max_{\vec{\alpha} \in \Delta_m} \sum_{i=1}^m \alpha_i(g^i \cdot x_T - L(g_1, \ldots, g_{T-1}, g^i)).$$

The set Δ_m is a compact convex set, and the objective $\sum_{i=1}^m \alpha_i(g^i \cdot x_T - L(g_1, \ldots, g_{T-1}, g^i))$ is linear in both x and $\vec{\alpha}$ hence we can apply Sion's Minimax theorem to obtain

$$V_{T-1}(g_1, \ldots, g_{T-1}) = \max_{\vec{\alpha} \in \Delta_m} \inf_{x_T} \left(\sum_i \alpha_i g^i \right) \cdot x_T - \sum_i \alpha_i L(g_1, \ldots, g_{T-1}, g^i).$$

Notice that if $\sum_i \alpha_i g^i \neq \mathbf{0}$ then the infimum is $-\infty$ since the player can make the objective arbitrarily small. Hence we can restrict the outer maximization to distributions $\vec{\alpha}$ such that $\sum_i \alpha_i g^i = \mathbf{0}$. This simplifies the expression to

$$V_{T-1}(g_1, \ldots, g_{T-1}) = \max_{\vec{\alpha} \in \Delta_m} -\sum_i \alpha_i L(g_1, \ldots, g_{T-1}, g^i) \qquad \text{s.t.} \quad \sum_i \alpha_i g^i = \mathbf{0}.$$

Notice that, by assumption, $-L$ is convex in each of its arguments, and hence $V_{T-1}(g_1, \ldots, g_{T-1})$ is also convex in each g_t independently, since the maximum of convex functions is convex.

The inductive argument follows identically to the base case, but where we replace $-L$ with V_t, since we can write

$$V_{t-1}(g_1, \ldots, g_{t-1}) = \inf_{x_t} \sup_{g_t \in \mathcal{G}} g_t \cdot x_t + V_t(g_1, \ldots, g_{t-1}, g_t).$$

$\square$

Theorem 3. *There exists a set of $n + 1$ distinct points $\{g^1, \ldots, g^{n+1}\} \subset \mathcal{G}$ whose convex hull is of full rank, and a distribution $\vec{\alpha} \in \Delta_{n+1}$ satisfying $\sum_{i=1}^{n+1} \alpha_i g^i = \mathbf{0}$, such that $V = \sum_{i=1}^{n+1} \alpha_i f(g^i)$. Moreover, an optimal choice for the infimum in (6) is the gradient of the unique linear interpolation of the pairs $\{(g^1, -f(g^1)), \ldots, (g^{n+1}, -f(g^{n+1}))\}$.*

We prove this theorem via a sequence of lemmas. We begin with the observation that we may assume, without loss of generality, that $\mathcal{G}$ is convex, and hence $\mathcal{G} = \text{ConvexHull}(\mathcal{G})$. This is because, for any x, the objective $\sup_{g \in \mathcal{G}} x \cdot g + f(g)$ will always be achieved at the boundary of $\mathcal{G}$ since the objective function $x \cdot g + f(g)$ is the sum of two convex functions and is thus convex.

Lemma 11. *The infimum in (6) is achieved in a bounded set.*

10

Proof. Let $M = \sup_{g \in \mathcal{G}} |f(g)|$ then clearly we have that $\inf_{x \in \mathbb{R}^n} \sup_{g \in \mathcal{G}} x \cdot g + f(g) \leq M$ since x can be chosen as $\mathbf{0}$. It is sufficient to show any x such that $\|x\| > 2M/\epsilon$ achieves a worse value than $\mathbf{0}$. Since $\mathbf{0}$ is in the interior of $\mathcal{G}$, there exists an $\epsilon > 0$ such that $g = \frac{\epsilon x}{\|x\|} \in \mathcal{G}$. Then, $\sup_{g \in \mathcal{G}} x \cdot g + f(g) \geq x \cdot \frac{\epsilon x}{\|x\|} + f(g) > 2M - M = M$. $\qquad \square$

The above lemma is useful since it lets us conclude that we need not necessarily assume x is unbounded. Moreover, since the $\inf$ is achieved on a compact set, then it has at least one solution x^* that we can analyze. Let $\Phi \subset \mathbb{R}^n$ denote the set of points x on which the infimum in (6) is achieved. For any x, let $\Gamma(x) \subset \mathcal{G}$ be the set of corners of the polytope $\mathcal{G}$ on which the supremum $\sup_{g \in \mathcal{G}} x \cdot g + f(g)$ is achieved for fixed x.

Lemma 12. *For any $x \in \Phi$, the set* $\mathrm{ConvexHull}(\Gamma(x))$ *must contain the origin.*

Proof. Let us assume that $\mathbf{0} \notin \mathrm{ConvexHull}(\Gamma(x))$, then I will show that this contradicts the assumption that x is optimal. If v is the value of the objective in (6), then define $\Gamma_\epsilon(x)$ to be the set of $g \in \mathcal{G}$ such that $g \cdot x + f(x) \geq v - \epsilon$. We claim that we can choose $\epsilon > 0$ small enough so that $\mathrm{ConvexHull}(\Gamma_\epsilon(x))$ also does not contain $\mathbf{0}$. This implies that there is some $\delta > 0$ such that $\|g\| > \delta$ for all $g \in \mathrm{ConvexHull}(\Gamma_\epsilon(x))$. Moreover, since $\mathrm{ConvexHull}(\Gamma_\epsilon(x))$ is a convex set there must be a separating hyperplane between $\mathbf{0}$ and $\mathrm{ConvexHull}(\Gamma_\epsilon(x))$, and hence there is some unit vector $z \in \mathbb{R}^n$ (the normal to the hyperplane) such that $z \cdot g < -\delta$ for all $g \in \mathrm{ConvexHull}(\Gamma_\epsilon(x))$.

Choose $B > 0$ such that $\|g\| \leq B$ for all $g \in \mathcal{G}$. We claim that the point $x' \equiv x + \frac{\epsilon}{2B} z$ has a strictly smaller objective value that x. Consider any $g \in \mathrm{ConvexHull}(\Gamma_\epsilon(x))$, then we have

$$ g \cdot x' + f(g) = g \cdot x + f(g) + \frac{\epsilon}{2B} z \cdot g < v - \frac{\epsilon \delta}{2B} < v. $$

On the other hand, for any $g \in \mathcal{G} \setminus \mathrm{ConvexHull}(\Gamma_\epsilon(x))$ we have

$$ g \cdot x' + f(g) = g \cdot x + f(g) + \frac{\epsilon}{2B} z < v - \epsilon + \frac{\epsilon}{2B} z \cdot g \leq v - \epsilon + \frac{\epsilon}{2B} \|g\| \leq v - \frac{\epsilon}{2} < v $$

where the first inequality follows because by assumption $g \notin \Gamma_\epsilon(x)$. It follows from these two expressions that $\sup_{g \in G} g \cdot x' + f(g) < \sup_{g \in G} g \cdot x + f(g)$, a contradiction. $\qquad \square$

Concluding that $\mathrm{ConvexHull}(H)$ contains the origin is actually surprisingly useful.

Lemma 13. *There is some $x \in \Phi$ such that* $\mathrm{ConvexHull}(\Gamma(x))$ *has a non-empty interior.*

Another way to put this is that $\Gamma(x)$ has at least $n + 1$ points such that none of these is a convex combination of the others.

Proof. Notice that Φ is a convex set and, via Lemma 11, is bounded and compact. We claim that any x on the boundary of Φ satisfies the goal of the lemma. Choose a boundary point $x \in \Phi$, and assume that $\mathrm{ConvexHull}(\Gamma(x))$ is not of full-rank. Via Lemma 12, this set contains the origin, and hence we can find some unit vector z such that $z \cdot g = 0$ for all $g \in \mathrm{ConvexHull}(\Gamma(x))$.

Since $\mathcal{G}$ is a polytope, we can describe it as the hull of a finite number of points $\mathcal{G}' \equiv \{g^1, \ldots, g^m\}$. For any $g^i \notin \Gamma(x)$ we have $g^i \cdot x + f(g^i) < v$. Choose some $\epsilon > 0$ so that $g^i \cdot x + f(g^i) < v - \epsilon$ for every $g^i \in \mathcal{G}' \setminus \Gamma(x)$, which is possible since this is a finite set. Let $B > 0$ be a bound on the norm of all points in $\mathcal{G}$. Then we claim that the points $x + \frac{\epsilon}{2B} z$ and $x - \frac{\epsilon}{2B} z$ are both members of Φ. Of course, the latter statement contradicts the assumption that x is at the boundary of Φ. To prove the final claim, notice that by the convexity of f we have

$$ \sup_{g \in \mathcal{G}} g \cdot \left(x + \frac{\epsilon}{2B} z \right) + f(g) = \max_{i = 1, \ldots, m} g^i \cdot \left(x + \frac{\epsilon}{2B} z \right) + f(g^i). $$

For the last expression, we can check two cases. If $g^i \in \Gamma(x)$ then $g^i \cdot z = 0$ in which case $g^i \cdot \left(x + \frac{\epsilon}{2B} z \right) + f(g^i) = g^i \cdot x + f(g^i)$. On the other hand, for $g^i \notin \Gamma(x)$ we have

$$ g^i \cdot \left(x + \frac{\epsilon}{2B} z \right) + f(g^i) = g^i \cdot x + f(g^i) = \frac{\epsilon}{2B} g \cdot z < v - \epsilon + \epsilon/2 < v. $$

Hence the value of the objective is the same for x and $x + \frac{\epsilon}{2B} z$. A similar argument follows for $x - \frac{\epsilon}{2B} z$. $\qquad \square$

Lemma 14. *If $x \in \Phi$ and we pick any full-rank set of points $g_1, \ldots, g_{n+1} \in \Gamma(x)$ whose hull contains the origin, then we may write x as the gradient of the linear interpolation of the points $\{(g_1, -f(g_1)), \ldots, (g_{n+1}, -f(g_{n+1}))\}$. Moreover, this implies that x is a subgradient of the function f restricted to the set $\mathcal{G}$.*

Proof. Let us notice that if we were to search for the linear interpolation of the points $\{(g_1, -f(g_1)), \ldots, (g_{n+1}, -f(g_{n+1}))\}$, then we would need to find a vector $m \in \mathbb{R}^n$ and an offset $b \in \mathbb{R}$ such that
$$m \cdot g_i + b = -f(g_i) \qquad \forall\, i = 1, \ldots, n+1,$$
and indeed since the set of g_i's is of full rank this has a unique solution. However, the point x also satisfies a similar set of equations:
$$x \cdot g_i + f(g_i) = c \qquad \forall\, i = 1, \ldots, n+1,$$
where c is the value of the objective in (6). Given the uniqueness of the above to systems of equations, we have that $m = x$. □

Now given the above results we can actually construct the optimal strategy for the adversary.

Lemma 15. *For any full-rank set of points $g_1, \ldots, g_{n+1} \in \Gamma(x)$ whose hull contains the origin, let $\vec{\alpha} \in \Delta_{n+1}$ be a set of weights such that $\sum_i \alpha_i g_i = \mathbf{0}$ (and indeed $\vec{\alpha}$ is unique). Then the value of the objective (6) is precisely $\sum_i \alpha_i f(g_i)$. Moreover, one optimal randomized strategy for the adversary is to choose g_i with probability α_i.*

Proof. Recall that the point x^* satisfies a system of linear equations
$$x^* \cdot g_i + f(g_i) = c \qquad \forall\, i = 1, \ldots, n+1,$$
where c is the value of the objective. Furthermore, it also satisfies any *mixture* of these equations. By taking an $\vec{\alpha}$ mixture of these equations we have
$$c = \sum_i \alpha_i (g_i \cdot x^* + f(g_i)) = \mathbf{0} \cdot x^* + \sum_i \alpha_i f(g_i) = \sum_i \alpha_i f(g_i).$$
□

A.2 Proofs from Section 3

Theorem 7. *The value of this game is $V^T = \mathbb{E}_{G \sim \mathcal{B}_T}\left[\frac{1}{2\sigma} G^2\right] = \frac{T}{2\sigma}$.*

Proof. Starting from Eq. (10),

$$\mathbb{E}_{G \sim \mathcal{B}_T}[G^2] = \frac{1}{2^T} \sum_{i=0}^{T} \binom{T}{i}(2i - T)^2 \qquad\qquad Eq.\ (10)$$

$$= \frac{1}{2^T}\left(4\sum_{i=0}^{T}\binom{T}{i}i^2 - 4T\sum_{i=0}^{T}\binom{T}{i}i + T^2\sum_{i=0}^{T}\binom{T}{i}\right)$$

and since $\sum_{t=0}^{T}\binom{T}{t} = 2^T$, $\sum_{t=0}^{T}\binom{T}{t}t = T2^{T-1}$, $\sum_{t=0}^{T}\binom{T}{t}t^2 = (T+T^2)2^{T-2}$,

$$= \frac{1}{2^T}\left(4(T+T^2)2^{T-2} - 4T(T2^{T-1}) + T^2 2^T\right)$$
$$= (T+T^2) - 2T^2 + T^2 = T.$$

The result then follows from linearity of expectation. □

Theorem 8. *Consider the game where $\mathcal{G} = [-1, 1]$ with benchmark $L(G) = -\exp(G/\sqrt{T})$. Then*
$$V^T = \left(\cosh \tfrac{1}{\sqrt{T}}\right)^T \le \sqrt{e}$$

with the bound tight as $T \to \infty$. Let $\tau = T - t$ and $G_t = g_{1:t}$, then the conditional value of the game is $V_t(G_t) = \left(\cosh \frac{1}{\sqrt{T}}\right)^{\tau} \exp\left(\frac{G_t}{\sqrt{T}}\right)$ and the player's minimax optimal strategy is:

$$x_{t+1} = -\exp\left(\frac{G_t}{\sqrt{T}}\right) \sinh \frac{1}{\sqrt{T}} \left(\cosh \frac{1}{\sqrt{T}}\right)^{\tau-1} \tag{12}$$

Proof. First, we compute the value of the game:

$$V^T = \mathop{\mathbb{E}}_{G \sim \mathcal{B}_T} \left[-L(G)\right] = 2^{-T} \sum_{i=0}^{T} \binom{T}{i} \exp\left(\frac{2i - T}{\sqrt{T}}\right)$$

$$= 2^{-T} \exp\left(-\sqrt{T}\right) \sum_{i=0}^{T} \binom{T}{i} \left(\exp\left(2/\sqrt{T}\right)\right)^i$$

$$= 2^{-T} \exp\left(-\sqrt{T}\right) \left(1 + \exp\left(2/\sqrt{T}\right)\right)^T,$$

where we have used the ordinary generating function, $\sum_{i=0}^{T} \binom{T}{i} x^i = (1+x)^T$. Manipulating the above expression for the value of the game, we arrive at $V^T = \cosh(1/\sqrt{T})^T$. Using the series expansion for $\cosh$ leads to the upper bound $\cosh(x) \le \exp(x^2/2)$,

from which we conclude

$$V_T = \left(\cosh\left(1/\sqrt{T}\right)\right)^T \le \exp\left(\frac{1}{2T}\right)^T = \sqrt{e}.$$

Using similar techniques, we can derive the conditional value of the game, letting $\tau = T - t$ be the number of rounds left to be played:

$$V_t(G_t) = 2^{-\tau} \sum_{i=0}^{\tau} \binom{\tau}{i} \exp\left(\frac{G_t + 2i - \tau}{\sqrt{T}}\right) = 2^{-\tau} \exp\left(\frac{G_t - \tau}{\sqrt{T}}\right) \left(1 + \exp\left(2/\sqrt{T}\right)\right)^{\tau}.$$

Following Eq. (9) and simplifying leads to the update of Eq. (12). It remains to show $\lim_{T \to \infty} V_T = \sqrt{e}$. Using the change of variable $x = 1/\sqrt{T}$, equivalently we have $\lim_{x \to 0} \cosh(x)^{\frac{1}{x^2}}$. Examining the log of this function,

$$\lim_{x \to 0} \log\left(\cosh(x)^{\frac{1}{x^2}}\right) = \lim_{x \to 0} \frac{1}{x^2} \log \cosh(x) = \lim_{x \to 0} \frac{1}{x^2} \left(\frac{x^2}{2} - \frac{x^4}{12} + \frac{x^6}{45} - \frac{17x^8}{2520} + \dots\right) = \frac{1}{2},$$

where we have taken the Maclaurin series of $\log \cosh(x)$. Using the continuity of $\exp$, we have against any adversary,

$$\lim_{x \to 0} \left(\cosh(x)^{\frac{1}{x^2}}\right) = \exp\left(\lim_{x \to 0} \log\left(\cosh(x)^{\frac{1}{x^2}}\right)\right) = \sqrt{e}.$$

$\square$

Theorem 9. *Consider a one dimensional game with $\mathcal{G} = [-1, 1]$ with benchmark function L non-positive on $\mathcal{G}^T$. Then for the optimal betting strategy we have that $|x_t| \le -\sum_{s=1}^{t} g_s x_s + V^T$, and further $V^T \ge \sum_{s=1}^{t} g_s x_s$ for any t and any sequence $g_1, \dots, g_t$.*

Proof. We need to prove

$$\sum_{s=1}^{t} g_s x_s \le V^T \tag{15}$$

and

$$|x_t| \le -\sum_{s=1}^{t} g_s x_s + V^T. \tag{16}$$

13

The definition of the value of the game and the fact the algorithm is minimax optimal ensures

$$\sum_{t=1}^{T} g_t x_t - L(G) \leq V^T$$

or, since $-L(G) \geq 0$,

$$\sum_{t=1}^{T} g_t x_t \leq V^T. \tag{17}$$

Now, suppose on some round t we have $\sum_{s=1}^{t} g_s x_s > V^T$. Then, the adversary can simply play $g_\tau = 0$ for rounds $t+1, \ldots, T$, which implies

$$\sum_{s=1}^{T} g_s x_s = \sum_{s=1}^{t} g_s x_s > V^T,$$

contradicting Eq. (17). Hence, Eq. (15) must hold. Further, if the player ever chose a bet so large it violated Eq. (16), the adversary could choose $g_t \in \{-1, 1\}$ in order to violate Eq. (17). $\qquad \square$

Theorem 10. *Consider the game between an adversary who chooses losses $g_t \in [-1, 1]$, and a player who chooses $x_t \in \mathbb{R}$. For a given sequence of plays, $x_1, g_1, x_2, g_2, \ldots, x_T, g_T$, the value to the adversary is $\sum_{t=1}^{T} g_t x_t - |g_{1:T}|$. Then, when T is even with $T = 2M$, the minimax value of this game is given by*

$$V_T = 2^{-T} \frac{2M \, T!}{(T-M)!M!} \leq \sqrt{\frac{2T}{\pi}}.$$

Further, as $T \to \infty$, $V_T \to \sqrt{\frac{2T}{\pi}}$. Let B be a random variable drawn from $\mathcal{B}_{T-t}$. Then the minimax optimal strategy for the player given the adversary has played $G_t = g_{1:t}$ is given by

$$x_{t+1} = \Pr(B < -G_t) - \Pr(B > -G_t) = 1 - 2\Pr(B > -G_t) \in [-1, 1]. \tag{14}$$

Proof. Letting $T = 2M$ and working from Eq. (10),

$$V^T = -\mathop{\mathbb{E}}_{G \sim \mathcal{B}_T}[L(G)] = \frac{2}{2^T} \sum_{i=0}^{T} \binom{T}{i} |i - M| = \frac{2M}{2^T} \binom{2M}{M} = 2^{-T} \frac{2M \, T!}{(T-M)!M!}, \tag{18}$$

where we have applied a classic formula of de Moivre [1718] for the mean absolute deviation of the binomial distribution (see also Diaconis and Zabell [1991]). Using a standard bound on the central binomial coefficient (based on Stirling's formula),

$$\binom{2M}{M} = \frac{4^M}{\sqrt{\pi M}} \left(1 - \frac{c_M}{M}\right) \tag{19}$$

where $\frac{1}{9} < c_M < \frac{1}{8}$ for all $M \geq 1$, we have

$$V^T \leq 2M \frac{1}{\sqrt{\pi M}} = \sqrt{\frac{2T}{\pi}}.$$

As implied by Eq. (19), this inequality quickly becomes tight as $T \to \infty$.

In order to compute the minimax algorithm, we would like a closed form for $V_t(G_t) = -\mathbb{E}_{G^\tau \sim \mathcal{B}_\tau} \left[L(G_t + G^\tau) \right]$, where $G_t = g_{1:t}$ is the sum of the gradients so far, $\tau = T - t$ is the number of rounds to go, and and $G^\tau = g_{t+1:T}$ is a random variable giving the sum of the remaining gradients. Unfortunately, the structure of the binomial coefficients exploited in the proof of Theorem 10 does not apply given an arbitrary offset G_t. Nevertheless, we will be able to derive a formula for the update that is readily computable. Letting B be a random variable with distribution $\mathcal{B}_\tau$, the update of Eq. (9) becomes

$$x_{t+1} = \frac{1}{2} \sum_{b=-\tau}^{\tau} \Pr(B = b) \Big(|G_t + b - 1| - |G_t + b + 1| \Big).$$

Whenever $G_t + b \geq 1$, the difference in absolute values is -2, and whenever $G_t + b \leq 1$, the difference is 2. When $G_t + b = 0$, the difference is zero. Thus,

$$x_{t+1} = \frac{1}{2} \left(\Pr(B > -G_t)(-2) + \Pr(B < -G_t)(2) \right) = \Pr(B < -G_t) - \Pr(B > -G_t).$$

$\qquad \square$

B A Symmetric Betting Algorithm

The one-sided algorithm of Theorem 8 has

$$\text{Loss} = V^T + L(G) \leq -\exp\left(\frac{G}{\sqrt{T}}\right) + \sqrt{e}.$$

In order to do well when $g_{1:T}$ is large and negative, we can run a copy of the algorithm on $-g_1, \ldots, -g_T$, switching the signs of each x_t it suggests. The combined algorithm then satisfies

$$\text{Loss} \leq -\exp\left(\frac{G}{\sqrt{T}}\right) - \exp\left(\frac{-G}{\sqrt{T}}\right) + 2\sqrt{e}$$

$$\leq -\exp\left(\frac{|G|}{\sqrt{T}}\right) + 2\sqrt{e},$$

and so following Eq. (13) and Theorem 1 of Streeter and McMahan [2012], we obtain the desired regret bounds. The following theorem implies the symmetric algorithm is in fact minimax optimal with respect to the combined benchmark

$$L_C(G) = -\exp\left(\frac{G}{\sqrt{T}}\right) - \exp\left(\frac{-G}{\sqrt{T}}\right).$$

Theorem 16. *Consider two 1-D games where the adversary plays from $[-1, 1]$, defined by concave functions L_1 and L_2 respectively. Let x_t^1 and x_t^2 be minimax-optimal plays for L_1 and L_2 respectively, given that $g_1, \ldots g_{t-1}$ have been played so far in both games. Then $x_1 + x_2$ is also minimax optimal for the combined game that uses the benchmark $L_C(G) = L_1(G) + L_2(G)$.*

Proof. First, taking $\tau = T - t$ and using Theorem 4 three times, we have

$$V^C(g_1, \ldots, g_t) = - \underset{G^\tau \sim \mathcal{B}_\tau}{\mathbb{E}} \left[L_1(g_{1:t} + G^\tau) + L_2(g_{1:t} + G^\tau)\right]$$

$$= - \underset{G^\tau \sim \mathcal{B}_\tau}{\mathbb{E}} \left[L_1(g_{1:t} + G^\tau)\right] - \underset{G^\tau \sim \mathcal{B}_\tau}{\mathbb{E}} \left[L_2(g_{1:t} + G^\tau)\right]$$

$$= V^1(g_1, \ldots, g_t) + V^2(g_1, \ldots, g_t),$$

using linearity of expectation. Then, using Eq. (9) for each of the three games, we have

$$x_t^C = \arg\min_x \max_g gx + V_C(g_1, \ldots, g_{t-1}, g)$$

$$= \frac{1}{2}\big(V_C(g_1, \ldots, g_{t-1}, -1) - V_C(g_1, \ldots, g_{t-1}, +1)\big)$$

$$= \frac{1}{2}\big(V_1(g_1, \ldots, g_{t-1}, -1) + V_2(g_1, \ldots, g_{t-1}, -1) - V_1(g_1, \ldots, g_{t-1,1}) - V_2(g_1, \ldots, g_{t-1}, +1)\big)$$

$$= x_t^1 + x_t^2.$$

$\square$