

# Repeated Games against Budgeted Adversaries

Jacob Abernethy\*  
Division of Computer Science  
UC Berkeley  
jake@cs.berkeley.edu

Manfred K. Warmuth†  
Department of Computer Science  
UC Santa Cruz  
manfred@cse.ucsc.edu

## Abstract

We study repeated zero-sum games against an adversary on a budget. Given that an adversary is constrained by the amount he can play each action, we consider what ought to be the player's best mixed strategy with knowledge of this budget. We show that, for a general class of normal-form games, the minimax strategy is indeed efficiently computable and relies on a simple random walk.

## 1 Introduction

Let us consider the following two-player zero-sum guessing game between an Adversary and a Player. The Adversary picks some number of \$1 coins from his pocket, say between 1 and 4, and the Player must guess how many coins are held. If the Player guesses the number correctly he keeps them; if he guesses too high then he must pay \$1 to the Adversary; if the guess was too low then nothing is won or lost.

Classical results in game theory immediately give us an efficient way to solve for the optimal strategy pair for this particular guessing game. We may write our payoff matrix as

$$M = \begin{pmatrix} 1 & -1 & -1 & -1 \\ 0 & 2 & -1 & -1 \\ 0 & 0 & 3 & -1 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

then we can solve for the Player's and Adversary's optimal mixed strategies  $\mathbf{p}^*$  and  $\mathbf{a}^*$  with the following linear programs:

$$\begin{array}{ll} \max_{\mathbf{p} \in \Delta_4, c} & c \\ \mathbf{p}^\top M \mathbf{e}_i \geq c \forall i & \end{array} \quad \begin{array}{ll} \min_{\mathbf{a} \in \Delta_r, d} & d, \\ \mathbf{e}_j M \mathbf{a} \leq d \forall j & \end{array}$$

where  $\Delta_4$  is the 4-simplex. In our example problem we find that the Player may as well guess uniformly amongst his choices, i.e.  $\mathbf{p}^* = \langle 1/4, 1/4, 1/4, 1/4 \rangle$ , and the Adversary will choose coins with probabilities  $\mathbf{a}^* = \langle 5/8, 5/24, 5/48, 1/16 \rangle$ .

Now consider playing the same game but in a repeated fashion. For each of a sequence of  $T$  rounds, the Adversary reaches into his pocket for some coins, the Player guesses, and the same rewards and costs are charged. What is the optimal strategy for each player? It is easy to see that each round may be treated in isolation, as the behavior on one round need not affect that on another. That is, this repeated game can simply be treated as  $T$  identical copies of the original.

---

\*Supported by DARPA grant FA8750-05-2-0249 and NSF grant DMS-0707060.

†Supported by NSF grant IIS 0325363.

We now make the problem significantly more difficult. Consider the same repeated game, but where the number of rounds *depends on the choices of the Adversary*. In other words, imagine that the Adversary is on a budget, and the repeated game will end as soon as the budget has been spent. As an example, we may know a priori how many coins in total the Adversary has in his pocket, and that the guessing game will end as soon as all coins have been selected once. Alternatively, the Adversary may simply stop playing once each action has been played at least a fixed number of times. The Player, aware of his opponent’s budget, can now adjust his strategy to exploit this constraint.

At first glance, the optimal strategy against a budgeted Adversary appears difficult: the game value can be computed recursively at every possible “state” using a linear program and then propagated backwards, but this leads to the standard difficulties associated with extensive form games, namely that the number of states is typically exponential.

In the present paper, we show that one can indeed efficiently estimate the minimax Player strategy for a certain class of zero-sum games, that we call *inverse-nonnegative games*. This class includes the coin-guessing game described above, several experts/decision games, and many other natural games. The results apply to essentially any budget scheme that can be described as a “stopping criterion” on the state space, and where the stopping criterion is “monotonic”. The optimal Player strategy relies simply on “simulating” the Adversary’s strategy by running a random walk through the state space until the budget is spent. The minimax value of the game (the total cost of the optimal Player against a worst-case Adversary) has an appealing interpretation: the expected length of this random walk.

## 1.1 Budgets and Learning

The concept of a budgeted adversary is not simply of interest in game theory, it has significant applications in many learning and decision problems. A natural model of learning involves an agent that makes a sequence of decisions and, on each round, the environment reveals an outcome, presenting the agent with some feedback and associated cost of his decision. Such a setting is often modeled probabilistically where, typically, the goal of the learner is to estimate the parameters of the process producing the outcome. But this approach has its limitations: finding an appropriate model can be a challenge, and the process itself may not even be i.i.d. An alternative framework that has gained popularity in recent years is to imagine that the outcome sequence is chosen by some arbitrary process but, at the same time, the sequence won’t be chosen “too adversarially”. This notion of an environment that is “arbitrary but not unduly adversarial” might well be reinterpreted as “adversarial yet budgeted”. It is natural to imagine that the world, or environment, may intentionally inflict a “bad” outcome on a particular round of the game, yet we don’t expect this to occur too many times.

This semi-adversarial “world view” is perhaps most salient in the setting of learning with experts, described as follows. On each of a sequence of rounds, the learner chooses a distribution on the experts, each expert reveals some  $[0, 1]$  loss value, and the learner suffers the expected loss [6, 7, 5]. In such a setting, the worst case outcome might be that each expert suffers loss 1 on each round, leading to loss 1 for the learner as well. From the perspective of decision-making, however, this is an uninteresting scenario as the learner gains nothing by making better decisions. Instead, what is generally assumed is that, while the losses may be chosen badly, there will be at least one “good expert” who will suffer but a small amount of loss. This is exactly a budgeted adversary problem: given a bound  $k$  on the number of losses of the best expert, losses are chosen repeatedly until the event that the best expert has exactly  $k + 1$ , at which point the game stops.

The minimax solution to this experts game was solved recently [1]. It is indeed directly a special case of the present work, in which the payoff matrix  $M$  is the identity and the budget/stopping criterion is exactly that mentioned above, when all “experts” incur more than  $k$  losses. The results herein are significantly

more general, as we allow a rich class of payoff matrices, arbitrary budget constraints. Notably, our results generalize to the case the budget is unknown but chosen from a known prior.

There is some recent related work [3, 4] that employs random walks in continuous time. This requires considerably more machinery such as Wiener processes. Here we characterize the minimax optimal solution using a discrete random walk and prove that this solution can be efficiently approximated.

## 2 Setting

### 2.1 Repeated Games with Budgets

We now define the problem setting at hand. We consider a repeated two player game between the Player and the Adversary. Each round of the game is identical: the Player chooses some distribution over actions  $\mathbf{w} \in \Delta_n$ , where  $\Delta_n$  is the  $n$ -simplex, the Adversary chooses some response  $i \in [n]$ , and the Player suffers the expected cost  $\mathbf{w}^\top M \mathbf{e}_i$ , where  $M$  is the  $n \times n$  payoff matrix defined in advance.

After a number of rounds, we can define the *state* of the repeated game as follows. Given that the Adversary has played a sequence of actions  $i_1, i_2, \dots, i_t$ , we define the state on round  $t$  as  $\mathbf{s}_t := \sum_{j=1}^t \mathbf{e}_{i_j}$ . The *state space* is  $\mathbb{N}_0^n$ , i.e. all tuples of non-negative integers. The game will be played repeatedly until the Adversary exceeds his budget, in which case the repeated game stops. We assume that the budget depends entirely on the state vector  $\mathbf{s}_t$ ; that is, when some condition  $\delta(\mathbf{s}_t)$  is met, the game ends, where  $\delta(\cdot)$  is some binary-valued *stopping criterion*:  $\delta(\mathbf{s}_t) = 0$  when the state  $\mathbf{s}_t$  is over-budget, and  $\delta(\mathbf{s}_t) = 1$  when the budget constraint has not yet been met. We require simply that  $\delta(\cdot)$  is *monotone* in the sense that if  $\delta(\mathbf{s}) = 0$ , then for any  $\mathbf{s}' \geq \mathbf{s}$ ,  $\delta(\mathbf{s}') = 0$  as well. This is a natural assumption and suggests that, once the budget is spent, more playing can not bring it back.

The cost of the Player of the repeated game is defined in the natural way. At each round  $t$  of the game, the player chooses  $\mathbf{w}_t$ , the adversary chooses  $i_t$ , and this continues until  $\mathbf{s}_t = \mathbf{e}_{i_1} + \dots + \mathbf{e}_{i_t}$  reaches the budget, i.e.  $\delta(\mathbf{s}_t) = 0$ . Let  $T$  be the first time where  $\delta(\mathbf{s}_t) = 0$ , then the Player's total cost is  $\sum_{t=1}^T \mathbf{w}_t^\top M \mathbf{e}_{i_t}$ . We can define the *value*  $V(M, \delta)$  of the game, when the payoff matrix is  $M$  and budget is  $\delta$ , as the smallest cost achievable against an optimal opponent. That is,

$$V(M, \delta) := \min_{\mathbf{w}: \mathbb{N}_0^n \rightarrow \Delta_n} \max_{i_1, i_2, \dots} \sum_{t=1}^{\infty} (\mathbf{w}(\mathbf{s}_{t-1})^\top M \mathbf{e}_{i_t}) \delta(\mathbf{s}_t) \quad (\text{where } \mathbf{s}_t := \mathbf{e}_{i_1} + \dots + \mathbf{e}_{i_t}).$$

Notice here that we characterize the Player's strategy in an oblivious way, i.e. as simply a function  $\mathbf{w} : \mathbb{N}_0^n \rightarrow \Delta_n$  which is independent of the past sequence of plays. It is easy to check that this is a legal simplification, as the state vector  $\mathbf{s}$  provides a "sufficient statistic" for the state of the game.

### 2.2 Inverse-nonnegative Matrices

In the present work, we consider only games whose payoff matrix is inverse-nonnegative.

**Definition 2.1.** A matrix  $M$  is inverse-nonnegative if all of the entries of  $M^{-1}$  are nonnegative. Equivalently,  $M$  is inverse-nonnegative if and only if for every  $\mathbf{y} \in \mathbb{R}_+^n$  the equation  $M\mathbf{x} = \mathbf{y}$  has a solution  $\mathbf{x} \in \mathbb{R}_+^n$ .

The inverse-nonnegative property is somewhat restrictive, yet as we show in Section 4 it encompasses a number of natural problems in online learning and other settings. We mention one characterization of inverse-nonnegativity which is known in the literature.

**Theorem 2.1** (From [2], Page 113, Theorem 2.6). *Let  $n \times n$  matrix  $A$  take the form  $A = \alpha I - B$  for a nonnegative matrix  $B$ . Then  $A$  is inverse-nonnegative if and only if  $\alpha$  is larger than the spectral radius of  $B$ .*

### 3 Minimax algorithm and their approximation

Our central results are proven in this Section. We begin in 3.1 by addressing the case when the budget  $\delta$  is known in advance to both players, and we provide precise proofs of the main result. In 3.2, we consider the case when  $\delta$  is unknown but drawn from a known prior. We state the corresponding Theorem regarding the minimax behavior of the game,

#### 3.1 Known Budget $\delta$

We now introduce a random process on the state space as follows, and we define a number of quantities associated with this process. At first, this may appear very mechanical, but it will soon become clear what role each object plays. Given the matrix  $M$ , we first define a number of quantities associated with  $M^{-1}$ .

$$\begin{aligned} |M^{-1}| &:= \mathbf{1}_n^\top M^{-1} \mathbf{1}_n = \sum_{i,j \in [n]} (M^{-1})_{i,j} \quad (\text{one-norm of } M^{-1}) \\ \mathbf{q} &:= \frac{M^{-1} \mathbf{1}_n}{|M^{-1}|} \quad (\text{row normalizers as a distribution}) \\ \mathbf{m}_i &:= \frac{\mathbf{e}_i M^{-1}}{\mathbf{e}_i^\top M^{-1} \mathbf{1}_n}, \quad \text{for } i = 1, \dots, n \quad (\text{normalized rows of } M^{-1}). \end{aligned}$$

With  $\mathbf{q}$  in hand, we define the following random process on the state space. Define a random sequence of indices  $I_1, I_2, \dots \in [n]$ , where each  $I_t$  is sampled according to the distribution  $\mathbf{q}$ . Now let  $S_t := \sum_{m=1}^t \mathbf{e}_{I_m}$ , where  $S_0 := \mathbf{0}$ . Assuming that we start at state  $\mathbf{s}$ , this induces a sequence of states

$$\mathbf{s} = \mathbf{s} + S_0 \rightarrow \mathbf{s} + S_1 \rightarrow \mathbf{s} + S_2 \rightarrow \dots \rightarrow \mathbf{s} + S_t \rightarrow \dots$$

It is understood that the Markov process  $\{\mathbf{s} + S_t\}_t$ , which we will refer to as the ‘‘random walk’’, is induced by the i.i.d. sequence  $I_1, I_2, \dots$ . The expectation will be with respect to these variables.

For the analysis we will need three more tools associated with this process.

$$\begin{aligned} \text{stop}(\mathbf{s}) &:= t \text{ s.t. } \delta(\mathbf{s} + S_t) = 0 \text{ and } \delta(\mathbf{s} + S_{t-1}) = 1 \\ &= \text{Time at which the budget was spent in random walk from } \mathbf{s} \\ I_{\text{stop}(\mathbf{s})} &= \text{The coordinate that spent the last of the budget} \\ \hat{p}_i(\mathbf{s}) &:= \Pr(I_{\text{stop}(\mathbf{s})} = i) \\ \Lambda_{\mathbf{q}, \delta}(\mathbf{s}) &:= \mathbb{E}_{\{I_j\} \sim \mathbf{q}} \left[ \sum_{t=0}^{\infty} \delta(\mathbf{s} + S_t) \right] = \mathbb{E}_{\{I_j\} \sim \mathbf{q}} [\text{stop}(\mathbf{s})] \\ &= \text{The expected length of the random walk before the budget is spent.} \end{aligned}$$

Notice that the random variables  $\text{stop}(\mathbf{s})$  and  $I_{\text{stop}(\mathbf{s})}$  are well-defined because the stopping criterion is monotone. It may not be immediately obvious why the quantity  $I_{\text{stop}(\mathbf{s})}$  is so important. However, as we will see, the optimal Player strategy will be to set  $\mathbf{w}(\mathbf{s}) = \mathbb{E}_{I_{\text{stop}(\mathbf{s})}} \mathbf{m}_i^\top = \sum_i \hat{p}_i(\mathbf{s}) \mathbf{m}_i^\top$ . We begin by expressing  $\hat{p}_i(\mathbf{s})$  in terms of the expected path length.

**Lemma 3.1.** For any state  $\mathbf{s}$  and any action  $i$ , we have

$$\widehat{p}_i(\mathbf{s}) = q_i (\Lambda_{\mathbf{q},\delta}(\mathbf{s}) - \Lambda_{\mathbf{q},\delta}(\mathbf{s} + \mathbf{e}_i)),$$

*Proof.* We can write the difference in expectations as

$$\begin{aligned} \Lambda_{\mathbf{q},\delta}(\mathbf{s}) - \Lambda_{\mathbf{q},\delta}(\mathbf{s} + \mathbf{e}_i) &= \mathbb{E}_{\{I_j\} \sim \mathbf{q}} \left[ \sum_{t=0}^{\infty} \delta(\mathbf{s} + S_t) - \delta(\mathbf{s} + \mathbf{e}_i + S_t) \right] \\ &= \sum_{t=0}^{\infty} \Pr(\delta(\mathbf{s} + S_t) > \delta(\mathbf{s} + \mathbf{e}_i + S_t)) \\ &= \sum_{t=0}^{\infty} \frac{\Pr(\delta(\mathbf{s} + S_t) > \delta(\mathbf{s} + \mathbf{e}_i + S_t)) \Pr(I_{t+1} = i)}{\Pr(I_{t+1} = i)} \\ &= \sum_{t=0}^{\infty} \frac{\Pr(\delta(\mathbf{s} + S_t) > \delta(\mathbf{s} + \mathbf{e}_i + S_t) \text{ and } I_{t+1} = i)}{\Pr(I_{t+1} = i)} \\ &= \sum_{t=0}^{\infty} \frac{\Pr(\text{Stop}(\mathbf{s}) = t + 1 \text{ and } I_{t+1} = i)}{q_i} \\ &= \frac{\Pr(I_{\text{Stop}(\mathbf{s})} = i)}{q_i} = \frac{\widehat{p}_i(\mathbf{s})}{q_i}. \end{aligned}$$

□

We call  $\frac{\Lambda_{\mathbf{q},\delta}(\mathbf{s})}{|M^{-1}|}$  the potential of the game at state  $\mathbf{s}$ . The following lemma shows that when the learner uses  $\mathbf{w}(\mathbf{s}) = \mathbb{E} \mathbf{m}_{I_{\text{Stop}(\mathbf{s})}}^\top$ , then for each action  $i$  of the Adversary, the cost is exactly the drop of the potential.

**Lemma 3.2.** For any state  $\mathbf{s}$  and any action  $i$ , we have

$$\mathbf{w}(\mathbf{s})^\top M \mathbf{e}_i = \frac{1}{|M^{-1}|} (\Lambda_{\mathbf{q},\delta}(\mathbf{s}) - \Lambda_{\mathbf{q},\delta}(\mathbf{s} + \mathbf{e}_i)), \quad \text{for } \mathbf{w}(\mathbf{s}) = \mathbb{E} \mathbf{m}_{I_{\text{Stop}(\mathbf{s})}}^\top.$$

*Proof.*

$$\begin{aligned} \mathbf{w}(\mathbf{s})^\top M \mathbf{e}_i &= \sum_i \widehat{p}_i(\mathbf{s}) \mathbf{m}_i M \mathbf{e}_i \\ &= \widehat{\mathbf{p}}(\mathbf{s})^\top \text{diag}^{-1}(M^{-1} \mathbf{1}_n) M^{-1} M \mathbf{e}_i \\ &= \frac{\widehat{p}_i(\mathbf{s})}{\mathbf{e}_i^\top M^{-1} \mathbf{1}_n} \\ &= \frac{1}{|M^{-1}|} \frac{\widehat{p}_i(\mathbf{s})}{q_i} \\ &= \frac{1}{|M^{-1}|} (\Lambda_{\mathbf{q},\delta}(\mathbf{s}) - \Lambda_{\mathbf{q},\delta}(\mathbf{s} + \mathbf{e}_i)). \end{aligned}$$

□

The following Theorem contains our main result. It gives a concise formula for the value of the game as well as the optimal probabilities of the Player.

**Theorem 3.1.** *The value of the game is the following normalized expected path length  $V(M, \delta) = \frac{\Lambda_{\mathbf{q}, \delta}(\mathbf{0})}{|M^{-1}|}$  and  $w(\mathbf{s}) = \mathbb{E} \mathbf{m}_{I_{\text{stop}(\mathbf{s})}}^\top$  is the optimal Player distribution at any state  $\mathbf{s}$ .*

*Proof.* We will first show that when the Player uses  $\mathbf{w}(\mathbf{s}) = \mathbb{E} \mathbf{m}_{I_{\text{stop}(\mathbf{s})}}^\top$ , then for every action sequence  $i_1, i_2, \dots$  chosen by the Adversary, the total cost is the total drop of the potential and this drop is always  $\frac{\Lambda_{\mathbf{q}, \delta}(\mathbf{s})}{|M^{-1}|}$ . This immediately shows that  $V(\mathbf{s}) \geq \frac{\Lambda_{\mathbf{q}, \delta}(\mathbf{s})}{|M^{-1}|}$  because the Player's choice may be non-optimal.

Indeed, assume we start at state  $\mathbf{s}_0 = \mathbf{0}$ , and the Adversary chooses an action sequence  $i_1, i_2, \dots, i_T$  at the end of which the budget is reached. Define the sequence of states  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_T$ , where  $\mathbf{s}_t = \mathbf{s}_{t-1} + \mathbf{e}_{i_t}$  for  $t = 1, \dots, T$ , and  $\delta(\mathbf{s}_T) = 0$  by assumption. Then by the previous lemma,

$$\begin{aligned} \sum_{t=1}^T \mathbf{w}(\mathbf{s}_{t-1})^\top M \mathbf{e}_{i_t} &= \sum_{t=1}^T \frac{1}{|M^{-1}|} (\Lambda_{\mathbf{q}}(\mathbf{s}_{t-1}) - \Lambda_{\mathbf{q}}(\mathbf{s}_t)) \\ &= \frac{\Lambda_{\mathbf{q}, \delta}(\mathbf{0}) - \Lambda_{\mathbf{q}, \delta}(\mathbf{s}_T)}{|M^{-1}|} \\ &= \frac{\Lambda_{\mathbf{q}, \delta}(\mathbf{0})}{|M^{-1}|}. \end{aligned}$$

To prove the matching lower bound on the value of the game, we imagine a randomized Adversary that ignores the Player's distribution and simply samples each action  $I_t$  according to the distribution  $\mathbf{q}$ . Then, for any distribution  $\mathbf{w}$  of the player, we see that

$$\mathbb{E}_{I_t \sim \mathbf{q}} \mathbf{w}^\top M \mathbf{e}_{I_t} = \mathbf{w}^\top M \mathbf{q} = \frac{\mathbf{w}^\top M M^{-1} \mathbf{1}_n}{|M^{-1}|} = \frac{1}{|M^{-1}|}$$

If the Adversary continues to use this oblivious randomized strategy throughout, then he is performing the random walk described above, generating a sequence of independently random indices  $I_1, I_2, \dots$  which induces the walk on the state space  $\{\mathbf{s} + S_t\}_t$ . We can then compute the expected cost of this strategy, against any Player strategy,

$$\sum_{t=1}^{\infty} \mathbb{E}_{I_t \sim \mathbf{q}} \mathbf{w}^\top M \mathbf{e}_{I_t} \delta(\mathbf{s} + S_t) = \mathbb{E} \left[ \sum_{t=0}^{\infty} \frac{\delta(\mathbf{s} + S_t)}{|M^{-1}|} \right] = \frac{\Lambda_{\mathbf{q}, \delta}(\mathbf{s})}{|M^{-1}|}$$

We conclude that  $\mathbf{w}(\mathbf{s}) = \mathbb{E} \mathbf{m}_{I_{\text{stop}(\mathbf{s})}}^\top$  is the optimal Player distribution and  $\mathbf{q}$  the optimum Adversary distribution.  $\square$

Before proceeding, we make an important observation. We make no claims as to the ease of computing  $\mathbf{w}(\mathbf{s}) = \mathbb{E} \mathbf{m}_{I_{\text{stop}(\mathbf{s})}}^\top$  as it requires finding the expected outcome of a random walk. On the other hand, it is typically assumed that the opponent in a zero-sum game can not see one's random coins. It thus suffices to sample  $I_{\text{stop}(\mathbf{s})}$  a *single* time, and play  $\mathbf{w}(\mathbf{s}) = \mathbf{m}_{I_{\text{stop}(\mathbf{s})}}^\top$ . This can be done very efficiently, and we provide the algorithm for completeness. Also it suffices to draw a single sequence  $\{I_t\}_t$  that is long enough. Even if the same sequence is used, we still have  $\mathbf{w}(\mathbf{s}) = \mathbb{E} \mathbf{m}_{I_{\text{stop}(\mathbf{s})}}^\top$ , but the variance might be larger in this case.

### 3.2 Unknown Budget $\delta$

The previous results are useful when the budget constraint  $\delta$  is known in advance. This may be a reasonable assumption in many cases, but it is unrealistic in many learning applications where we can only expect to

---

**Algorithm 1** RandWalk

---

Input: state  $\mathbf{s}$  s.t.  $\delta(\mathbf{s}) = 1$   
Initialize the random walk to  $\tilde{\mathbf{s}} := \mathbf{s}$   
**while**  $\delta(\tilde{\mathbf{s}}) = 1$  **do**  
    Increment  $t$  and sample  $I_t \sim \mathbf{q}$   
     $\tilde{\mathbf{s}} := \tilde{\mathbf{s}} + \mathbf{e}_{I_t}$   
**end while**  
Return( $\mathbf{m}_{I_t}^\top$ )

---

have a modest a priori knowledge of the Adversary's true budget. In the present Section, we consider a generalization in which the Adversary's budget is unknown but drawn, at the outset, from a known distribution  $\rho$ . Precisely, the game proceeds as follows:

1. Before the first round of the game, a budget  $\delta$  is drawn from  $\rho$  and given to the Adversary (but not to the Player)
2. Play proceeds in the usual fashion: at round  $t$ , the Player chooses  $\mathbf{w}(\mathbf{s})$ , the Adversary chooses  $\mathbf{e}_{i_t}$ , the Player suffers  $\mathbf{w}_t^\top M \mathbf{e}_{i_t}$ , and the state vector updates  $\mathbf{s} \leftarrow \mathbf{s} + \mathbf{e}_{i_t}$
3. The repeated game ends as soon as  $\delta(\mathbf{s}) = 0$ .

We require two restrictions on  $\rho$ . First, every  $\delta$  in the support of  $\rho$  has the monotonicity property as before. Second, we must have the ability to sample efficiently from  $\rho$ , as well as from the conditional distribution  $\rho|_{\delta(\mathbf{s})=1}$  for each  $\mathbf{s} \in \mathcal{S}$  (by  $\rho|_{\delta(\mathbf{s})=1}$  we mean the distribution conditioned on the event that  $\delta(\mathbf{s}) = 1$ ).

At first glance, it would appear that not knowing  $\delta$  is a significant handicap and should make the problem considerably more difficult. To the contrary, the random walk underlying the minimax solution only requires a very simple modification: at  $\mathbf{s}$ , simply “guess” the budget constraint  $\delta$ , by sampling from the *conditional* distribution  $\rho|_{\delta(\mathbf{s})=1}$ , and then proceed as though this were the true budget.

We need to slightly redefine several of our random objects. The coordinate sequence  $I_1, I_2, \dots$  drawn according to  $\mathbf{q}$ , and the associated random walk  $\{\mathbf{s} + S_t\}_t$ , will remain the same as it only depends on the game matrix  $M$ . However, the random variable  $\text{stop}(\mathbf{s})$  and the associated stopping coordinate  $I_{\text{stop}(\mathbf{s})}$  now depend on the random  $\delta$  which is drawn according to the conditional  $\rho|_{\delta(\mathbf{s})=1}$ . This induces the distribution  $\hat{p}(\mathbf{s})$ , defined by  $\hat{p}_i(\mathbf{s}) = \Pr(I_{\text{stop}(\mathbf{s})} = i)$ , as before. The expected random-walk path-length is defined as

$$\Lambda_{\mathbf{q}, \rho}(\mathbf{s}) = \mathbb{E}_{\delta \sim \rho|_{\delta(\mathbf{s})=1}} \mathbb{E}_{\{I_j\}} \left[ \sum_{t=0}^{\infty} \delta(\mathbf{s} + S_t) \right].$$

We define the value  $\hat{V}$  of the generalized zero-sum game as follows:

$$\hat{V}(M, \rho) := \min_{\mathbf{w}: \mathbb{N}_0^n \rightarrow \Delta_n} \mathbb{E} \max_{\delta \sim \rho} \sum_{t=1}^{\infty} (\mathbf{w}(\mathbf{s}_{t-1})^\top M \mathbf{e}_{i_t}) \delta(\mathbf{s}_t) \quad (\text{where } \mathbf{s}_t := \mathbf{e}_{i_1} + \dots + \mathbf{e}_{i_t}).$$

Notice the primary difference is that  $\delta$  is drawn after the player has committed to his strategy  $\mathbf{w}()$ .

**Theorem 3.2.** *The value of the generalized game is the following normalized expected path length  $\widehat{V}(M, \rho) = \frac{\Lambda_{\mathbf{q}, \rho}(\mathbf{0})}{|M^{-1}|}$  and the optimal Player distribution at state  $\mathbf{s}$  is*

$$\mathbf{w}(\mathbf{s}) = \mathbb{E}_{\delta \sim \rho | \delta(\mathbf{s})=1} \mathbb{E}_{\{I_j\}} \mathbf{m}_{I_{\text{stop}(\mathbf{s})}}^\top.$$

We omit the proof of the Theorem, as it follows very closely to that of Theorem 3.1. The main difference is in the following modified Lemma, which requires only a straightforward modification from Lemma 3.1.

**Lemma 3.3.** *For any state  $\mathbf{s}$  and any  $i$ , we have*

$$\frac{\widehat{p}_i(\mathbf{s})}{q_i} = \Lambda_{\mathbf{q}, \rho}(\mathbf{s}) - \Lambda_{\mathbf{q}, \rho}(\mathbf{s} + \mathbf{e}_i) \Pr_{\delta \sim \rho}(\delta(\mathbf{s} + \mathbf{e}_i) = 1 | \delta(\mathbf{s}) = 1).$$

*Proof.* We can rewrite  $\Lambda_{\mathbf{q}, \rho}(\mathbf{s} + \mathbf{e}_i) \Pr_{\delta \sim \rho}(\delta(\mathbf{s} + \mathbf{e}_i) = 1 | \delta(\mathbf{s}) = 1)$  as

$$\begin{aligned} & \mathbb{E}_{\{I_j\} \sim \mathbf{q}} \left[ \mathbb{E}_{\delta \sim \rho | \delta(\mathbf{s} + \mathbf{e}_i) = 1} \left[ \sum_{t=0}^{\infty} \delta(\mathbf{s} + \mathbf{e}_i + S_t) \Pr_{\delta \sim \rho}(\delta(\mathbf{s} + \mathbf{e}_i) = 1 | \delta(\mathbf{s}) = 1) \right] \right] \\ &= \mathbb{E}_{\{I_j\} \sim \mathbf{q}} \left[ \mathbb{E}_{\delta \sim \rho | \delta(\mathbf{s}) = 1} \left[ \sum_{t=0}^{\infty} \delta(\mathbf{s} + \mathbf{e}_i + S_t) \right] \right] \end{aligned}$$

We can now write the difference

$$\begin{aligned} & \Lambda_{\mathbf{q}, \rho}(\mathbf{s}) - \Lambda_{\mathbf{q}, \rho}(\mathbf{s} + \mathbf{e}_i) \Pr_{\delta \sim \rho}(\delta(\mathbf{s} + \mathbf{e}_i) = 1 | \delta(\mathbf{s}) = 1) \\ &= \mathbb{E}_{\{I_j\} \sim \mathbf{q}} \mathbb{E}_{\delta \sim \rho | \delta(\mathbf{s}) = 1} \left[ \sum_{t=0}^{\infty} \delta(\mathbf{s} + S_t) - \delta(\mathbf{s} + \mathbf{e}_i + S_t) \right]. \end{aligned}$$

The remainder of the proof follows exactly from Lemma 3.1. □

Again we can sample  $\mathbf{m}_{I_{\text{stop}(\mathbf{s})}}$  and use this as our estimate for  $\mathbf{w}(\mathbf{s})$ . Note that in our above formula for

---

### Algorithm 2 RandWalk

---

```

Input: state  $\mathbf{s}$ 
Sample  $\delta \sim \rho | \delta(\mathbf{s})=1$ 
Initialize the random walk to  $\tilde{\mathbf{s}} := \mathbf{s}$ 
while  $\delta(\tilde{\mathbf{s}}) = 1$  do
    Increment  $t$  and sample  $I_t \sim \mathbf{q}$ 
     $\tilde{\mathbf{s}} := \tilde{\mathbf{s}} + \mathbf{e}_{I_t}$ 
end while
Return( $\mathbf{m}_{I_t}$ )

```

---

the value of the game we are sampling  $\delta$  according to  $\rho$ . However since the summands of the inner max are multiplied by  $\delta(\mathbf{s})$  this is the same as sampling  $\delta$  conditional  $\rho | \delta(\mathbf{s})=1$ .

## 4 Applications

The present work was originally motivated by the minimax solution to the “experts game” [1]. In each trial the Learner chooses a distribution  $\mathbf{w}$  over the  $n$  experts, the adversary picks a loss vector  $\ell_t \in \{0, 1\}^n$  (specifying the loss of each of the  $n$  experts), and the Learner incurs the loss  $\mathbf{w} \cdot \ell_t$ . The game finishes as soon as the best expert incurs more than  $k$  losses, i.e. when the minimum coordinate of  $\sum_t \ell_t$  exceeds  $k$ .

This repeated experts game is close to the present framework, where one could define the state  $\mathbf{s}$  as  $\sum_t \ell_t$ , but this is not quite right as the Adversary has  $2^n$  actions and not  $n$ . The reduction can be fixed, however, with an additional lemma: it is easy to show that the game in which we restrict  $\ell \in \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  is no harder than the game where  $\ell \in \{0, 1\}^n$ . With this in mind, we can define the budget as

$$\delta(\mathbf{s}) = \begin{cases} 0 & \text{if } \mathbf{s} \geq (k+1, k+1, \dots, k+1) \\ 1 & \text{otherwise} \end{cases}.$$

The value of the game in this case is

$$V(\delta) := \min_{\mathbf{w}: \mathbb{N}_0^n \rightarrow \Delta_n} \max_{\ell_1, \ell_2, \dots} \sum_{t=1}^{\infty} \mathbf{w}(\mathbf{s}_{t-1})^\top \ell_t \delta(\mathbf{s}_t) \quad (\text{where } \mathbf{s}_t := \ell_1 + \dots + \ell_t).$$

Letting  $M$  be the identity, we have exactly a repeated game against a budgeted Adversary.

The techniques given in Section 3 provide a number of new algorithms for modified settings. For example,

1. Imagine that each expert  $i$  does not suffer a unit loss, but some arbitrary value  $c_i$ . In other words, some experts may be riskier than others. In our framework, we can simply set  $M = \text{diag}(c_1, \dots, c_n)$ . In addition, we can re-scale the budget criterion appropriately:  $\delta(\mathbf{s}) = 0$  when  $\min_i c_i s_i > k$ .
2. A different assumption on the experts is not that “there will be a good expert” but that “some pair of experts will be good”, for example. This can be modeled in the stopping criterion by setting  $\delta(\mathbf{s}) = 0$  whenever  $\min_{i,j} s_i + s_j > k$ .
3. Assuming that the loss of the best expert is known and fixed at  $k$  is generally unreasonable in many learning settings. A more natural assumption is that the loss bound  $k$  comes from some distribution, and the learner only knows the distribution but not the outcome. The techniques in Section 3.2 immediately give us an algorithm for this case, assuming that one can sample efficiently from the posterior.

## 5 Open problems

In this paper we showed that for a certain class of budgeted games, the optimal strategy is defined by an expected outcome at the end of a random walk. Fortunately, we can efficiently produce an unbiased estimate of the strategy based on a single random walk. However, we believe that computing the precise distribution for a general stopping criterion  $\delta$  can be hard computationally and might even be #P-complete. Determining the exact complexity is an open problem.

We presented the minimax strategy for the case when the budget is unknown but there is a prior on the budget. Typically, however, the budget is not actually drawn from a prior but is simply fixed and unknown; the prior assumption is used to hedge against all such possible budgets. We pose the following question.

What is the price paid for not knowing  $k$ ? That is, when the prior is chosen well, how much more, in terms of  $k$ , does the informed Player pay versus the uniformed Player? We do not have an answer to this question, and we conjecture that it is  $O(\log k)$ .

## References

- [1] J. Abernethy, M. K. Warmuth, and J. Yellin. Optimal strategies from random walks. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT 08)*, pages 437–445, July 2008.
- [2] Abraham Berman and Robert J. Plemmons. *Nonnegative matrices in the mathematical sciences*. SIAM, 1994.
- [3] K. Chaudhuri, Y. Freund, and D. Hsu. A parameter-free Hedging algorithm. Arxiv/0903.2851, May 2009.
- [4] Y. Freund. A method for Hedging in continuous time. Arxiv/0904.3356, May 2009.
- [5] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to Boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, 1997. Special Issue for EuroCOLT '95.
- [6] N. Littlestone and M. K. Warmuth. The Weighted Majority algorithm. *Inform. Comput.*, 108(2):212–261, 1994. Preliminary version in FOCS 89.
- [7] V. Vovk. Aggregating strategies. In *Proceedings of the Third Annual Workshop on Computational Learning Theory*, pages 371–383. Morgan Kaufmann, 1990.