The Big Match in Small Space

Kristoffer Arnsfelt Hansen*[1], Rasmus Ibsen-Jensen[2], and Michal Koucký†[3]

[1]Aarhus University
[2]IST Austria
[3]Charles University, Prague

April 26, 2016

Abstract

In this paper we study how to play (stochastic) games optimally using little space. We focus on repeated games with absorbing states, a type of two-player, zero-sum concurrent mean-payoff games. The prototypical example of these games is the well known Big Match of Gillete (1957). These games may not allow *optimal* strategies but they always have ε-optimal strategies. In this paper we design ε-optimal strategies for Player 1 in these games that use only $O(\log \log T)$ space. Furthermore, we construct strategies for Player 1 that use space $s(T)$, for an arbitrary small unbounded non-decreasing function s, and which guarantee an ε-optimal value for Player 1 in the limit superior sense. The previously known strategies use space $\Omega(\log T)$ and it was known that no strategy can use constant space if it is ε-optimal even in the limit superior sense. We also give a complementary lower bound.

Furthermore, we also show that no Markov strategy, even extended with finite memory, can ensure value greater than 0 in the Big Match, answering a question posed by Abraham Neyman.

*The author acknowledges support from the Danish National Research Foundation and The National Science Foundation of China (under the grant 61361136003) for the Sino-Danish Center for the Theory of Interactive Computation and from the Center for Research in Foundations of Electronic Markets (CFEM), supported by the Danish Strategic Research Council.

†The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 616787. Supported in part by grant from Neuron Fund for Support of Science, and by the Center of Excellence CE-ITI under the grant P202/12/G061 of GA ČR.

Contents

1 Introduction

In game theory there has been considerable interest in studying the complexity of strategies in infinitely repeated games. A natural way how to measure the complexity of a strategy is by the number of states of a finite automaton implementing the strategy. A common theme is to consider what happens when some or all players are restricted to play using a strategy given by an automaton of a certain bounded complexity.

Asymptotic view. Previous works have mostly been limited to dichotomy results: either there is a good strategy implementable by finite automaton or there is no such strategy. Our goal here is to refine this picture. We do this by taking the asymptotic view: measuring the complexity as a function of the number of rounds played in the game. Now when the strategy no longer depends just on a finite amount of information about the history of the game it could even be a computationally difficult problem to decide the next move of the strategy. But we focus on investigating how much information a good strategy must store about the play so far to decide on the next move; in other words, we study how much space the strategy needs.

Game classes. The class of games we study is that of repeated zero-sum games with absorbing states. These form a special case of undiscounted stochastic games. Stochastic games were introduced by Shapley [14], and they constitute a very general model of games proceeding in rounds. We consider the basic version of two-player zero-sum stochastic games with a constant number of states and a constant number of actions. In a given round t the two players simultaneously choose among a number of different *actions* depending on the current *state*. Based on the choice of the pair (i, j) of actions as well as the current state k, Player 1 receives a *reward* $r_t = a_{ij}^k$ from Player 2, and the game proceeds to the next state ℓ according to probabilities $p_{ij}^{k\ell}$.

Limit-average rewards. In Shapley's model, in every round the game stops with non-zero probability, and the payoff assigned to Player 1 by a play is simply the sum of rewards r_i. The stopping might be viewed as *discounting* later rewards by a discounting factor $0 < \beta < 1$. Gillette [5] considered the more general model of undiscounted stochastic games where all plays are infinite. He is interested in the average reward $\frac{1}{T}\sum_{t=1}^{T} r_t$ to Player 1 as T tends to infinity. As the limit may not exist one needs to consider $\liminf$, $\limsup$, or some Banach limit [15] of the sums. In many cases the particular choice of the limit does not matter much, but it turns out that for our results it has interesting consequences. For this reason we consider both $\liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} r_t$ and $\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} r_t$.

Note that both these notions have natural interpretations. For instance, the $\liminf$ notion suits the setup where the infinite repeated game actually models a game played repeatedly for an unspecified (but large) number of rounds, where one thus desires a guarantee on the average reward after a certain number of rounds. The $\limsup$ notion on the other hand models the ability to always recover from arbitrary losing streaks in the repeated game.

The Big Match. A prototypical example of an undiscounted stochastic game is the well-known Big Match of Gillette [5] (see Figure 1 for an illustration of the Big Match). This game fits also into an important special subclass of undiscounted stochastic games: the *repeated games with absorbing states*, defined by Kohlberg [11]. In a repeated game with absorbing states there is only one state that can be left; all the other states are *absorbing*, i.e., the probability of leaving them is zero regardless of the actions of the players. Even in these games, as for general undiscounted stochastic games, there might not be an optimal strategy for the players [5]. On the other hand there always exist ε-optimal strategies [11], which are strategies guaranteeing the value of the game up to an additive term ε. The Big Match provides such an example: the value of the game is 1/2, but Player 1 does not have an optimal strategy, and must settle for an ε-optimal strategy [2]. On the other hand, it is known that such ε-optimal strategies in the Big Match must have a certain level of complexity. More precisely, for any $\varepsilon < \frac{1}{2}$, an ε-optimal strategy can neither be implemented by a finite automaton nor take the form of a Markov strategy (a strategy whose only dependence on the history is the number of rounds played) [16].

In this paper we consider the Big Match in particular and then generalize our results to general repeated games with absorbing states.

The model under consideration. We are interested in the *space complexity* of ε-optimal strategies in repeated games with absorbing states. A general strategy of a player in a game might depend on the whole history of the play up to the current time step. Moreover the decision about the next move might depend arbitrarily on the history. This provides the strategies with lots of power. There are two natural ways how to restrict the strategies: one can put computational restrictions on how the next move is decided based on the history of the play, or one can put a limit on how much information can the strategy remember about the history. One can also combine both types of restrictions, which leads to an interactive Turing machine based model, modelling a dynamic algorithm.

In this paper we mainly focus on restricting the amount of information the strategy can remember. This restriction is usually studied in the form of how *large size a finite automaton* (transducer) for the strategy has to be, and we follow this convention. By the size of a finite automaton we mean the number of states. The automatons we consider can make use of probabilistic transitions, and we will not consider the describtion of these probabilities as part of the size of the automaton. We do address these separately, however.

History of the model. The idea of measuring complexity of strategies in repeated games in terms of automata was proposed by Aumann [1]. The survey by Kalai [10] further discuss the idea in several settings of repeated games. However in this line of research the finite automata is assumed to be fixed for the duration of the game. This represents a considerable restriction as for many games there is no good strategy that could be described in this setting. Hence we consider strategies in which the automata can grow with time. To be more precise we consider infinite automata and measure how many different states we could have visited during the first T steps of the play. The logarithm of this number corresponds to the amount of space one would need to keep track of the current state of the automaton. We are interested in how this space grows with the number of rounds of the play.

Comparison of our model with a Turing machine based model. To impose also computational restrictions on the model, one can consider the usual Turing machine with one-way input and output tapes that work in lock-step and that record the play: whenever the machine writes its next action on the output tape it advances the input head to see the corresponding move of the other player. The space usage of the model is then the work space used by the machine, growing with the number of actions processed. The Turing machine can be randomized to allow for randomized strategies. The main differences between this model and the automaton based model we focus on in this paper is that in the case of infinite automata the strategy can be *non-uniform* and use *arbitrary* probabilities on its transitions whereas the Turing machine is *uniform* in the sense that it has a finite program that is fixed for the duration play and in particular, all transition probabilities are explicitly generated by the machine.

Bounds for strategies with deterministic update. Trivially, any strategy needs space at most

$O(T)$, since such memory would suffice to remember the whole history of the play. It is not hard to see (cf. [9, Chap. 3.2.1]) that if a strategy is not restricted to a finite number of states, then the number of reachable states by round T must be at least T. This means that the space needed by any such strategy is $\Omega(\log T)$. However this provides only worst-case answer to our question, since for randomized strategies it might happen that only negligible fraction of the states can be reached with reasonable probability. Indeed, it might be that with probability close to 1 the strategy reaches only a very limited number of states. This is the setup we are interested in. As we will see in a moment the strategies we consider use substantially less space than $O(\log T)$ with high probability (and $O(\log T)$ space in the worst case).

Relationship to data streaming We find that our question is naturally related to algorithmic questions in data streaming. In data streaming one tries to estimate on-line various properties of a data stream while minimizing the amount of information stored about the stream. As we will see our solutions borrow ideas from data streaming in particular, we use sampling to estimate properties of the play so far. It is rather interesting that this is sufficient for a large class of games.

1.1 Our results

We provide two types of results. We show that there are ε-optimal strategies for repeated games with absorbing states, and we also show that there are limits on how small space such strategies could possibly use. Our strategies are first constructed for the Big Match. Then, following Kohlberg [11] these strategies are extended to general repeated games with absorbing states.

Upper bounds on space usage. Our first results concern the Big Match. We show that for all $\varepsilon > 0$, there exists an ε-optimal strategy that uses $O(\log\log T)$ space with probability $1-\delta$ for any $\delta > 0$. We note that the previous constructed strategies of Blackwell and Ferguson [2] and Kohlberg [11] uses space $\Theta(\log T)$.

Theorem 1. *For all $\varepsilon > 0$, there is an ε-optimal strategy σ_1 for Player 1 in the Big Match such that for any $\delta > 0$ with probability at least $1-\delta$, the strategy σ_1 uses $O(\log\log T)$ space in round T.*

Remark. We would like to stress the order of quantification above and their impact on the big-O notation used above for conciseness. The strategy we build depends on the choice of ε, but only for the actions made – the memory updates are independent thereof, and thus likewise is the space usage. The dependence of δ is also very benign. More precisely, there exists a constant $C > 0$ independent of ε and δ, and an integer T_0 depending on δ, but independent of ε, in such a way that with probability at least $1-\delta$, the strategy σ_1 uses at most space $C\log\log T$, for all $T \geq T_0$. The same remark holds elsewhere in our statements.

Our results translated to the Turing based model. After a slight modification our ε-optimal strategy can be implemented by a Turing machine so that (1) it processes T actions in time $O(T)$; and (2) each time it processes an action, all randomness used comes from at most 1 unbiased coin flip; and (3) it, for all $\delta > 0$, uses $O(\log\log T + \log\log\varepsilon^{-1})$ space with probability $1-\delta$, before round T. See Corollary 19.

Arbitrary small, but growing space for $\limsup$**.** For the case of $\limsup$ evaluation of the average rewards we can design strategies that uses even less space, in fact arbitrarily small, but growing, space.

Theorem 2. *For any non-decreasing* unbounded *function s, there exists an ε-supremum-optimal strategy σ_1 for Player 1 in the Big Match such that for each $\delta > 0$, with probability at least $1 - \delta$, strategy σ_1 uses $O(s(T))$ space in round T.*

We may for instance think of s as the inverse of the Ackermann function. Although the strategy from this theorem has very uniform description it might not always be implementable by a Turing machine running in the same space since the machine needs to sample probabilistic events comparable to $1/T$ or smaller. That might not be achievable in small space using just a fair coin.

Our strategy that is ε-optimal is actually an instantiation of the ε-supremum-optimal strategy to the setting of $O(\log \log T)$ space. We are unable to achieve ε-optimality in less space, and this seems to be inherent to our techniques.

Generalization to repeated games with absorbing states. We can generalize the above statements to the case of general repeated games with absorbing states.

Theorem 3. *For all $\varepsilon > 0$ and any repeated game with absorbing states G, there is an ε-optimal strategy σ_k for Player k in G such that, for each $\delta > 0$, with probability at least $1 - \delta$, the strategy σ_k uses $O(\log \log T + \log 1/\varepsilon \cdot \mathrm{poly}(|G|))$ space in round T.*

Theorem 4. *For all $\varepsilon > 0$, any repeated game with absorbing states G, and any non-decreasing unbounded function s, there exists an ε-supremum optimal strategy σ_k for Player k in G such that for each $\delta > 0$, with probability at least $1-\delta$, the strategy σ_1 uses $O(s(T) + \log 1/\varepsilon \cdot \mathrm{poly}(|G|))$ space in round T.*

These strategies are obtained by reducing to a special simple case of repeated games with absorbing states, generalized Big Match games, to which our Big Match strategies can be generalized. This reduction can furthermore be done effectively by a polynomial time algorithm.

Lower bound on space usage. We provide two lower bounds on space addressing different aspects of our strategies. One property of our strategies is that the smaller the space used is, the smaller the probabilities of actions employed are. The reciprocal of the smallest non-zero probability is the *patience* of a strategy. This is a parameter of interest for strategies. We show that the patience of our strategies is close to optimal. In particular, we show that the first $f(T)$ memory states must use probabilities close to $1/T^{f(T)}$, where $s(T) = \log f(T)$ is the space usage. We can almost match this bound by our strategies.

Finite-memory deterministic-update Markov strategies are no good. Beside the lower bound on patience we investigate the possibility of using a good strategy for Player 1 which would use only a constant number of states but where the actions could also depend on the round number. This is what we call a *finite-memory Markov strategy*. We show that such a strategy which also updates its memory state deterministically cannot exist. This answers a question posed by Abraham Neyman.

Theorem 5. *For all $\varepsilon < \frac{1}{2}$, there exists no finite-memory deterministic-update ε-optimal Markov strategy for Player 1 in the Big Match.*

1.2 Our techniques

The previously given strategies for Player 1 in the Big Match [2, 11] use space $\Theta(\log T)$ as they maintain the count of the number of different actions taken by the other player. There are two principal ways how one could try to decrease the number of states for such randomized strategies: either to use approximate counters [13, 4], or to sub-sample the stream of actions of the other player and use a good strategy on the sparse sample. In this paper we use the latter approach.

Overview over our strategy for the Big Match. Our strategies for Player 1 proceed by observing the actions of Player 2 and collecting statistics on the payoff. Based on these statistics Player 1 adjusts his actions. The statistics is collected at random sample points and Player 1 plays according to a "safe" strategy on the points not sampled and plays according to a good (but space-inefficient) strategy on the sample points. If the space of Player 1 is at least $\log\log T$ then Player 1 is able to collect sufficient statistics to accurately estimate properties of the actions of Player 2. Namely, substantial dips in the average reward given to Player 1 can be detected with high probability and Player 1 can react accordingly. Thus that during infinite play, the average reward will not be able drop for extended periods of time, and this will guarantee that $\liminf$ evaluation of the average rewards is close to the value of the game.

The bottle-neck in the $\liminf$ case. However, if our space is considerably less than $\log\log T$ we do not know how to accurately estimate these properties of the actions of Player 2. Thus, long stretches of actions of Player 2 giving low average rewards might go undetected as long as they are accompanied by stretches of high average rewards. Thus one could design a strategy for Player 2 that has low $\liminf$ value of the average rewards, but has large $\limsup$ value. Against such a strategy, our space-efficient strategy for Player 1 is unlikely to stop. So during infinite play, while our strategy guarantees that the $\limsup$ evaluation of the average rewards is close to the value of the game, it performs poorly under $\liminf$ evaluation. It is not clear whether this is an intrinsic property of all very small space strategies for Player 1 or whether one could design a very small space strategy achieving that the $\liminf$ evaluation of the average rewards is close to the value of the game. We leave this as an interesting open question.

Generalizing to repeated games with absorbing states. Our extension to general repeated games with absorbing states follow closely the work of Kohlberg [11]. He showed that all such games have a value and constructed ε-optimal strategies for them, building on the work of Blackwell and Ferguson [2]. His construction is in two steps: The question of value and of ε-optimal strategies are solved for a special case of repeated games with absorbing states, generalized Big Match games, that are sufficiently similar to the Big Match game that one of the strategies given by Blackwell and Ferguson [2] can be extended to this more general class of games. Having done this, Kohlberg shows how to reduce general repeated games with absorbing states to generalized Big Match games.

In a similar way we can extend our small-space strategies for the Big Match to the larger class of generalized Big Match games. These can then directly be used for Kohlberg's reduction. This reduction is however only given as an existence statement. We show how the reduction can be made explicit and computed by a polynomial time algorithm. This is done using linear programming formulations and fundamental root bounds of univariate polynomials. This also provides explicit bounds on the bitsize of the reduced generalized Big Match games. We also give a simple polynomial time algorithm for approximating the value of any repeated game with absorbing states based on bisection and linear programming.

2 Definitions

Probability distributions. A *probability distribution* over a finite set S, is a map $d : S \to [0,1]$, such that $\sum_{s\in S} d(s) = 1$. Let $\Delta(S)$ denote the set of all probability distributions over S.

Repeated games with absorbing states. The games we consider are special cases of two player, zero-sum concurrent mean-payoff games in which all states except at most one are *absorbing*, i.e. never left if entered (note also that an absorbing state can be assumed to have just a single action for each player). We restrict our definitions to this special case, introduced by Kohlberg [11] as repeated games with absorbing states. Such a game G is given by sets of actions A_1 and A_2 for each player together with maps $\pi : A_1 \times A_2 \to \mathbb{R}$ (the stage payoffs) and $\omega : A_1 \times A_2 \to [0, 1]$ (the absorption probabilities).

The game G is played in rounds. In every round $T = 1, 2, 3, \ldots$, each player $k \in \{1, 2\}$ independently picks an action $a_k^T \in A_k$. Player 1 then receives the stage payoff $\pi(a_1^T, a_2^T)$ from Player 2. Then, with probability $\omega(a_1^T, a_2^T)$ the game stops and all payoffs of future rounds are fixed to be $\omega(a_1^T, a_2^T)$ (we may think of this as the game proceeding to an absorbing state where the (unique) stage payoff for future rounds is $\pi(a_1^T, a_2^T)$). Otherwise, the game just proceeds to the next round.

The sequence $(a_1^1, a_2^1), (a_1^2, a_2^2), (a_1^3, a_2^3), \ldots$ of actions taken by the two players is called a *play*. A finite play occurs when the game stops after the last pair of actions. Otherwise the play is infinite. To a given play P we associate an infinite sequence of rewards $(r_T)_{T \geq 1}$ received by Player 1. If $P = (a_1^1, a_2^1), (a_1^2, a_2^2), \ldots, (a_1^\ell, a_2^\ell)$ is a finite play of length ℓ we let $r_T = \pi(a_1^T, a_2^T)$ for $1 \leq T \leq \ell$, and $r_T = \pi(a_1^\ell, a_2^\ell)$ for $T > \ell$. In this case we say that the game stops with *outcome* r_ℓ.

Otherwise, if $P = (a_1^1, a_2^1), (a_1^2, a_2^2), \ldots$ is infinite we simply let $r_T = \pi(a_1^T, a_2^T)$ for all $T \geq 1$.

To evaluate the sequence of the rewards we consider both the lim inf and lim sup value of the average reward $\frac{1}{T} \sum_{t=1}^{T} r_t$. We thus define the limit-infimum payoff to Player 1 of the play as

$$u_{\inf}(P) = \liminf_{n \to \infty} \frac{1}{n} \sum_{T=1}^{n} r_T \quad ,$$

and similarly the limit-supremum payoff to Player 1 of the play as

$$u_{\sup}(P) = \limsup_{n \to \infty} \frac{1}{n} \sum_{T=1}^{n} r_T \quad .$$

Strategies. A *strategy* for Player k is a function $\sigma_k : (A_1 \times A_2)^* \to \Delta(A_k)$ describing the probability distribution of the next chosen action after each finite play. We say that Player k *follows* a strategy σ_k if for every finite play P of length $T - 1$, at round T Player k picks the next action according to $\sigma_k(P)$. We say that a strategy σ_k is *pure* if for every finite play P the distribution $\sigma_k(P)$ assigns probability 1 to one of the actions of A_k (i.e. the next action is uniquely determined). Also, we say that a strategy σ_k is a *Markov strategy* if for every T and every play P of length $T-1$, the distribution $\sigma_k(P)$ does not depend on the particular actions during the first $T - 1$ rounds but is just a function of T. Thus Markov strategy σ_k can be viewed as a map $\mathbb{Z}_+ \to \Delta(A_k)$ or simply a sequence of distributions over A_k.

A *strategy profile* σ is a pair of strategies (σ_1, σ_2), one for each player. A strategy profile σ defines a probability measure on plays in the natural way. We define the expected limit-infimum payoff to Player 1 of the strategy profile $\sigma = (\sigma_1, \sigma_2)$ as $u_{\inf}(\sigma) = u_{\inf}(\sigma_1, \sigma_2) = \mathrm{E}_{P \sim (\sigma_1, \sigma_2)}[u_{\inf}(P)]$ and similarly the expected limit-supremum payoff to Player 1 of the strategy profile σ as $u_{\sup}(\sigma) = u_{\sup}(\sigma_1, \sigma_2) = \mathrm{E}_{P \sim (\sigma_1, \sigma_2)}[u_{\sup}(P)]$.

Values and near-optimal strategies. We define the *lower values* of G by $\underline{\mathrm{v}}_{\inf} = \sup_{\sigma_1} \inf_{\sigma_2} u_{\inf}(\sigma_1, \sigma_2)$ and $\underline{\mathrm{v}}_{\sup} = \sup_{\sigma_1} \inf_{\sigma_2} u_{\sup}(\sigma_1, \sigma_2)$, and we define the *upper values* of G by $\overline{\mathrm{v}}_{\inf} = \inf_{\sigma_2} \sup_{\sigma_1} u_{\inf}(\sigma_1, \sigma_2)$ and $\overline{\mathrm{v}}_{\sup} = \inf_{\sigma_2} \sup_{\sigma_1} u_{\sup}(\sigma_1, \sigma_2)$. Clearly $\underline{\mathrm{v}}_{\inf} \leq \underline{\mathrm{v}}_{\sup} \leq \overline{\mathrm{v}}_{\sup}$ and $\underline{\mathrm{v}}_{\inf} \leq \overline{\mathrm{v}}_{\inf} \leq \overline{\mathrm{v}}_{\sup}$. Kohlberg showed that all these values coincide and we call this common number $\mathrm{v}(G)$ the value v of G.

Theorem 6 (Kohlberg, Theorem 2.1)**.** $\underline{\mathrm{v}}_{\inf} = \overline{\mathrm{v}}_{\sup}$.

Note that this also shows that for the purpose of defining the value of G the choice of the limit of the average rewards does not matter. But a given strategy σ_1 for Player 1 could be close to guaranteeing the value with respect to lim sup evaluation of the average rewards, while being far from doing so with respect to the more restrictive lim inf evaluation. We shall hence distinguish between these different guarantees.

Let $\varepsilon > 0$ and let σ_1 be a strategy for Player 1. We say that σ_1 is *ε-supremum-optimal*, if

$$\mathrm{v}(G) - \varepsilon \leq \inf_{\sigma_2} u_{\sup}(\sigma_1, \sigma_2)$$

and that σ_1 is *ε-optimal*, if

$$\mathrm{v}(G) - \varepsilon \leq \inf_{\sigma_2} u_{\inf}(\sigma_1, \sigma_2) \ .$$

Observation 1. *Clearly it is sufficient to take the infimum over just pure strategies σ_2 for Player 2, and hence when showing that a particular strategy σ_1 is ε-supremum-optimal or ε-optimal we may restrict our attention to pure strategies σ_2 for Player 2.*

One can naturally make similar definitions for Player 2, where the roles of lim inf and lim sup would then be interchanged, but we shall restrict ourselves here to the perspective of Player 1.

If the strategy σ_1 is 0-supremum-optimal (0-optimal) we simply say that σ_1 is supremum-optimal (optimal). The Big Match gives an example where Player 1 does not have a supremum-optimal strategy [2].

Memory and memory-based strategies. A *memory configuration* or *state* is simply a natural number. We will often think of memory configurations as representing discrete objects such as tuples of integers. In such a case we will always have a specific encoding of these objects in mind.

Let $\mathcal{M} \subseteq \mathbb{N}$ be a set of memory states. A *memory-based strategy* σ_1 for Player 1 consists of a *starting state* $m_s \in \mathcal{M}$ and two maps, the *action map* $\sigma_1^a : \mathcal{M} \to \Delta(A_1)$ and the *update map* $\sigma_1^u : A_1 \times A_2 \times \mathcal{M} \to \Delta(\mathcal{M})$. We say that Player 1 *follows* the memory-based strategy σ_1 if in every round T when the game did not stop yet, he picks his next move a_1^T at random according to $\sigma_1^a(m_T)$, where the sequence $m_1, m_2, \ldots$ is given by letting $m_1 = m_s$ and for $T = 1, 2, 3, \ldots$ choosing m_{T+1} at random according to $\sigma_k^u(a_1^T, a_2^T, m_T)$, where a_2^T is the action chosen by Player 2 at round T.

The strategies we construct in this paper have the property that their action maps do not depend on the action a_1^T of Player 1. In these cases we simplify notation and write just $\sigma_1^u(a_2^T, m^T)$.

Since each finite play can be encoded by a binary string, and thus a natural number, we can view any strategy σ_k for Player k as a memory-based strategy. One can find similarly defined types of strategies in the literature, but typically, the function corresponding to the update function is deterministic.

	L	**R**
L	1	0
R	0*	1*

Figure 1: The Big Match in a matrix form.

Memory sequences and space usage of memory-based strategies. Let σ_1 be a memory-based strategy for Player 1 on memory states $\mathcal{M}$ and σ_2 be a strategy for Player 2. Assume that Player 1 follows σ_1 and Player 2 follows σ_2. The strategy profile (σ_1, σ_2) defines a probability measure on (finite and infinite) sequences over $\mathcal{M}$ in the natural way. For a (finite) sequence $M \in \mathcal{M}^*$, let $\omega_1(M)$ denote the probability that Player 1 follows this sequence of memory states during the first $|M|$ rounds of the game, while the game does not stop before round $|M|$.

Fix a non-decreasing function $f : \mathbb{N} \to \mathbb{N}$ and a probability p. The strategy σ_1 *uses* $\log f(T)$ *space with probability at least* p *against* σ_2, if for all T, the probability $\Pr_{(\sigma_1,\sigma_2)}[\forall i \le T : M_i \le f(T)] \ge p$ (i.e., with probability at least p, the current memory has stayed below that of $f(T)$ before round T, for all T). If σ_1 uses $\log f(T)$ space with probability at least p against every strategy σ_2', then we say that σ_1 *uses* $\log f(T)$ *space with probability at least* p.

The Big Match. The Big Match, introduced by Gillette [5] is a simply defined repeated game with absorbing states, where each player has only two actions. In each round Player 1 has the choice to stop the game (action **R**), or continue with the next round (action **L**). Player 2 has the choice to declare the round safe (action **L**) or unsafe (action **R**). If play continues in a round declared safe, or if play stops in a round declared unsafe, Player 2 must give Player 1 a reward 1. In the other two cases no reward is given.

More formally, let the action sets be $A_1 = A_2 = \{\mathbf{L}, \mathbf{R}\}$. The rewards are given by $\pi(a_1, a_2) = 1$ if $a_1 = a_2$ and $\pi(a_1, a_2) = 0$ if $a_1 \ne a_2$. The stopping probabilities are given by $\omega(\mathbf{R}, a_2) = 1$ and $\omega(\mathbf{L}, a_2) = 0$.

We can illustrate this game succinctly in a matrix form as shown in Figure 1, where rows are indexed by the actions of Player 1, columns are indexed by the actions of Player 2, entries give the rewards, and a star on the reward means that the game stops with probability 1 (See Section 8.1 for a general definition of the matrix form of a repeated game with absorbing states).

Density of pure Markov Strategies in the Big Match. When constructing strategies for Player 1 in the Big Match, not only is it sufficient to consider only pure strategies for Player 2 as noted in Observation 1, but we may restrict our consideration to pure Markov strategies, since Player 2 only ever observes the action **L** of Player 1. An important property of a pure Markov strategy σ for Player 2 in the Big Match is the *density* of **L** actions of a prefix of σ.

Denote by $\sigma^T \in \{\mathbf{L}, \mathbf{R}\}^T$, the length T prefix of σ. The *density of* ***L*** in σ^T, denoted $\mathrm{dens}(\sigma^T)$, is defined by

$$\mathrm{dens}(\sigma^T) = \frac{|\{i \mid (\sigma^T)_i = \mathbf{L}\}|}{T} \quad .$$

Further, for $T' < T$ we define

$$\mathrm{dens}(\sigma, T', T) = \frac{|\{i \ge T' \mid (\sigma^T)_i = \mathbf{L}\}|}{T - T' + 1} \quad .$$

Observation 2. *Suppose Player 2 follows a pure Markov strategy σ. Then for any play P and $T < |P|$ we have*

$$\mathrm{dens}(\sigma^T) = \frac{1}{T} \sum_{T'=1}^{T} r_T \ ,$$

where r_T is the reward given to Player 1 in round T. In particular, when P is infinite, we have

$$u_{\inf}(P) = \liminf_{T \to \infty} \mathrm{dens}(\sigma^T) \ ,$$

and

$$u_{\sup}(P) = \limsup_{T \to \infty} \mathrm{dens}(\sigma^T) \ .$$

3 Small space ε-supremum-optimal strategies in the Big Match

For given $\varepsilon > 0$, let $\xi = \varepsilon^2$. For any non-decreasing and *unbounded* function $f : \mathbb{Z}_+ \to \mathbb{Z}_+$, we will now give an ε-supremum optimal strategy σ_1^* for Player 1 in the Big Match that for all $\delta > 0$ with probability $1 - \delta$ uses $O(\log f(T))$ space. Let $\overline{f}$ be a strictly increasing unbounded function from $\mathbb{Z}_+$ to $\mathbb{R}_+$, such that $\overline{f}(x) \leq f(x)$ for all $x \in \mathbb{Z}_+$, and let F be the inverse of $\overline{f}$. For simplicity, and without loss of generality, we assume that $F(1) = 1$ and $F(T+1) \geq 2 \cdot F(T)$. Note that in particular $\overline{f}(2 \cdot T) \leq 2 \cdot \overline{f}(T)$.

Intuitive description of the strategy and proof. The main idea for building the strategy is to partition the rounds of the game into epochs, such that epoch i has expected length $F(i)$. The i'th epoch is further split into i sub-epochs. In each sub-epoch j of the i-th epoch we sample i^2 rounds uniformly at random. In every round not sampled we simply stay in the same memory state and play **L** with probability 1. We view the i^2 samples as a stream of actions chosen by Player 2. We then follow a particular ξ-optimal base strategy $\sigma_1^{i,\xi}$ for the Big Match on the samples of sub-epoch j. This strategy $\sigma_1^{i,\xi}$ is a suitably modified version of a strategy by Blackwell and Ferguson [2] and Kohlberg [11].

More precisely, if $\sigma_1^{i,\xi}$ stops in its k-th round when run on the samples of sub-epoch j, the strategy σ_1^* stops on the k-th sample in sub-epoch j. This will ensure that if σ_1^* stops with probability at least $\sqrt{\xi}$, the outcome is at least $\frac{1}{2} - \xi$.

Also, for any $0 < \delta < \frac{1}{2}$ and for sufficiently large i, depending on δ, if the samples have density of **L** at most $\frac{1}{2} - \delta$ then $\sigma_1^{i,\xi}$ stops on the samples with a positive probability depending only on ξ, namely ξ^4. For $f(T) = \Theta(\log T)$, the division into sub-epochs ensures that if $\liminf_{T\to\infty} \mathrm{dens}(\sigma^T) < \frac{1}{2}$ then infinitely many sub-epochs have density of **L** smaller than $1/2$, and thus the play stops with probability 1 in one of such epochs. This is not necessarily true for $f(T)$ smaller than $\log T$.

The base strategy. The important inner part of our strategy is a ξ-optimal strategy $\sigma_1^{i,\xi}$ parametrized by a non-negative integer i. These strategies are similar to ξ-optimal strategies given by Blackwell and Ferguson [2] and Kohlberg [11] (in fact, setting $i = 0$ and replacing ξ^4 by ξ^2 below one obtains the strategy used by Kohlberg).

The strategy $\sigma_1^{i,\xi}$ uses deterministic updates of memory, and uses integers as memory states (we think of the memory as an integer counter). The memory update function is given by

$$\sigma_1^{i,u}(a,j) = \begin{cases} j+1 & \text{if } a = \mathbf{L} \\ j-1 & \text{if } a = \mathbf{R} \end{cases}$$

and the action function is given by

$$\sigma_1^{i,a}(j)(\mathbf{R}) = \begin{cases} \xi^4(1-\xi)^{i+j} & \text{if } i+j > 0 \\ \xi^4 & \text{if } i+j \le 0 \end{cases}$$

The complete strategy We are now ready to define σ_1^*. The memory states of this strategy are 5-tuples $(i,j,k,\ell,b) \in \mathbb{Z}_+ \times \mathbb{Z}_+ \times \mathbb{Z} \times \mathbb{N} \times \{0,1\}$. Here i denotes the current epoch and j denotes the current sub-epoch of epoch i. The number of samples already made in the current sub-epoch is k. The memory state of the inner strategy is stored as ℓ. Finally b is 1 if and only if the strategy will sample to the inner strategy in the next step.

The memory update function $\sigma_1^{*,u}$ is as follows. Let (i,j,k,ℓ,b) be the current memory state and let a be the action of Player 2 in the current step. We then describe the distribution of the next memory state (i',j',k',ℓ',b').

- The current step is not sampled if $b=0$. In that case we keep $i'=i$, $j'=j$, $k'=k$, and $\ell'=\ell$.
- The current epoch is ending if $j=i$, $k=i^2-1$, and $b=1$. In that case $i'=i+1$, $j'=1$, $k'=0$, and $\ell'=0$.
- A sub-epoch is ending within the current epoch if $j<i$, $k=i^2-1$, and $b=1$. In that case $i'=i$, $j'=j+1$, $k'=0$, and $\ell'=0$.
- We sample within a sub-epoch if $k<i^2-1$ and $b=1$. In that case $i'=i$, $j'=j$, $k'=k+1$, and $\ell'=\sigma_1^{i,u}(a,\ell)$.

Finally, in every case, we make a probabilistic choice whether to sample in the next step by letting $b'=1$ with probability $\frac{(i')^3}{F(i')}$.

The action function $\sigma_1^{*,a}$ is given by

$$\sigma_1^{*,a}((i,j,k,\ell,b))(a) = \begin{cases} \sigma_1^{i,\xi}(j)(a) & \text{if } b=1 \\ 1 & \text{if } b=0 \text{ and } a=\mathbf{L} \\ 0 & \text{otherwise} \end{cases}.$$

In other words, if the current step is sampled, Player 1 follows the current base strategy, and otherwise always plays $\mathbf{L}$.

The starting memory state is $m_s = (1,1,1,0,0)$. The states that can be reached in sub-epoch j of epoch i are of the form (i,j,k,ℓ,b) where $0 \le k < i^2$ and $-i^2 < \ell < i^2$. Thus at most $4i^4$ states can be reached. The states are mapped to the natural numbers as follows: The memory $(1,1,1,0,0)$ is mapped to 0 and for each epoch i, all states in epoch i are mapped to the numbers (in an arbitrary order) following the numbers mapped to by epoch $i-1$.

Proof preliminaries It will be useful to consider the strategy modified to never stop. Thus denote by $\widetilde{\sigma}_1$ the strategy for Player 1, where $\widetilde{\sigma}^u = \sigma_1^{*,u}$ and $\widetilde{\sigma}_1^a(m)(\mathbf{L}) = 1$ for all memory states m.

We next define random variables indicating the locations of the sample steps. Fix some strategy σ_2 for Player 2. Let $\mathcal{M}_{m_s}^{\tilde{\sigma}_1,\sigma_2}$ be the memory sequence assigned to Player 1 when Player 1 follows $\tilde{\sigma}_1$ and Player 2 follows σ_2. For positive integers i, j, k let $t(i, j, k)$ be the random variable indicating the round in which we sample the k'th time in sub-epoch j of epoch i. For simplicity of notation we let $t(i, 0, i^2)$ denote $t(i-1, i-1, (i-1)^2)$, and we let $t(i, j) = t(i, j, 0)$ denote $t(i, j-1, i^2)$.

3.1 Space usage of the strategy

We will here consider the space usage of σ_1^*. First we will argue that with high probability, for all large enough i and any j, the length of sub-epoch j of epoch i, $t(i, j, i^2) - t(i, j-1, i^2)$, is close to $F(i)$.

Lemma 7. *For any $\gamma, \delta \in (0, 1/4)$, there is a constant M such that with probability at least $1-\gamma$, for all $i \geq M$ and all $j \in \{1, \ldots, i\}$, we have that*

$$t(i, j, i^2) - t(i, j-1, i^2) \in [(1-\delta)F(i)/i, (1+\delta)F(i)/i] \ .$$

Proof. The expected number of times we sample during $(1-\delta)F(i)/i$ steps of the i'th epoch is $(1-\delta)i^2$. If $t(i, j, i^2) - t(i, j-1, i^2) < (1-\delta)F(i)/i$ then we sampled at least i^2 times during these $(1-\delta)F(i)/i$ steps of epoch i. This means that the actual number of samples is larger than its expectation by a factor $\delta/(1-\delta)$. Thus by the multiplicative Chernoff bound, Theorem 45, we see that,

$$\Pr[t(i, j, i^2) - t(i, j-1, i^2) < (1-\delta)F(i)/i] < \exp\left(-c(1-\delta)i^2\right) \ ,$$

where $c = (\frac{\delta}{1-\delta})^2/(2 + \frac{\delta}{1-\delta})$. Similarly, if $t(i, j, i^2) - t(i, j-1, i^2) > (1+\delta)F(i)/i$ then we sampled less than i^2 times during $(1+\delta)F(i)/i$ steps which is less than the expected by a factor $\delta/(1+\delta)$ of its expectation. Again, by the multiplicative Chernoff bound,

$$\Pr[t(i, j, i^2) - t(i, j-1, i^2) > (1+\delta)F(i)/i] < \exp\left(-\frac{\delta^2}{2}(1+\delta)i^2\right) \ .$$

Thus for any M, we can bound from above the probability of any of the differences for $i \geq M$ being outside of the required range by:

$$\sum_{i=M}^{\infty} i \cdot \left(\exp\left(-c(1-\delta)i^2\right) + \exp\left(-\frac{\delta^2}{2}(1+\delta)i^2\right)\right) \ .$$

This sum is convergent so for sufficiently large M it can be bounded by γ. The lemma follows. $\square$

Now we bound the space usage of the strategy σ_1^*.

Lemma 8. *For all constants $\gamma > 0$, with probability at least $1-\gamma$, the space usage of σ_1^* is $O(\log f(T))$.*

Proof. Recall that there are at most $4 \cdot i^4$ distinct possible memory states reachable during the j'th sub-epoch of epoch i, for all i, j. Thus, since there are i sub-epochs in epoch i there are at most $4 \cdot i^5$ distinct possible memory states reachable during the i'th epoch. It is then clear that there are at most $\sum_{r=0}^{i} 4 \cdot r^5 \leq 4 \cdot i^6$ distinct possible memory states that can have been reached before the end of epoch i, for all i. Because memory states in earlier epochs are mapped to smaller numbers than latter epochs, we have that the strategy has not been in any state above that of $4 \cdot i^6$ before the end of epoch i.

Fix some $\gamma > 0$. By Lemma 7, with probability at least $1 - \gamma$, there is a M such that for all $i \geq M$, the number $t(i, i, i^2) - t(i, 1, i^2)$ is greater than $\frac{F(i)}{2}$. Thus also $t(i, i, i^2) \geq \frac{F(i)}{2}$. Consider any round T, for $T \geq t(M, M, M^2)$. Let i be the epoch containing round T. By the preceding we know that, before time step $t(i+1, i+1, (i+1)^2)$ (which is greater than T), the strategy have only been in memory states below $4 \cdot (i+1)^6$. We also have that $T \geq t(i, i, i^2) \geq \frac{F(i)}{2}$ and hence $2\overline{f}(T) \geq \overline{f}(2T) \geq i$, where the first inequality is by our assumption on F. Thus the strategy can only have been in states below that of $4 \cdot (2 \cdot \overline{f}(T) + 1)^6$ before time step T. This is true for all sufficiently large T and thus the strategy uses at most $O(\log f(T))$ space with probability $1 - \gamma$. $\square$

3.2 Play stopping implies good outcome

We first establish some properties of the base strategy $\sigma_1^{i,\xi}$. The proof of these uses ideas similar to proofs by Blackwell and Ferguson [2] and Kohlberg [11], where they showed ε-optimality of their strategies.

Lemma 9. *Let $T, i \geq 1$ be integers and $0 < \xi < 1$ be a real number. Let $\sigma \in \{\boldsymbol{L}, \boldsymbol{R}\}^T$ be a arbitrary prefix of a pure Markov strategy for Player 2. Consider the first T rounds where the players play the Big Match following $\sigma_1^{i,\xi}$ and σ respectively. Let p_{win} be the probability that Player 1 stops the game (i.e. plays $\boldsymbol{R}$) and wins. Let p_{loss} be the probability that Player 1 stops the game and loses. Then we have:*

1. $$p_{\text{loss}} \leq (1-\xi)^i \xi^3 + (1-\xi)^{-1} p_{\text{win}} \enspace .$$

2. *For any $0 < \delta \leq \frac{1}{2}$ and any $T > i/(2\delta)$, if* $\text{dens}(\sigma) \leq \frac{1}{2} - \delta$ *then*
$$p_{\text{win}} + p_{\text{loss}} \geq \xi^4 \enspace .$$

Proof. Define $d_\ell = |\{\ell' < \ell \mid \sigma_{\ell'} = \mathbf{L}\}| - |\{\ell' < \ell \mid \sigma_{\ell'} = \mathbf{R}\}|$. Note that d_ℓ is precisely the value of the counter used by $\sigma_1^{i,\xi}$ as memory in step ℓ. For integer d, define

$$K_d = \{\ell \in \{1, \ldots, T\} \mid (d_\ell = d \,\&\, \sigma_\ell = \mathbf{L}) \text{ or } (d_\ell = d+1 \,\&\, \sigma_\ell = \mathbf{R})\} \enspace .$$

There is an illustration of how d_ℓ could evolve through the steps in Figure 2. The set K_d is then the times the counter moves between the pair of rows d and $d+1$. Observe that the counter is alternately moving up and down in each K_d (for instance, in the gray row, the arrows are wider and first moves up, then down and then up again).

Notice that K_d partitions $\{1, \ldots, T\}$. Let $p_{\text{loss},d}$ be the probability that the game stops at some step $\ell \in K_d$ where $\sigma_\ell = \mathbf{L}$, and let $p_{\text{win},d}$ be the probability that the game stops at some step $\ell \in K_d$ where $\sigma_\ell = \mathbf{R}$. We see that p_{loss} is the sum of $p_{\text{loss},d}$, and that p_{win} is the sum of $p_{\text{win},d}$. Let

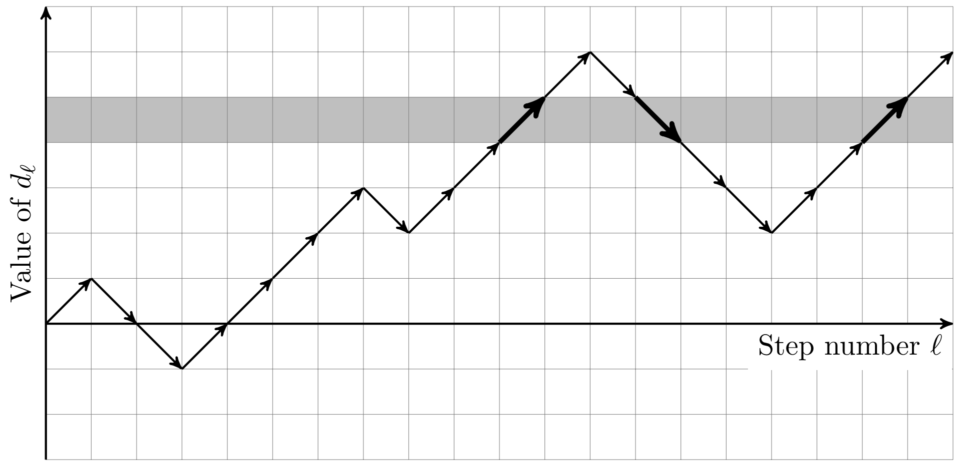

Figure 2: Possible movement of the counter $\sigma_1^{i,\xi}$ uses as memory through the steps.

$K_d = \{k_1 < k_2 < \cdots < k_m\}$, for some m. Observe that for any j, $\sigma_{k_j} = \mathbf{L}$ if and only if $\sigma_{k_{j+1}} = \mathbf{R}$, so $\sigma_{k_1}, \sigma_{k_2}, \ldots, \sigma_{k_m}$ is an alternating sequence (as mentioned in relation to the illustration) starting with $\mathbf{L}$ when $d \geq 0$ and starting with $\mathbf{R}$ otherwise. For any $d \geq -i$, the probability that the game stops in round $k_j \in K_d$, conditioned on the event that it did not stop before round k_j, is $\xi^4(1-\xi)^{i+d_{k_j}}$ so, it is $\xi^4(1-\xi)^{i+d}$ when $\sigma_{k_j} = \mathbf{L}$, and it is $\xi^4(1-\xi)^{i+d+1}$ when $\sigma_{k_j} = \mathbf{R}$.

Hence, for $d \geq 0$, $p_{\text{loss},d} \leq \xi^4(1-\xi)^{i+d} + (1-\xi)^{-1}p_{\text{win},d}$. This is because we stop at k_1 with probability at most $\xi^4(1-\xi)^{i+d}$ (in which case Player 1 loses) and then for each even j, the probability of stopping at step k_j (and winning) is at least $(1-\xi)$-times the probability of stopping at step k_{j+1} (and losing). (Indeed, the probability of stopping at step k_{j+1} might be even substantially smaller as the probability of stopping between k_j and k_{j+1} might be non-zero.)

For $d \in \{-i, -i+1, \ldots, -1\}$, $p_{\text{loss},d} \leq (1-\xi)^{-1}p_{\text{win},d}$ as for each *odd* j, the probability of stopping at step k_j (and winning) is at least $(1-\xi)$-times the probability of stopping at step k_{j+1} (and losing).

Finally, for $d < -i$, $p_{\text{loss},d} \leq p_{\text{win},d}$, as for each odd j, the probability of stopping at step k_j (and winning) is at least the probability of stopping at step k_{j+1} (and losing). (The probability of stopping at any such k_j conditioned on not stopping sooner is ξ^4 in this case.)

Hence,

$$\begin{aligned} p_{\text{loss}} = \sum_d p_{\text{loss},d} &\leq \sum_{d\geq 0} \xi^4(1-\xi)^{i+d} + (1-\xi)^{-1}\sum_d p_{\text{win},d} \\ &\leq \xi^3(1-\xi)^i + (1-\xi)^{-1}p_{\text{win}}. \end{aligned}$$

For the second part, if the density of $\sigma_1, \ldots, \sigma_T$ is at most $\frac{1}{2} - \delta$ then it must contain at most $T/2 - \delta T < T/2 - i/2$ occurrences of the letter $\mathbf{L}$. Hence, it contains more than $t/2 + i/2$ occurrences of the letter $\mathbf{R}$. This implies that when the game reaches round T, we have that the memory state j (recalling that the memory states are integers and in any round corresponds to the difference between the number of times Player 2 has played $\mathbf{R}$ minus the time he played $\mathbf{L}$ up till now) is such that $j \leq -i$ and hence Player 1 plays $\mathbf{R}$ at step t with probability ξ^4 if the play did not stop, yet. $\square$

We can now prove the main statement of this subsection, that if the probability of stopping is not too small, Player 1 wins with probability close to 1/2 if play the stops.

Lemma 10. *Let $\xi \in (0,1)$. Let σ be an pure Markov strategy for Player 2. If the probability that play stops is at least $\sqrt{\xi}$, then we have that*

$$\Pr^{\sigma_1^*,\sigma}[\text{ Player 1 wins } \mid \text{ play stops }] \geq \frac{1}{2} - \sqrt{\xi} \ .$$

Proof. We will in this proof continue playing in a state v, even if Player 1 plays **R**. Note that σ_1^* can still generate a choice in this case. Let $A_{i,j}$ be set of plays in which, between round 1 and round $t(i,j,1)-1$, Player 1 does not play **R**. Let $W_{i,j}$ be the set of plays, in which, between round $t(i,j,1)$ and round $t(i,j,i^2)$, Player 1 plays **R** and when he plays **R** for the first time between these rounds, Player 2 plays **R** as well. Similarly, let $L_{i,j}$ be the set of plays, in which, between round $t(i,j,1)$ and round $t(i,j,i^2)$, Player 1 plays **R** and when he does so for the first time Player 2 plays **L**. Let S be the set of plays, in which Player 1 plays **R** in some round. Let W be the set of plays, in which Player 1 plays **R** in some round and the first time he does so Player 2 also plays **R**. Let L be the set of plays, in which Player 1 plays **R** in some round and the first time he does so Player 2 plays **L**. We see that $S = W \cup L$. Clearly, $\Pr^{\sigma_1^*,\sigma}[W] = \sum_{i,j} \Pr^{\sigma_1^*,\sigma}[W_{i,j} \ \& \ A_{i,j}]$ and $\Pr^{\sigma_1^*,\sigma}[L] = \sum_{i,j} \Pr^{\sigma_1^*,\sigma}[L_{i,j} \ \& \ A_{i,j}]$.

Fix a possible value of all $t(i,j,k)$'s and denote by Y the event that these particular values actually occurs. Fix i and j. Conditioned on Y and $A_{i,j}$, between time $t(i,j,1)$ and $t(i,j,i^2)$ Player 1 plays $\sigma_1^{i,\xi}$ against a fixed strategy $\sigma_{t(i,j,1)}, \sigma_{t(i,j,2)}, \ldots, \sigma_{t(i,j,i^2)}$ for Player 2. By Lemma 9, the probability of losing such a game for Player 1 is at most $\xi^3(1-\xi)^i$ plus the probability of winning in this game divided by $(1-\xi)$. Hence,

$$\Pr[L_{i,j} \mid Y, A_{i,j}] \leq \xi^3(1-\xi)^i + (1-\xi)^{-1}\Pr[W_{i,j} \mid Y, A_{i,j}].$$

Since the above inequality is true conditioned on arbitrary values of $t(i,j_1,j_2)$'s, it is true also without the conditioning:

$$\Pr[L_{i,j} \mid A_{i,j}] \leq \xi^3(1-\xi)^i + (1-\xi)^{-1}\Pr[W_{i,j} \mid A_{i,j}].$$

Thus,

$$\begin{aligned}
\Pr^{\sigma_1^*,\sigma}[L] &= \sum_{i=1}^{\infty}\sum_{j=1}^{i} \Pr[L_{i,j} \ \& \ A_{i,j}] \\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{i} \Pr[L_{i,j} \mid A_{i,j}] \cdot \Pr[A_{i,j}] \\
&\leq \sum_{i=1}^{\infty}\sum_{j=1}^{i} \Big(\xi^3(1-\xi)^i + (1-\xi)^{-1}\Pr^{\sigma_1^*,\sigma}[W_{i,j} \mid A_{i,j}]\Big) \cdot \Pr[A_{i,j}] \\
&\leq \sum_{i=1}^{\infty}\sum_{j=1}^{i} \xi^3(1-\xi)^i + (1-\xi)^{-1}\sum_{i=1}^{\infty}\sum_{j=1}^{i}\Pr^{\sigma_1^*,\sigma}[W_{i,j} \ \& \ A_{i,j}] \\
&= \sum_{i=1}^{\infty} i \cdot \xi^3(1-\xi)^i + (1-\xi)^{-1}\sum_{i=1}^{\infty}\sum_{j=1}^{i}\Pr^{\sigma_1^*,\sigma}[W_{i,j} \ \& \ A_{i,j}] \\
&\leq \xi + (1-\xi)^{-1}\Pr^{\sigma_1^*,\sigma}[W] \ .
\end{aligned}$$

Hence, $(1-\xi)\Pr^{\sigma_1^*,\sigma}[L] \leq \xi - \xi^2 + \Pr^{\sigma_1^*,\sigma}[W]$, and so $\Pr^{\sigma_1^*,\sigma}[L] - \Pr^{\sigma_1^*,\sigma}[W] \leq 2\xi$. By our assumption we also have that $\Pr^{\sigma_1^*,\sigma}[L] + \Pr^{\sigma_1^*,\sigma}[W] = \Pr^{\sigma_1^*,\sigma}[S] \geq \sqrt{\xi}$.

We want to find the minimum probability that we might win, conditioned on us stopping

$$\Pr^{\sigma_1^*,\sigma}[W \mid S] = \frac{\Pr^{\sigma_1^*,\sigma}[W]}{\Pr^{\sigma_1^*,\sigma}[S]} = \frac{\Pr^{\sigma_1^*,\sigma}[W]}{\Pr^{\sigma_1^*,\sigma}[W] + \Pr^{\sigma_1^*,\sigma}[L]}$$

We see that it is greater than the solution of:

$$\begin{array}{lll} \min & \frac{x}{x+y} & \\ \text{s.t.} & x+y & \geq \sqrt{\xi} \\ & y-x & \leq 2\xi \end{array}$$

Solving the above, we see that $\Pr^{\sigma_1^*,\sigma}[W \mid S] \geq \frac{1}{2} - \sqrt{\xi}$ and the lemma follows. □

3.3 Low density means play stops

In this subsection we will prove that play will stop with probability 1 if the density of prefixes of the pure Markov strategy used by Player 2 is not infinitely often at least 1/2. First, we will show that low density implies that some sequence of sub-epochs also have low density.

Lemma 11. *Let $\delta \in (0,1)$. Let σ be an arbitrary pure Markov strategy for Player 2. Let $a_{i,j}$ be some numbers. Consider the event Y where $t(i,j) = a_{i,j}$ for all i, j. If $\limsup_{T\to\infty} \operatorname{dens}(\sigma^T) \leq 1/2 - \delta$, then conditioned on Y, there is an infinite sequence of sub-epochs and epochs $(i_n, j_n)_n$ such that* $\operatorname{dens}(\sigma, a_{i_n,j_n} + 1, a_{i_n,(j_n+1)}) \leq 1/2 - \delta/4$.

Proof. Let M be such that for every $T' \geq M$ we have that $\operatorname{dens}(\sigma^{T'}) \leq 1/2 - \delta/2$. Let $(T_n)_n$ be a sequence such that $T_1 \geq M$ and for all $n \geq 1$ we have that $T_{n+1} \cdot \delta/4 \geq T_n$ and $T_n = a_{i,j}$ for some i, j. Let $(i'_n, j'_n)_n$ be the sequence such that $T_n = a_{i'_n,j'_n}$. This means that even if $\operatorname{dens}(\sigma^{T_n}) = 0$, the density $\operatorname{dens}(\sigma, T_n + 1, T_{n+1})$ is at most $1/2 - \delta/4$, because $\operatorname{dens}(\sigma^{T_{n+1}}) \leq 1/2 - \delta/2$. But, we then get that there exists some sub-epoch j_n in epoch i_n, such that $j'_n \leq j_n \leq j'_{n+1}$ and such that $i'_n \leq i_n \leq i'_{n+1}$ for which the density of that sub-epoch $\operatorname{dens}(\sigma, a_{i_n,j_n} + 1, a_{i_n,(j_n+1)})$ is at most $1/2 - \delta/4$, because not all sub-epochs can have density below that of the average sub-epoch. But then $(i_n, j_n)_n$ satisfies the lemma statement. □

We are now ready to prove the main statement of this subsection.

Lemma 12. *Let σ be an arbitrary pure Markov strategy for Player 2. If*

$$\limsup_{T\to\infty} \operatorname{dens}(\sigma^T) < 1/2 \ ,$$

then when played against σ_1^* the play stops with probability 1.

Proof. Let $\delta > 0$ be such that $\limsup_{T\to\infty} \operatorname{dens}(\sigma^T) \leq 1/2 - \delta$. Consider arbitrary numbers $a_{i,j}$ and the event Y stating that $t(i,j) = a_{i,j}$ for all i, j. Let $(i_n, j_n)_n$ be the sequence of sub-epochs and epochs shown to exists by Lemma 11 with probability 1. That is, for each (i_n, j_n) we have that sub-epoch j_n of epoch i_n has density at most $1/2 - \delta/4$. We see that, conditioned on Y that each sample are sampled uniformly at random in each sub-epoch j of each epoch i, except for the last sample. Now consider some fixed n. By Hoeffding's inequality, Theorem 46 (setting $a_i = 0$ and

$b_i = 1$, and letting c_i be i'th payoff for Player 1), the probability that in sub-epoch j_n of epoch i_n that among our $i_n^2 - 1$ first samples we have $\frac{\delta}{8} \cdot ((i_n)^2 - 1)$ additional $\mathbf{L}$ on top of the expectation (which is at most $(1/2 - \delta/4) \cdot ((i_n)^2 - 1)$) is bounded as

$$\Pr\left[\mathrm{dens}(\sigma_{t(i_n,j_n,1)}, \sigma_{t(i_n,j_n,2)}, \ldots, \sigma_{t(i_n,j_n,(i_n)^2-1)}) \geq \frac{1}{2} - \frac{\delta}{8}\right]$$
$$\leq 2\exp\left(-\frac{2(\frac{\delta}{8}((i_n)^2-1))^2}{(i_n)^2-1}\right) = 2\exp\left(-\frac{\delta^2}{32}((i_n)^2-1)\right) .$$

For sufficiently large i_n, this is less than $\frac{1}{2}$. If on the other hand the number of $\mathbf{L}$'s we sample is less than $(\frac{1}{2} - \frac{\delta}{8})((i_n)^2 - 1)$, we see that $(i_n)^2 - 1 > i_n/(16\delta)$ for large enough i_n and in that case we have, by Lemma 9, that we stop with probability at least ξ^4 in sub-epoch j_n of epoch i_n. Thus, for each of the infinitely many n's for which i_n is sufficiently high, we have a probability of at least $\frac{\xi^4}{2}$ of stopping. Thus play must stop with probability 1.

The argument was conditioned on some fixed assignment of endpoints of sup-epochs and epochs, but since there is such a assignment with probability 1 (since they are finite with probability 1), we conclude that the proof works without the condition. □

3.4 Proof of main result

Theorem 13. *The strategy σ_1^* is $\sqrt{\xi}$-supremum-optimal, and for all $\delta > 0$, with probability at least $1 - \delta$ it uses space $O(\log f(T))$.*

Proof. The space usage follows from Lemma 8. Let s be the probability that the play stops. We now consider three cases, either (i) $s = 1$; or (ii) $\sqrt{\xi} < s < 1$; or (iii) $s \leq \sqrt{\xi}$. In case (i), if $s = 1$, Player 1 wins with probability $\frac{1}{2} - \sqrt{\xi}$, by Lemma 10. In case (ii), if $\sqrt{\xi} < s < 1$, then, by Lemma 10, conditioned on the play stopping, Player 1 wins with probability $W_s = \frac{1}{2} - \sqrt{\xi}$ and, since $s < 1$, by Lemma 12 $\widetilde{W_s} = \limsup_{T\to\infty} \mathrm{dens}(\sigma^T) \geq \frac{1}{2}$. Thus, Player 1 wins with probability

$$s \cdot W_s + (1-s) \cdot \widetilde{W_s} \geq s \cdot (\frac{1}{2} - \sqrt{\xi}) + (1-s)\frac{1}{2} \geq \frac{1}{2} - \sqrt{\xi}$$

In case (iii), if $s \leq \sqrt{\xi}$, then by Lemma 12 $\widetilde{W_s} = \limsup_{T\to\infty} \mathrm{dens}(\sigma^T) \geq \frac{1}{2}$. The winning probability is then at least

$$s \cdot 0 + (1-s) \cdot \frac{1}{2} \geq \frac{1-\sqrt{\xi}}{2} \geq \frac{1}{2} - \sqrt{\xi}$$

□

4 An ε-optimal strategy that uses $\log\log T$ space for the Big Match

For given $\varepsilon > 0$, let $\xi = \varepsilon^2$. In this section we give a ε-optimal strategy σ_1^* for Player 1 in the Big Match that for all $\delta > 0$ with probability $1 - \delta$ uses $O(\log\log T)$ space. The strategy is simply an instantiation of the strategy σ_1^* from Section 3, setting $f(T) = \lceil \log T \rceil$. We can then let $\overline{f} = \log T$ and $F(T) = 2^T$.

The claim about the space usage of σ_1^* is thus already established in Section 3. To obtain the stronger property of ε-optimality rather than just ε-supremum-optimality, we just need to establish a lim inf version of Lemma 12.

First we show a technical lemma. For a pure Markov strategy σ and a sequence of integers $I = \{i_1, i_2, \ldots, i_m\}$, let σ_I be the sequence, $\sigma_{i_1}, \sigma_{i_2}, \ldots, \sigma_{i_m}$. Note that $\sigma^k = \sigma_{\{1,\ldots,k\}}$.

Lemma 14. *Let σ be a pure Markov strategy for Player 2, $\delta < 1/4$ be a positive real, and M be a positive integer. Let $\liminf_{T\to\infty} \mathrm{dens}(\sigma^T) \le \frac{1}{2} - \delta$. Let $\ell_1, \ell_2, \ldots$ be such that for all $i \ge M$, we have that $\ell_i \in [(1-\delta)\cdot(2^{i+1}-1), (1+\delta)\cdot(2^{i+1}-1)]$. Then there exists a sequence $k_2, k_3, \ldots$ such that for infinitely many $i > M$, we have that $\ell_{i-1} + \delta 2^{i-2} \le k_i \le \ell_i$ and $\mathrm{dens}(\sigma_{\{\ell_{i-1}+1,\ldots,k_i\}}) \le \frac{1}{2} - \frac{\delta}{4}$.*

Proof. Let ℓ_i be as required. If there are infinitely many i such that $\mathrm{dens}(\sigma_{\{\ell_{i-1}+1,\ldots,\ell_i\}}) \le \frac{1}{2} - \frac{\delta}{4}$ then set $k_i = \ell_{i+1}$ and the lemma follows by observing $(k_i - \ell_{i-1}) \ge (1-\delta)(2^{i+1}-1) - (1+\delta)(2^i - 1) = (1-3\delta)2^i \ge \delta 2^{i-2}$, for $i > M$. So assume that only for finitely many i, $\mathrm{dens}(\sigma_{\{\ell_{i-1}+1,\ldots,\ell_i\}}) \le \frac{1}{2} - \frac{\delta}{4}$. Thus the following claim can be applied for arbitrary large i_0.

Claim 15. *Let $i_0 \ge M$ be given. If for every $i \ge i_0$, $\mathrm{dens}(\sigma_{\{\ell_{i-1}+1,\ldots,\ell_i\}}) > \frac{1}{2} - \frac{\delta}{4}$ then there exist $j > i_0$ and k such that $\ell_{j-1} + \delta 2^{j-2} \le k \le \ell_j$ and $\mathrm{dens}(\sigma_{\{\ell_{j-1}+1,\ldots,k\}}) \le \frac{1}{2} - \delta$.*

We can use the claim to find $k_2, k_3, \ldots$ inductively. Start with large enough $i_0 \ge M$ and set $k_i = \ell_i$ for all $i \le i_0$. Then provided that we already inductively determined $k_2, k_3, \ldots, k_{i_0}$, we apply the above claim to obtain j and k, and we set $k_j = k$ and $k_i = \ell_i$, for all $i = i_0 + 1, \ldots, j-1$.

So it suffices to prove the claim. For any $d \ge 1$, $\ell_{i_0} \cdot 2^{d-1} \le \ell_{i_0+d}$ and

$$\mathrm{dens}(\sigma^{\ell_{i_0+d}}) \ge \frac{(\frac{1}{2} - \frac{\delta}{4})(\ell_{i_0+d} - \ell_{i_0})}{\ell_{i_0+d}} \ .$$

Furthermore, if $d \ge 1 + \log(4/\delta)$ then $\ell_{i_0} \le \frac{\delta}{4}\ell_{i_0+d}$ and

$$\mathrm{dens}(\sigma^{\ell_{i_0+d}}) \ge \left(\frac{1}{2} - \frac{\delta}{4}\right) - \frac{\delta}{4} = \frac{1}{2} - \frac{\delta}{2} \ .$$

Since $\liminf_{k\to\infty} \mathrm{dens}(\sigma^k) \le \frac{1}{2} - \delta$, there must be k and $d \ge 1 + \log(4/\delta)$ such that $\ell_{i_0+d-1} \le k \le \ell_{i_0+d}$ and $\mathrm{dens}(\sigma^k) \le \frac{1}{2} - \delta$. Set $j = i_0 + d$. Also

$$\mathrm{dens}(\sigma^k) = \frac{\mathrm{dens}(\sigma^{\ell_{j-1}})\ell_{j-1} + \mathrm{dens}(\sigma_{\{\ell_{j-1}+1,\ldots,k\}})(k - \ell_{j-1})}{\ell_{j-1} + (k - \ell_{j-1})} \ ,$$

which means

$$\begin{aligned}\left(\mathrm{dens}(\sigma^k) - \mathrm{dens}(\sigma_{\{\ell_{j-1}+1,\ldots,k\}})\right)(k - \ell_{j-1}) &= \left(\mathrm{dens}(\sigma^{j-1}) - \mathrm{dens}(\sigma^k)\right)\ell_{j-1} \\ &\ge \left[\left(\frac{1}{2} - \frac{\delta}{2}\right) - \left(\frac{1}{2} - \delta\right)\right]\ell_{j-1} = \frac{\delta}{2}\ell_{j-1} \ .\end{aligned}$$

Thus $\mathrm{dens}(\sigma_{\{\ell_{j-1}+1,\ldots,k\}}) \le \mathrm{dens}(\sigma^k)$ which in turn is less than $\frac{1}{2} - \delta$. Furthermore, $k - \ell_{j-1} \ge \frac{\delta}{2}\ell_{j-1} \ge \frac{\delta}{2}(1-\delta)(2^j - 1) \ge \frac{\delta}{4}2^j$, provided that $j \ge 2$. Hence, k and j have the desired properties. □

We are now ready to prove the lim inf version of Lemma 12.

Lemma 16. *Let σ be an arbitrary pure Markov strategy for Player 2. If*

$$\liminf_{t\to\infty} \mathrm{dens}(\sigma_1, \ldots, \sigma_t) < 1/2 \ ,$$

then when played against σ_1^* the play stops with probability 1.

Proof. Let $0 < \delta < \frac{1}{4}$ be such that

$$\liminf_{t\to\infty} \mathrm{dens}(\sigma_1, \ldots, \sigma_t) \leq 1/2 - \delta$$

Pick arbitrary $\gamma \in (0, 1)$. We will show that with probability at least $1 - \gamma$ the game stops, and this implies the statement. Let M be given by Lemma 7 applied for γ and $\delta/2$. Then we have that with probability at least $1 - \gamma$, for all $i \geq M$ and $j \in \{1, \ldots, i\}$,

$$t(i, j, i^2) - t(i, j-1, i^2) \in [(1 - \frac{\delta}{2})2^i/i, (1 + \frac{\delta}{2})2^i/i] \ .$$

Pick $t_{i,j} \in \mathbb{N}$, for $i = 1, 2, \ldots$ and $j \in \{1, \ldots, i\}$, so that $t_{i,j-1} < t_{i,j}$ where $t_{i,0}$ stands for $t_{i-1,i-1}$. Let $t_{i,j} - t_{i,j-1} \in [(1 - \frac{\delta}{2})2^i/i, (1 + \frac{\delta}{2})2^i/i]$, for all $i \geq M$ and $j \in \{1, \ldots, i\}$. Pick M' so that $\frac{\delta}{2}(2^{M'} - 1) \geq \max\{t_{M,0}, (1 - \frac{\delta}{2})2^M\}$. Define $\ell_i = t_{i,i}$ for all $i \geq 1$. Then for all $i \geq M'$, $\ell_i \in [(1 - \delta) \cdot (2^{i+1} - 1), (1 + \delta) \cdot (2^{i+1} - 1)]$ as

$$\begin{aligned}
\ell_i = t_{i,i} &= t_{M,0} + \sum_{M \leq i' \leq i, 1 \leq j \leq i'} t_{i',j} - t_{i',j-1} \\
&\leq \frac{\delta}{2}(2^{M'} - 1) + \sum_{M \leq i' \leq i} i' \cdot (1 + \frac{\delta}{2})2^{i'}/i' \\
&\leq \frac{\delta}{2}(2^{M'} - 1) + (1 + \frac{\delta}{2}) \cdot (2^{i+1} - 1) \\
&\leq (1 + \delta) \cdot (2^{i+1} - 1),
\end{aligned}$$

and similarly for the lower bound: $\ell_i \geq \sum_{i',j} t_{i',j} - t_{i',j-1} \geq (1 - \frac{\delta}{2})(2^{i+1} - 2^M) \geq (1 - \delta) \cdot (2^{i+1} - 1)$. Thus Lemma 14 is applicable on ℓ_i with M set to M', and we obtain a sequence $k_2, k_3, \ldots$ such that $\mathrm{dens}(\sigma_{\ell_{i-1}+1,\ldots,k_i}) \leq \frac{1}{2} - \frac{\delta}{4}$ and $k_i - \ell_{i-1} \geq \delta 2^{i-2}$ for infinitely many i. Pick any of the infinitely many $i \geq \max\{M', 32(1 + \delta)/\delta\}$ for which $k_i - \ell_{i-1} \geq \delta 2^{i-2}$ and $\mathrm{dens}(\sigma_{\ell_{i-1}+1,\ldots,k_i}) \leq \frac{1}{2} - \frac{\delta}{4}$. Since $\delta 2^{i-3} \geq (1 + \delta)2^i/i$, there is some $j \in \{1, \ldots, i\}$ such that $\ell_{i-1} + \delta 2^{i-3} \leq k_i - (1 + \delta)2^i/i \leq t_{i,j} \leq k_i$. Fix such j. Since $k_i \leq t_{i,j} + (1 + \delta)2^i/i$, we have

$$\begin{aligned}
\mathrm{dens}(\sigma_{\ell_{i-1}+1,\ldots,t_{i,j}}) &= \frac{\mathrm{dens}(\sigma_{\ell_{i-1}+1,\ldots,k_i})(k_i - \ell_{i-1})}{t_{i,j} + \ell_{i-1}} \\
&\leq \frac{\mathrm{dens}(\sigma_{\ell_{i-1}+1,\ldots,k_i})((1 + \delta)2^i/i + t_{i,j} - \ell_{i-1})}{t_{i,j} + \ell_{i-1}} \\
&\leq \left(\frac{1}{2} - \frac{\delta}{4}\right) \cdot \left(1 + \frac{8(1 + \delta)}{i}\right) \\
&\leq \frac{1}{2} - \frac{\delta}{4} + \frac{4(1 + \delta)}{i} \leq \frac{1}{2} - \frac{\delta}{8}.
\end{aligned}$$

Hence, $\mathrm{dens}(\sigma_{\ell_{i-1}+1,\ldots,t_{i,j}}) \leq \frac{1}{2} - \frac{\delta}{8}$. So for some $j' \in \{1, \ldots, j\}$, $\mathrm{dens}(\sigma_{t_{i,j'-1}+1,\ldots,t_{i,j'}}) \leq \frac{1}{2} - \frac{\delta}{8}$. We can state the following claim.

Claim 17. *For i large enough, conditioned on $t(a, b, a^2) = t_{a,b}$, for all $a \geq M$ and all b, and conditioned on that the game did not stop before the time $t_{i,j'-1} + 1$, the game stops during times $t_{i,j'-1} + 1, \ldots, t_{i,j'}$ with probability at least $\xi^4/2$.*

Conditioned on $t(a,b,a^2) = t_{a,b}$, for all a, b, the claim implies that the game stops with probability 1. Note that the condition is true for some valid choice of $t_{a,b}$ with probability $1-\gamma$. This is because the claim can be invoked for infinitely many i's and for each such i we will have $\xi^4/2$ chance of stopping.

It remains to prove the claim. Assume $t(i,j'-1,i^2) = t_{i,j'-1}$ and $t(i,j',i^2) = t_{i,j'}$. Clearly,

$$\mathrm{dens}(\sigma_{t_{i,j'-1}+1,\ldots,t_{i,j'}-1}) \le \mathrm{dens}(\sigma_{t_{i,j'-1}+1,\ldots,t_{i,j'}}) \cdot (2^{i-1}/(2^{i-1}-1)) \le \frac{1}{2} - \frac{\delta}{16} \ ,$$

for i large enough. So if we sample i^2-1 times from $\sigma_{t_{i,j'-1}+1}, \ldots, \sigma_{t_{i,j'}-1}$ we expect at most $(\frac{1}{2} - \frac{\delta}{16})(i^2-1)$ of the letters to be **L**. By Hoeffding's inequality, Theorem 46 (again setting $a_i = 0$ and $b_i = 1$, and letting c_i be i'th payoff for Player 1), the probability that we get at least $\frac{\delta}{32}(i^2-1)$ additional **L** items on top of the expectation is given by

$$\begin{aligned} &\Pr\left[\mathrm{dens}(\sigma_{t(i,j',1)}, \sigma_{t(i,j',2)}, \ldots, \sigma_{t(i,j',i^2-1)}) \ge \frac{1}{2} - \frac{\delta}{32}\right] \\ &\le 2\exp\left(-\frac{2(\frac{\delta}{32}(i^2-1))^2}{i^2-1}\right) = 2\exp\left(-\frac{\delta^2}{512}(i^2-1)\right) \ . \end{aligned}$$

The probability is taken over the possible choices of $t(i,j',1) < t(i,j',2) < \cdots < t(i,j',i^2-1)$ assuming $t(i,j'-1,i^2) = t_{i,j'-1}$ and $t(i,j',i^2) = t_{i,j'}$. For i sufficiently large, $2e^{-\delta^2(i^2-1)/512} \le 1/2$. Whenever

$$\mathrm{dens}(\sigma_{t(i,j',1)}, \sigma_{t(i,j',2)}, \ldots, \sigma_{t(i,j',i^2-1)}) \le \frac{1}{2} - \frac{\delta}{32}$$

we have at least ξ^4 chance of stopping by Lemma 9, as Player 1 plays $\sigma_1^{i+j',\xi}$ against

$$\sigma_{t(i,j',1)}, \sigma_{t(i,j',2)}, \ldots, \sigma_{t(i,j',i^2-1)}$$

and $i^2 - 1 > i/\delta \ge (i+j')/2\delta$ for sufficiently large i. Hence, the game stops with probability at least $(1-1/2)\cdot\xi^4 = \xi^4/2$. The claim, and thus the lemma, follows. $\square$

We can now conclude with the main result of this section.

Theorem 18. *The strategy σ_1^* is $\sqrt{\xi}$-optimal, and for all $\delta > 0$, with probability at least $1-\delta$ it uses space $O(\log\log T)$.*

Proof. This is proved just like Theorem 13, except that Lemma 16 is used in place of Lemma 12. $\square$

We can also improve the strategy and get the following corollary.

Corollary 19. *For any natural number k, there is a strategy which is 2^{-k}-optimal, has patience 2 and can be implemented on a Turing machine, using at most 1 random bit and amortized constant time*[1] *per round and with probability at least $1-\delta$ does it use tape space $O(\log\log(T) + \log k)$ upto round T.*

[1] i.e. for all T, let $c(T)$ be the computation used for the first T rounds, then $\frac{c(T)}{T}$ is some constant

Proof. We will only sketch the necessary ingredients for the proof. We will modify the strategy σ_1^* as defined in Section 4 to use less patience. First, it is easy to see that the probability that two sample points are ever within a polynomial in i distance of each other is small, in any sufficiently high epoch i, and thus we can ignore this event. The idea is as follows: Each sub-epoch of epoch i is split into blocks of length $B = i^{O(1)} \cdot k$ (where we simply keep track of the current location within the current block, but not the current block number). The purpose of each block is to provide all the randomness needed for the next block. Hence, it needs to (1) decide if there is a sample point in the next block; and (2) if so where; and (3) if we should stop on that sample point. Observe that we are ignoring the possibility that there are more than 1 sample point in a block. It is easy to see that the number of random bits needed in total to generate all those events is less than $i^{O(1)} \cdot k$ and thus it can be done with at most 1 random bit per round, implying the patience bound. To get the tape space bound, we utilize the three following simple ideas:

Idea 1: To get a probability like $\prod a_j$, for some sequence of probabilities a_j of length ℓ, one can test if sub-events that happens with probability a_i for all i all happens. This requires only space for a counter counting up to ℓ and space for the event that uses the most space (by reusing the space for the event).

Idea 2: To get an probability like 2^{-x} (or similarly $1 - 2^{-x}$) one can simply flip x coins and if all comes up tails, then the event happens. This requires only space for a counter counting up to x.

Idea 3: To get an probability like $\frac{y}{2^x}$, for any natural number x and y, one can simply use x random bits. This uses x many bits of space.

Direct application of these three ideas suffice to get all three events in $O(\log i + \log k)$ many bits. Observe that we can do this easily on a two tape Turing machine (one is used for events following idea 3 and one for the constant number of counters) and we only need amortized constant time in each round (to increment and/or reset some subset of the counters on the first tape and perhaps to add one more random bit or reset the second tape). □

5 Lower bound on patience

When considering a strategy of a player one may want to look at how small or large the probabilities occurring in that strategy are. The parameter of interest is the *patience* of the strategy which is the reciprocal of the smallest non-zero probability occurring in the strategy. Patience is closely related to the expected length of finite plays as small probability events will not occur if the play is too short so they will have little influence on the overall outcome [11, 6]. Care has to be taken how to define patience for strategies with infinitely many possible events. One thing to note of our space efficient strategies is that the patience of the states in which we are with high probability during the first T steps is approximately T, for rounds T close to the end of an epoch. In this section we show that this is essentially necessary. So if the space used by the strategy with high probability is $\log f(T)$, then the first $f(T)$ states must have patience about T. Thus the smaller the space the strategy uses the larger the patience the states must have.

The main theorem of this section states that if the patience of the $f(T)$ states in which Player 1 is with high probability is less than about $T^{1/f(T)}$ then the strategy is bad for Player 1. It is easy to observe that events with probability substantially less than $1/T$ are unlikely to occur during the first T steps of the game so one should expect the patience of a good strategy for Player 1 to be in the range between roughly $T^{1/f(T)}$ and T.

One may wonder whether the exponent $1/f(T)$ in $T^{1/f(T)}$ is necessary. It turns out that it is, so our lower bound is close to optimal using a technique like in Corollary 19.

We use the following definitions to deal with the fact that our strategies use infinitely many transitions so their overall patience is infinite.

For a memory based strategy σ_1 of Player 1 in a repeated zero-sum game with absorbing states, the *patience* of a set of memory states M is defined as:

$$\operatorname{pat}(M) = \max \left\{ \frac{1}{\sigma_1^a(m)(a_1)}, \frac{1}{\sigma_1^u(a_1, a_2, m)(m')},\ m, m' \in M, a_1 \in A_1, a_2 \in A_2 \right\} .$$

The patience of the other player is defined similarly.

Theorem 20. *Let $\delta, \varepsilon > 0$ be reals and $f : \mathbb{N} \to \mathbb{N}$ be an unbounded non-decreasing function such that $f(T) \leq \frac{1}{4} \log_{1/\varepsilon} T$ for all large enough T. If a strategy σ_1 of Player 1 uses space $\log f(T)$ before time T with probability at least $1 - \delta$, and the patience of the set of lexicographically first $f(T)$ memory states is at most $T^{1/(2f(T))}$ for all T large enough, then there is a strategy σ_2 of Player 2 such that $u(\sup, (\sigma_1, \sigma_2)) \leq \delta + 2\varepsilon$.*

Proof. Assume that $\varepsilon < 1/2$ and pick an integer k sufficiently large. For $i > 0$, define $\ell_i = \varepsilon^{-i}$ and $T_i = \sum_{j=1}^{i} \ell_j$. The strategy σ_2 of Player 2 proceeds in phases, each phase i is of length ℓ_i. In the first k phases, Player 2 plays $\mathbf{L}$ with probability $1 - \varepsilon$ and $\mathbf{R}$ with probability ε. In each phase $i > k$, Player 2 plays $\mathbf{R}$ for the first $(1 - \varepsilon)\ell_i$ steps, and afterwards he plays $\mathbf{L}$ with probability $1 - \varepsilon$ and $\mathbf{R}$ with probability ε. Notice, if the game does not stop by Player 1 playing $\mathbf{R}$ at some point then the expected $\limsup$ payoff to Player 1 is at most ε.

Our goal is to show that if the game stops then the expected payoff to Player 1 is at most $\delta + 2\varepsilon$. If the game stops during the last $\varepsilon\ell_i$ steps of a phase i, then the expected payoff to Player 1 is ε as the probability of Player 2 playing $\mathbf{R}$ at that time is ε. If the game stops during the first $(1 - \varepsilon)\ell_i$ steps of a phase $i > k$, then the payoff of Player 1 is 1. Our goal is to argue that the overall probability that Player 1 stops during the first $(1 - \varepsilon)\ell_i$ steps of some phase $i > k$ is small.

For any $t > 0$, denote by $M(t)$ the set of the lexicographically first $f(t)$ memory states (i.e. those mapped to a number below $f(t)$. Let C be the event that for all steps t, Player 1 is in one of the states in $M(t)$. For $t < t'$, let $S(t, t')$ be the event that Player 1 plays $\mathbf{R}$ in one of the steps $[t, t')$. Let $A_i(t)$ be the event that at time t, Player 1 is in one of the states in $M(T_i)$ and there is some memory state in $M(T_i)$ that can be reached from the state current at time t and in which there is a non-zero probability of playing $\mathbf{R}$ (i.e., stopping).

The probability that C does not occur is at most δ so for the rest of the proof we will assume that C occurs. Let k be large enough, and $i > k$. It is clear that if $S(T_{i-1}, T_{i-1} + (1 - \varepsilon)\ell_i)$ occurs then $A_i(T_{i-1})$ must have occurred as well so:

$$\Pr[S(T_{i-1}, T_{i-1} + (1 - \varepsilon)\ell_i)] \leq \Pr[A_i(T_{i-1})] .$$

Furthermore, if $A_i(T_{i-1})$ occurs then $A_i(t)$ occurs for all $t < T_{i-1}$. For $t \in [T_{i-1} - \varepsilon\ell_{i-1}, T_{i-1} - f(T_i))$, if $A_i(t)$ occurs then within the next $f(T_i)$ steps the strategy of Player 1 might reach a state in which Player 1 chooses the stopping action $\mathbf{R}$ with non-zero probability. Because of the patience of $M(T_i)$

and the fact that Player 2 plays each of his possible actions with probability at least ε during that time steps we have for any $t \in [T_{i-1} - \varepsilon\ell_{i-1}, T_{i-1} - f(T_i))$,

$$\Pr[S(t, t + f(T_i)) \mid A_i(t)] \geq \left(\frac{\varepsilon}{\mathrm{pat}(M(T_i))}\right)^{f(T_i)} .$$

Hence, the probability that $A_i(T_{i-1})$ occurs and the game did not stop yet is at most:

$$\left(1 - \left(\frac{\varepsilon}{\mathrm{pat}(M(T_i))}\right)^{f(T_i)}\right)^{\varepsilon\ell_{i-1}/f(T_i)} \leq \mathrm{e}^{-\left(\frac{\varepsilon}{\mathrm{pat}(M(T_i))}\right)^{f(T_i)} \cdot \varepsilon\ell_{i-1}/f(T_i)}.$$

For sufficiently large T, $f(T) \leq \frac{1}{4}\log_{1/\varepsilon} T$. Furthermore, $T_i \leq 2\ell_i$ and $\varepsilon T_i - 1 \leq T_{i-1}$. Since i is sufficiently large we have:

$$\begin{aligned}
\left(\frac{\varepsilon}{\mathrm{pat}(M(T_i))}\right)^{f(T_i)} \cdot \varepsilon \cdot \frac{\ell_{i-1}}{f(T_i)} &\geq \left(\frac{\varepsilon}{T_i^{\frac{1}{2f(T_i)}}}\right)^{f(T_i)} \cdot \varepsilon \cdot \frac{T_{i-1}}{2f(T_i)} \\
&\geq \varepsilon^{f(T_i)} \cdot \varepsilon^2 \cdot \frac{T_i^{1/2}}{2f(T_i)} - 1 \\
&\geq \varepsilon^2 \cdot \frac{T_i^{1/4}}{2f(T_i)} - 1 \geq T_i^{1/5}.
\end{aligned}$$

Since $T_i^{1/5} \gg i$ for i large enough, we get

$$\Pr[S(T_{i-1}, T_{i-1} + (1-\varepsilon)\ell_i)] \leq \varepsilon \cdot \mathrm{e}^{-i}.$$

We set k to be large enough so that the above analysis would work for $i > k$. Thus except for probability at most $\delta + \varepsilon$, Player 1 stops in a step when Player 2 plays **R** with probability only ε. Thus the expected payoff to Player 1 is at most $\delta + 2\varepsilon$. □

6 No finite-memory ε-optimal deterministic-update Markov strategy exists

A *memory-based Markov strategy* is an extension of a memory-based strategy that may also depend on the round number. More precisely, for Player 1, the action map σ_1^a for a memory-based Markov strategy σ_1 is a map from $\mathbb{Z}_+ \times \mathcal{M}$ to $\Delta(A_1)$ and the update map σ_1^u for memory-based Markov strategies is a map from $\mathbb{Z}_+ \times A_1 \times A_2 \times \mathcal{M}$ to $\Delta(\mathcal{M})$. We say that Player 1 *follows* the memory-based Markov strategy σ_1 if in every round T when the game did not stop yet, he picks his next move a_1^T at random according to $\sigma_1^a(T, m_T)$, where the sequence $m_1, m_2, \ldots$ is given by letting $m_1 = m_s$ and for $T = 1, 2, 3, \ldots$ choosing m_{T+1} at random according to $\sigma_k^u(a_1^T, a_2^T, m_T, T)$, where a_2^T is the action chosen by Player 2 at round T. The definition of memory-based Markov strategies is similar for Player 2. Note that memory-based Markov strategies are more general than memory-based strategies.

A memory-based (resp. Markov) strategy σ_1 for Player 1 has *deterministic-update*, if for all $a_1 \in A_1$, all $a_2 \in A_2$ and all $m \in \mathcal{M}$ (resp. all $T \in Z_+$) the distribution $\sigma_1^u(a_1, a_2, m)$ (resp. $\sigma_1^u(T, a_1, a_2, m)$) is deterministic.

In this section we will argue that no finite-memory ε-optimal deterministic-update Markov strategy for Player 1 in the Big Match exists, for $\varepsilon < \frac{1}{2}$. Let n be some positive integer. Let σ_1 be some Markov strategy using at most n memory for Player 1. We will show that for all $\delta > 0$, there exists a strategy σ_2 for Player 2 that ensures that $u(\inf, (\sigma_1, \sigma_2)) < \delta$. This shows that σ_1 can only ensure value 0.

Construction of σ_2 and the sequence of strategies σ_2^k. We will now describe the construction of σ_2. The strategy σ_2 will be the final strategy in a finite sequence of strategies $(\sigma_2^k)_k$. Each of the strategies σ_2^k (and thus also σ_2) is a deterministic memory-based Markov strategy and will use the same set of memory states $\mathcal{M}$ of size n and update map σ_1^u as σ_1. Observe that the action map σ_2^a for such a strategy can be thought of as a (∞, n)-matrix A over $\{\mathbf{L}, \mathbf{R}\}$, where $A_{T,m} = \sigma_2^a(T, m)$.

- Let $\sigma_2^{k,a}$ be the action map for σ_2.
- Let $S_{\mathbf{R}}^k = \{(T, m) \mid \sigma_2^{k,a}(T, m) = \mathbf{R}\}$ (i.e. the pairs under which σ_2^k plays $\mathbf{R}$).
- For all T, let $\mathcal{M}^T = \{(T', m) \mid T' \leq T\}$ (i.e. the memory states before round T).
- For all T, let $S_{\mathbf{R}}^{k,T} = S_{\mathbf{R}}^k \cap \mathcal{M}^T$ (i.e. the pairs under which σ_2^k plays $\mathbf{R}$ before round T).

Properties of strategies in the sequence σ_2^k. Besides ensuring that the last strategy σ_2 in the sequence is such that $u(\inf, (\sigma_1, \sigma_2)) < \delta$, our construction of σ_2^k will ensure the following properties:

1. **Property 1.** The probability to stop (using union bound) while Player 2 plays $\mathbf{R}$ is at most

$$\sum_{(T,m) \in S_{\mathbf{R}}^k} \sigma_1^a(T, m)(\mathbf{R}) \leq (1 - 2^{-k})\delta \ .$$

2. **Property 2.** The infimum limit, for T going to infinite, of the fraction of all pairs before round T under which σ_2^k plays $\mathbf{R}$ is at least

$$\liminf_{t \to \infty} \frac{|S_{\mathbf{R}}^{k,T}|}{n \cdot T} \geq \frac{\delta \cdot k}{n} \ .$$

The sequence has finite length. Observe that Property 2 ensures that the strategy σ_2^k cannot exists, for $k > \frac{n}{\delta}$, implying that the sequence has finite length. This is because σ_2^k, for such k, otherwise would require that there is some T, such that the number of pairs such that σ_2^k plays $\mathbf{R}$ before round T is strictly more than the number of pairs before round T.

The strategy σ_2^0. The action map $\sigma_2^{0,a}$ is such that $\sigma_2^{0,a}(T, m) = \mathbf{L}$ for all $T \in \mathbb{N}$ and $m \in \mathcal{M}$. The strategy has the wanted properties (because $S_{\mathbf{R}}^0$ is the empty set).

Informal inductive construction of the sequence of strategies $(\sigma_2^k)_k$. Consider playing σ_1 against σ_2^k and let the play be P. If the outcome is above δ (that is, σ_2^k does not satisfy the properties of σ_2 and we therefore need to construct σ_2^{k+1}) we know two things: 1) The probability that P stops is not 1, since otherwise the outcome is at most $(1-2^{-k})\delta$, by Property 1. 2) Player 2 plays $\mathbf{L}$ above a δ fraction of the time (in lim-inf). In that case, since the probability to stop is below 1, there is a round M, such that if the play reaches that round, the probability to stop in the rest of the play is below $2^{-(k+1)}\delta$. Let the sequence of memory states the players uses in the play be $(m_T^k)_T$. The strategy σ_2^{k+1} is then like the strategy σ_2^k, except for playing $\mathbf{R}$ then in memory m_T^k in round T. By choice of M, we get Property 1. Also, since the fraction of the time player 2 played $\mathbf{L}$ in P was above δ, then σ_2^{k+1} contains $\frac{\delta}{n}$ more $\mathbf{R}$'s than σ_2^k, which is Property 2. In Figure 3, we have an illustration of the action map of a possible sequence of strategies $(\sigma_2^k)_k$ for $n=4$ and some strategy σ_1. The line illustrates the memory sequence $(m_T^k)_T$ (for instance, the sequence $(m_T^1)_T$ is 41321144423333313131...) where the probability to stop somewhere on the solid line is below $2^{-(k+1)}\delta$.

Formal inductive construction of σ_2^{k+1}. We will now show that, given σ_2^k, we either have that $u(\inf,(\sigma_1,\sigma_2^k))<\delta$ or we can construct σ_2^{k+1} with the wanted properties. In the first case σ_2^k has the properties we wanted from σ_2 and can then stop the sequence. We now consider the case where $u(\inf,(\sigma_1,\sigma_2^k))\geq\delta$.

- Let p be the probability that the play stops, when the players follows (σ_1,σ_2^k).
- For all T, let p_T be the probability that the play stops in round T, conditioned on it not stopping before round T.
- Let $(m_T^k)_T$ be the sequence of memory states associated with σ_1 (or equally σ_2^k), then played against σ_2^k. Note that given σ_1 and σ_2^k, the sequence is deterministic, because the choices of Player 2 are deterministic and if the game has not stopped, Player 1 has played $\mathbf{L}$ at all earlier times and he updates its memory state deterministically.
- For all T, let $S_{\text{tail}}^{k,T}=\{(T',m_{T'}^k)\mid T'\geq T\}$ (i.e. the pairs in the memory state sequence after round T).
- Let v be the value of the game if it does not stop, i.e.

  $$v=\liminf_{T\to\infty}\frac{|(S_{\text{tail}}^{k,0}\setminus S_{\mathbf{R}}^k)\cap\mathcal{M}^T|}{T} \quad .$$

  Note that the set $(S_{\text{tail}}^{k,0}\setminus S_{\mathbf{R}}^k)\cap\mathcal{M}^T$ is the set of pairs, in the memory state sequence in which Player 2 plays $\mathbf{L}$ before round T.

Using the above definitions, we see that

$$u(\inf,(\sigma_1,\sigma_2^k))\leq p(1-2^{-k})\delta+(1-p)v \quad .$$

Thus, since $u(\inf,(\sigma_1,\sigma_2^k))\geq\delta$ we must have that $p<1$ and $\delta\leq v$.

We will now use the following mathematical lemma.

Lemma 21. *Let $x<1$ be some real number. Let $(x_T)_T$ be an infinite sequence, in which $x_T\in[0,1]$ and $1-\prod_T(1-x_T)=x$, then $\sum_T x_T\leq-\log(1-x)<\infty$.*

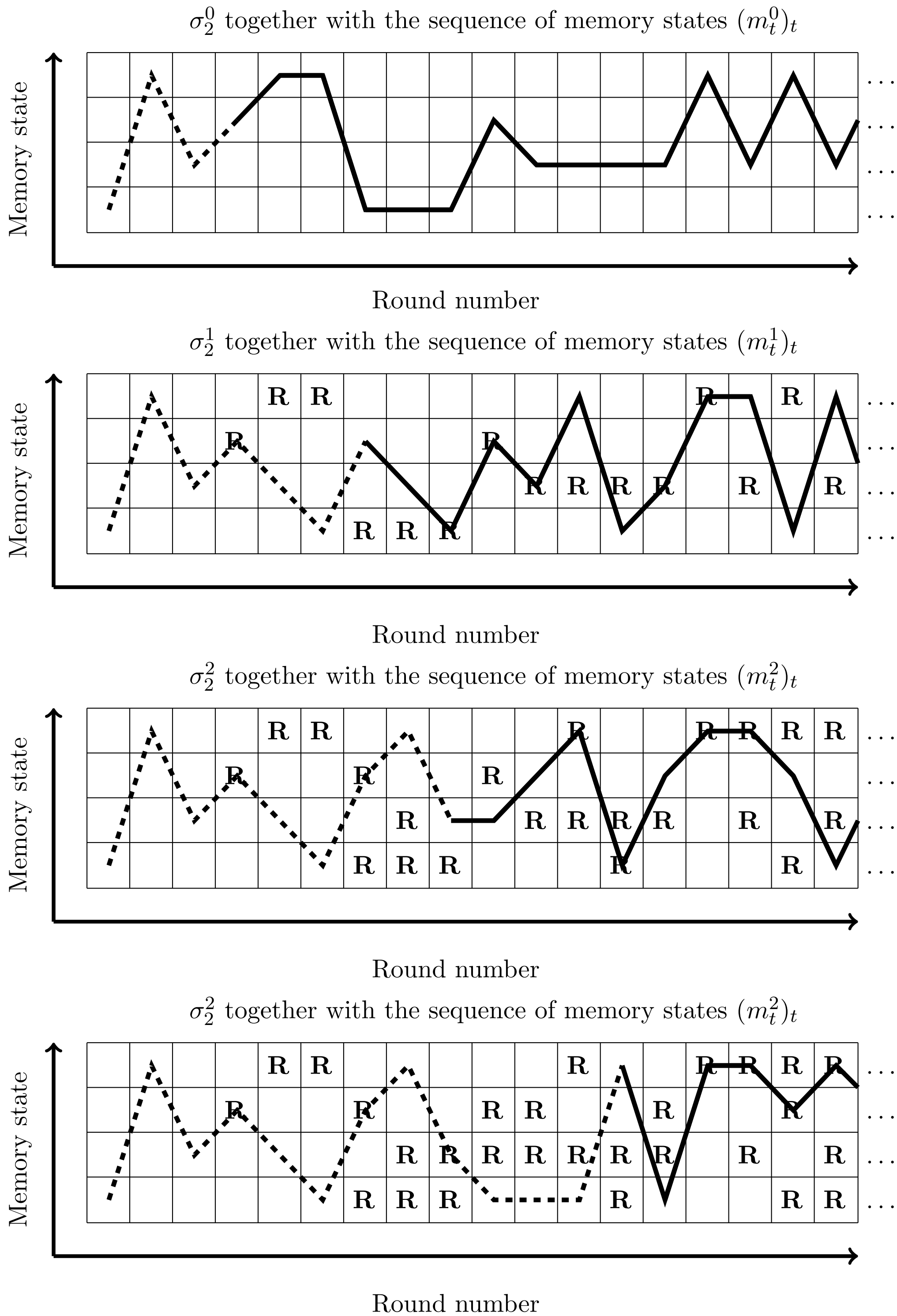

Figure 3: Posible sequence of strategies $(\sigma_2^k)_k$ and corresponding sequence of memory states $(m_t^k)_t$.

Proof. We have that all $x_T < 1$, since otherwise the product would be 0, and hence $x = 1$. Also, we get that

$$\prod_T (1 - x_T) = 1 - x \Rightarrow \sum_T \log(1 - x_T) = \log(1 - x)$$
$$\Rightarrow - \sum_T \log(1 - x_T) = - \log(1 - x) \Rightarrow \sum_T x_T \leq - \log(1 - x) \ .$$

The last inequality comes from that $f(y) = -\log(1-y) \geq y$ for $y \in [0, 1)$. This follows from $f(0) = 0$ and $f'(y) = \frac{1}{1-y} \geq 1$, for $y \in [0, 1)$. □

We have that $1 - \prod_T (1 - p_T) = p$ and by Lemma 21 we then get that $\sum_T p_T$ is finite. Hence, there exists some M such that $\sum_{T=M}^{\infty} p_T \leq \frac{\delta}{2^{k+1}}$.

We can now define $\sigma_2^{k+1,a}$. Let

$$\sigma_2^{k+1,a}(T, m) = \begin{cases} \mathbf{R} & \text{if } (T, m) \in S_{\text{tail}}^{k,M} \\ \sigma_2^{k,a}(T, m) & \text{otherwise} \end{cases}$$

The strategy σ_2^{k+1} satisfies the wanted properties. We will now show that σ_2^{k+1} satisfies the wanted properties.

1. That Property 1 is satisfied comes from that

$$\sum_{(T,m) \in S_{\mathbf{R}}^{k+1}} \sigma_1^a(T, m)(\mathbf{R}) \leq \sum_{(T,m) \in S_{\mathbf{R}}^{k}} \sigma_1^a(T, m)(\mathbf{R}) + \sum_{(T,m) \in S_{\text{tail}}^{k,M}} \sigma_1^a(T, m)(\mathbf{R})$$
$$\leq (1 - 2^{-k})\delta + 2^{-k-1}\delta = (1 - 2^{-k-1})\delta \ .$$

2. That Property 2 is satisfied can be seen as follows. We have that

$$\liminf_{T \to \infty} \frac{|S_{\mathbf{R}}^{k+1,T}|}{n \cdot T} = \liminf_{T \to \infty} \frac{|S_{\mathbf{R}}^{k,T}| + |(S_{\text{tail}}^{k,M} \setminus S_{\mathbf{R}}^{k}) \cap \mathcal{M}^T|}{n \cdot T}$$
$$\geq \liminf_{T \to \infty} \frac{|S_{\mathbf{R}}^{k,T}|}{n \cdot T} + \liminf_{T \to \infty} \frac{|(S_{\text{tail}}^{k,M} \setminus S_{\mathbf{R}}^{k}) \cap \mathcal{M}^T|}{n \cdot T} \ .$$

Since the properties are satisfied for σ_2^k, we get that $\liminf_{T\to\infty} \frac{|S_{\mathbf{R}}^{k,T}|}{n \cdot T} \geq \frac{\delta k}{n}$. We thus just need to argue that

$$\liminf_{T \to \infty} \frac{|(S_{\text{tail}}^{k,M} \setminus S_{\mathbf{R}}^{k}) \cap \mathcal{M}^T|}{n' \cdot T} \geq \frac{\delta}{n}$$

and we are done. That statement can be seen as follows

$$\delta \leq v \Rightarrow \delta \leq \liminf_{T \to \infty} \frac{|(S_{\text{tail}}^{k,0} \setminus S_{\mathbf{R}}^{k}) \cap \mathcal{M}^T|}{T} \Rightarrow \frac{\delta}{n} \leq \liminf_{T \to \infty} \frac{|(S_{\text{tail}}^{k,0} \setminus S_{\mathbf{R}}^{k}) \cap \mathcal{M}^T|}{n' \cdot T} \Rightarrow$$
$$\frac{\delta}{n} \leq \liminf_{T \to \infty} \frac{|(S_{\text{tail}}^{k,M} \setminus S_{\mathbf{R}}^{k}) \cap \mathcal{M}^T|}{n \cdot T} \ ,$$

where the last inequality comes from that $S_{\text{tail}}^{k,0}$ consists of the same pairs as $S_{\text{tail}}^{k,M}$ and then M more.

The above leads to the following theorem.

Theorem 22. *For all $\varepsilon < \frac{1}{2}$, there exists no finite-memory deterministic-update ε-optimal Markov strategy for Player 1 in the Big Match.*

7 Generalized Big Match Games

In order to generalize our results to arbitrary repeated games with absorbing states, we follow the approach of Kohlberg and consider first the subset of such games where Player 1, like in the Big Match, has just the choice whether to declare that the game should stop or continue. But unlike the Big Match, in case Player 1 declares the game should stop, the game will only stop with some non-zero probability (that depends on the action of Player 2). Furthermore, Player 2 can have any number of actions and the rewards can be arbitrary. We call such games *generalized* Big Match games.

More formally, a generalized Big Match game G is specified as follows. Let $A_1 = \{\mathbf{L}, \mathbf{R}\}$ and let A_2 be any finite set. The rewards $\pi(a_1, a_2)$ are arbitrary, but the stop probabilities $\omega(a_1, a_2)$ must satisfy that $\omega(\mathbf{L}, a_2) = 0$ and $\omega(\mathbf{R}, a_2) > 0$ for all $a_2 \in A_2$.

Our strategies for generalized Big Match games will follow the same template as those given in Sections 3 and 4. The change required is a modification of the base strategies $\sigma_1^{i,\xi}$. The proof that these new base strategies $\tau_1^{i,\xi}$ have the desired properties follow those for $\sigma_1^{i,\xi}$, but uses also additional ideas similar to those of Kohlberg[11].

Given G we define the *derived matrix game* $\widetilde{G}$ by

$$\widetilde{G}_{a_1,a_2} = \begin{cases} \pi(a_1, a_2) & \text{if } a_1 = \mathbf{L} \\ \omega(a_1, a_2) \cdot \pi(a_1, a_2) & \text{if } a_1 = \mathbf{R} \end{cases} .$$

Assumption 1. In the remainder of this section we make the following assumptions about the given generalized Big Match game G:

- The entries of $\widetilde{G}$ are integer.
- The value $\widetilde{G}$ is 0.
- Player 1 does not have a pure optimal strategy in $\widetilde{G}$.

We observe that the last requirement means that in the matrix game $\widetilde{G}$ Player 1 has a unique optimal strategy and it plays each action with non-zero probability.

Define $\omega = \min_{a_2} \omega(\mathbf{R}, a_2)$ to be the minimum non-zero stop probability of G and $K = \max_{a_1,a_2} |\widetilde{G}_{a_1,a_2}|$ be the maximum magnitude of an entry of $\widetilde{G}$.

Remark. As noted in Observation 1, it is sufficient to consider only pure strategies for Player 2. Unlike for the Big Match, here play may continue even when Player 1 chooses the action $\mathbf{R}$, and hence it is not sufficient to consider only pure Markov strategies for Player 2. We shall however (as done also by Kohlberg) give only the proof for this special case and just note that the proof for general pure strategies of Player 2 is done along the same lines

Generalized density of pure Markov strategies In order to generalize our strategies given in Sections 3 and 4, we must first generalize the notion of density of a pure Markov strategy σ for Player 2. We then define

$$\mathrm{gdens}(\sigma^T) = \frac{\sum_{i=1}^{T} \pi(\mathbf{L}, \sigma_i)}{T} \quad ,$$

and further, for $T' < T$ we define

$$\mathrm{gdens}(\sigma, T', T) = \frac{\sum_{i=T'}^{T} \pi(\mathbf{L}, \sigma_i)}{T - T'} \quad .$$

We can make a similar correspondence between the generalized density of a play and the average of the rewards received by Player 1. The main difference here is that the play may possibly continue whenever Player 1 plays the action $\mathbf{R}$. However, the event that Player 1 plays action $\mathbf{R}$ an infinite number of times happens with probability 0, since each time Player 1 does play action $\mathbf{R}$, the game stops with probability at least $\omega > 0$ by definition.

Observation 3. *Suppose Player 2 follows a pure Markov strategy σ. Consider a play P in which Player 1 plays only the action $\boldsymbol{L}$. Then for any $T < |P|$ we have*

$$\mathrm{gdens}(\sigma^T) = \frac{1}{T} \sum_{T'=1}^{T} r_T \quad ,$$

where r_T is the reward given to Player 1 in round T. Consider now an infinite play P in which Player 1 plays action $\boldsymbol{R}$ only a finite number of times. Then we have

$$u_{\inf}(P) = \liminf_{T \to \infty} \mathrm{gdens}(\sigma^T) \quad ,$$

and

$$u_{\sup}(P) = \limsup_{T \to \infty} \mathrm{gdens}(\sigma^T) \quad .$$

We need the following simple statement, which can be viewed as a quantified version of [11, Lemma 2.5].

Lemma 23. *Let σ be a pure Markov strategy for Player 2, and let j be any integer. If $T \cdot \mathrm{gdens}(\sigma^T) \leq -j \cdot 2K$ then*

$$\sum_{i=1}^{T} \omega(\boldsymbol{R}, \sigma_i^T) \cdot \pi(\boldsymbol{R}, \sigma_i^T) \geq j \quad .$$

Proof. Let σ_1 be the (unique) optimal strategy in $\widetilde{G}$. By Assumption 1 we have that $\sigma_1(\mathbf{L}) > 0$ and that the value of $\widetilde{G}$ is 0. This immediately implies that $\sigma_1(\mathbf{L}) > 1/2K$, since $p = \sigma_1(\mathbf{L})$ must satisfy an equation of the form

$$p \cdot \widetilde{G}_{\mathbf{L},a_2} + (1-p) \cdot \widetilde{G}_{\mathbf{R},a_2} = 0 \quad ,$$

for some $a_2 \in A_2$, where the entries $\widetilde{G}_{\mathbf{L},a_2}$ and $\widetilde{G}_{\mathbf{R},a_2}$ are non-zero and integers of magnitude at most K.

	L	**R**
L	1	-1
R	-1^*	1^*

Figure 4: The Big Match as a generalized Big Match game.

Now, by optimality of σ_1 we have

$$\sigma_1(\mathbf{L}) \cdot \pi(\mathbf{L}, a_2) + \sigma_1(\mathbf{R}) \cdot \omega(\mathbf{R}, a_2) \cdot \pi(\mathbf{R}, a_2) \geq 0 \ ,$$

for any $a_2 \in A_2$, and thus also

$$\sum_{i=1}^{T} \sigma_1(\mathbf{L}) \cdot \pi(\mathbf{L}, \sigma_i^T) + \sigma_1(\mathbf{R}) \cdot \omega(\mathbf{R}, \sigma_i^T) \cdot \pi(\mathbf{R}, \sigma_i^T)) \geq 0 \ .$$

By assumption $T \cdot \mathrm{gdens}(\sigma^T) = \sum_{i=1}^{T} \pi(\mathbf{L}, \sigma_i^T) \leq -j \cdot 2K$. Hence

$$\sum_{i=1}^{T} \omega(\mathbf{R}, \sigma_i^T) \cdot \pi(\mathbf{R}, \sigma_i^T) \geq -\frac{\sigma_1(\mathbf{L})}{\sigma_1(\mathbf{R})} \sum_{i=1}^{T} \pi(\mathbf{L}, \sigma_i^T) \geq \frac{1}{2K} \cdot j \cdot 2K = j \ .$$

□

The Big Match as a generalized Big Match game The Big Match as defined in Section 2 does not immediately fit our definition of generalized Big Match games, since the value of the derived matrix game is 1/2 rather than 0. To achieve this we may simply replace the rewards of value 0 with value -1. Doing this we see that

$$\begin{aligned}\mathrm{gdens}(\sigma^T) &= \frac{|\{i \mid (\sigma^T)_i = \mathbf{L}\}| - |\{i \mid (\sigma^T)_i = \mathbf{R}\}|}{T} \\ &= \frac{2|\{i \mid (\sigma^T)_i = \mathbf{L}\}| - T}{T} = 2\,\mathrm{dens}(\sigma^T) - 1\end{aligned}$$

for any σ^T.

7.1 Small space ε-supremum-optimal strategies in generalized Big Match games

For given ε, let $\xi = \varepsilon^2/(4K^4)$. Thus $\varepsilon = 2K^2\sqrt{\xi}$. The strategy σ_1^* for Player 1 is obtained from the strategy of Section 3, by exchanging the base strategy $\sigma_1^{i,\xi}$ with the new base strategy $\tau_1^{i,\xi}$ defined next.

The base strategy. Similar to $\sigma_1^{i,\xi}$, the strategy $\tau_1^{i,\xi}$ uses deterministic updates of memory, and uses integers as memory states The memory update function is changed to

$$\tau_1^{i,\xi,u}(a, j) = j - \widetilde{G}_{\mathbf{R},a} = j - \omega(\mathbf{R}, a)\pi(\mathbf{R}, a) \ ,$$

whereas the action function is unchanged as

$$\tau_1^{i,\xi,a}(j)(\mathbf{R}) = \begin{cases} \xi^4(1-\xi)^{i+j} & \text{if } i+j>0 \\ \xi^4 & \text{if } i+j \le 0 \end{cases} .$$

Note that when the Big Match is redefined as a generalized Big Match game, we have that $\tau_1^{i,\xi} = \sigma_1^{i,\xi}$.

We will next generalize the statements of Section 3 in the following paragraphs.

7.1.1 Space usage of the strategy

We will here consider the space usage of σ_1^*. Like for the strategy σ_1^* for the Big Match we are interested in the length of the epochs and can generalize Lemma 7 to generalized Big Match.

Lemma 24. *For any $\gamma, \delta \in (0, 1/4)$, there is a constant M such that with probability at least $1-\gamma$, for all $i \ge M$ and all $j \in \{1, \ldots, i\}$, we have that*

$$t(i,j,i^2) - t(i,j-1,i^2) \in [(1-\delta)F(i)/i, (1+\delta)F(i)/i] \;.$$

Proof. The proof is precisely the same as for Lemma 7, since it only concerns itself with the length of epochs and not with the base strategy, and only the base strategy has changed as compared to section Section 3. □

We next give a generalization of Lemma 8.

Lemma 25. *For all constants $\gamma > 0$, with probability at least $1-\gamma$, the space usage of σ_1^* is $O(\log f(T) + \log K)$.*

Proof. Observe that the base strategy $\tau_1^{i,\xi}$ can reach a factor of $2K$ more memory states upto round T than $\sigma_1^{i,\xi}$, since $\sigma_1^{i,\xi}$ changes its counter by ± 1 in each round while $\tau_1^{i,\xi}$ changes its counter by some number in $\{-K, \ldots, K\}$. Thus, σ_1^* uses at most a factor $2K$ more memory states then using $\tau_1^{i,\xi}$ as the base strategy instead of $\sigma_1^{i,\xi}$ in round T for any T. The statement then follows from a similar proof as the one for Lemma 8. □

7.1.2 Play stopping implies good outcome

Lemma 26. *Let $T, i \ge 1$ be integers and $0 < \xi < 1$ be a real number. Let $\sigma \in (A_2)^T$ be an arbitrary prefix of a pure Markov strategy for Player 2. Consider the first T rounds where the players play G following $\tau_1^{i,\varepsilon}$ and σ respectively.*

Let $S \in \{1, \ldots, T\} \cup \{\infty\}$ be a random variable in case the game stops in the first T rounds denotes that round, and is ∞ otherwise. Let U be the random variable that denotes the outcome in case the game stops in the first T rounds, and is 0 otherwise. Then

1. $$-\mathrm{E}[U \mid U<0] \Pr[U<0] \le \xi^3(1-\xi)^{i-K+1} + (1-\xi)^{1-2K}\mathrm{E}[U \mid U>0]\Pr[U>0] \;.$$

2. *If* $\mathrm{gdens}(\sigma^T) \le -i \cdot 2K/T$ *then*
$$\Pr[S<\infty] \ge \xi^4 \cdot \omega \;.$$

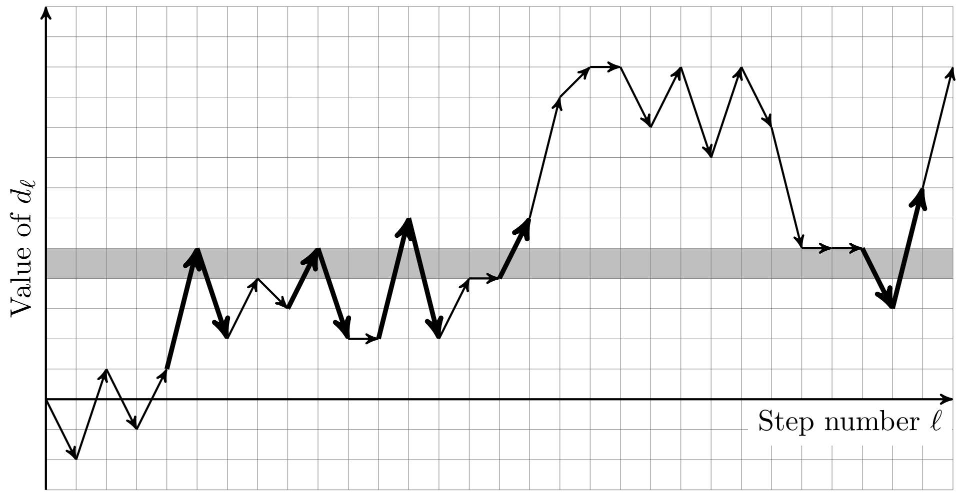

Figure 5: Possible movement of the counter $\tau_1^{i,\varepsilon}$ uses as memory through the rounds.

Proof. Define

$$d_\ell = -\sum_{\ell' < \ell} \widetilde{G}_{\mathbf{R},\sigma_{\ell'}} \quad .$$

and note that d_ℓ is the value of the counter used by $\tau_1^{i,\xi}$ as memory in step ℓ. There is an illustration of how d_ℓ could evolve through the steps in Figure 5. Let $I = \{\ell \in \{1,\dots,T\} \mid \widetilde{G}_{\mathbf{R},\sigma_\ell} < 0\}$ and $D = \{\ell \in \{1,\dots,T\} \mid \widetilde{G}_{\mathbf{R},\sigma_\ell} > 0\}$ be the sets of times where the counter is incremented and decremented, respectively. For integer d, define sets K_d by

$$\begin{aligned} K_d = \{\ell \in \{1,\dots,T\} \,|(&\widetilde{G}_{\mathbf{R},\sigma_\ell} < 0 \ \&\ d_\ell \le d < d_\ell - \widetilde{G}_{\mathbf{R},\sigma_\ell}) \text{ or} \\ &(\widetilde{G}_{\mathbf{R},\sigma_\ell} > 0 \ \&\ d_\ell - \widetilde{G}_{\mathbf{R},\sigma_\ell} \le d < d_\ell)\} \quad . \end{aligned}$$

Intuitively, K_d is now the set of times where the counter is either incremented to pass trough the value d and end above, or is decreased from above d passing through the value d. For instance, the gray row of Figure 5 corresponds to a set K_d, which consists of the edges that starts below or on the bottom of the gray row and ends over or on the top of the gray row. Each such edge are wider, in the figure, then the remaining edges. Note that horizontal edges are not in K_d. Notice each $\ell \in \{1,\dots,T\}$ belongs to exactly $|\widetilde{G}_{\mathbf{R},\sigma_\ell}|$ many of the sets K_d. Also if $K_d = \{k_1 < k_2 < \cdots < k_m\}$, for some m, then $k_j \in I$ if and only if $k_{j+1} \in k_{j+1} \in D$. Thus K_d is an sequence of elements alternately from I and D, and starts with an element of I when $d \ge 0$ and starts with an element of D otherwise.

Finally let, $A_\ell \in \{\mathbf{L},\mathbf{R}\}$ be the random variable indicating the action of Player 1 at time step

ℓ, and define $p_\ell = \Pr[S \geq \ell \;\&\; A_\ell = \mathbf{R}]$. We now have

$$\begin{aligned}
\mathrm{E}[U] &= \sum_{\ell=1}^{T} \Pr[S=\ell] \cdot \pi(\mathbf{R}, \sigma_\ell) = \sum_{\ell=1}^{T} p_\ell \cdot \omega(\mathbf{R}, \sigma_\ell)\pi(\mathbf{R}, \sigma_\ell) \\
&= \sum_{\ell=1}^{T} p_\ell \cdot \widetilde{G}_{\mathbf{R},\sigma_\ell} = \sum_{\ell=1}^{T} \sum_{d:\ell \in K_d} \operatorname{sgn}(\widetilde{G}_{\mathbf{R},\sigma_\ell}) \cdot p_\ell \\
&= \sum_{d} \sum_{\ell \in K_d} \operatorname{sgn}(\widetilde{G}_{\mathbf{R},\sigma_\ell}) \cdot p_\ell \ .
\end{aligned}$$

Similarly,

$$\mathrm{E}[U \mid U < 0] \Pr[U < 0] = -\sum_{d} \sum_{\ell \in K_d \cap I} p_\ell \ ,$$

and

$$\mathrm{E}[U \mid U > 0] \Pr[U > 0] = \sum_{d} \sum_{\ell \in K_d \cap D} p_\ell \ .$$

For integer d, define $E_{\mathrm{loss},d} = \sum_{\ell \in K_d \cap I} p_\ell$ and $E_{\mathrm{win},d} = \sum_{\ell \in K_d \cap D} p_\ell$. Recall that by definition of $\tau^{i,\xi}$ we have $\Pr[A_\ell = \mathbf{R} \mid S \geq \ell] = \xi^4(1-\xi)^{\max(0, i+d_\ell)}$.

Consider now $d \geq 0$. Then $p_{k_1} \leq \Pr[A_\ell = \mathbf{R} \mid S \geq k_1] \leq \xi^4(1-\xi)^{i+d-K+1}$, since $d_{k_1} > d - K$. For even j we have

$$\begin{aligned}
p_{k_j} &= \Pr[A_{k_j} = \mathbf{R} | S \geq k_j] \Pr[S \geq k_j] \\
&\geq (1-\xi)^{2K-1} \Pr[A_{k_j} = \mathbf{R} | S \geq k_{j+1}] \Pr[S \geq k_{j+1}] = (1-\xi)^{2K-1} p_{k_{j+1}} \ ,
\end{aligned}$$

since $d_{k_j} \leq d + K < d_{k_{j+1}} + 2K$. It follows that

$$E_{\mathrm{loss},d} \leq \xi^4(1-\xi)^{i+d-K+1} + (1-\xi)^{1-2K} E_{\mathrm{win},d} \ .$$

For $d < 0$, $E_{\mathrm{loss},d} \leq (1-\xi)^{1-2K} E_{\mathrm{win},d}$, since for *odd* j, $p_{k_j} \geq (1-\xi)^{2K-1} p_{k_{j+1}}$ as above. For $d < -i$ we can give better estimates, like in the case of Lemma 9, but it is not needed.

Taking the summation over d then gives

$$\begin{aligned}
-\mathrm{E}[U \mid U < 0] \Pr[U < 0] &= \sum_{d} E_{\mathrm{loss},d} \\
&\leq \sum_{d \geq 0} \xi^4 (1-\xi)^{i+d-K+1} + (1-\xi)^{1-2K} \sum_{d} E_{\mathrm{win},d} \\
&= \xi^3 (1-\xi)^{i-K+1} + (1-\xi)^{1-2K} \mathrm{E}[U \mid U > 0] \Pr[U > 0] \ .
\end{aligned}$$

For the second part, if $\mathrm{gdens}(\sigma^T) \leq -i \cdot 2K/T$ then Lemma 23 gives $\sum_{\ell=1}^{T} \widetilde{G}_{\mathbf{R},\sigma_\ell} \geq i$, and hence $d_T \leq -i$. This implies that if the game reaches round T, Player 1 plays $\mathbf{R}$ with probability ξ^4, which means the game stops with probability at least $\xi^4 \cdot \omega$. □

Lemma 27. *Let $0 < \xi < 0$. Let σ be a pure Markov strategy for Player 2. Let S be the event that the game stops. Let U be the random variable that denotes the outcome in case the game stops, and is 0 otherwise. If $\Pr[S] \geq K\sqrt{\xi}$ then $\mathrm{E}[U \mid S] \geq -2K\sqrt{\xi}$.*

Proof. The proof follows along that of Lemma 27. Let $A_{i,j}$ be the set of plays in which between round 1 and round $t(i,j,1)-1$, the game does not stop. Let $U_{i,j}$ be the random variable that denotes the outcome in case the game stops between round $t(i,j,1)$ and round $t(i,j,i^2)$, and is 0 otherwise. Note that $\mathrm{E}[U] = \sum_{i,j} \mathrm{E}[U_{i,j} \mid A_{i,j}] \Pr[A_{i,j}]$.

Fix a possible value of all $t(i,j,k)$'s and denote by Y the event that these particular values actually occur. Fix i and j. Conditioned on Y and $A_{i,j}$, between time $t(i,j,1)$ and $t(i,j,i^2)$ Player 1 plays $\tau_1^{i,\xi}$ against a fixed strategy $\sigma_{t(i,j,1)}, \sigma_{t(i,j,2)}, \ldots, \sigma_{t(i,j,i^2)}$ for Player 2. By Lemma 26 we thus have

$$\begin{aligned}
&- \mathrm{E}[U_{i,j} \mid U_{i,j} < 0, Y, A_{i,j}] \Pr[U_{i,j} < 0 \mid Y, A_{i,j}] \\
&\qquad \le \xi^3 (1-\xi)^{i-K+1} + (1-\xi)^{1-2K} \mathrm{E}[U_{i,j} \mid U_{i,j} > 0, Y, A_{i,j}] \Pr[U_{i,j} > 0 \mid Y, A_{i,j}] \ .
\end{aligned}$$

Since the above inequality is true conditioned on arbitrary values of $t(i,j_1,j_2)$'s, it is true also without the conditioning:

$$\begin{aligned}
&- \mathrm{E}[U_{i,j} \mid U_{i,j} < 0, A_{i,j}] \Pr[U_{i,j} < 0 \mid A_{i,j}] \\
&\qquad \le \xi^3 (1-\xi)^{i-K+1} + (1-\xi)^{1-2K} \mathrm{E}[U_{i,j} \mid U_{i,j} > 0, A_{i,j}] \Pr[U_{i,j} > 0 \mid A_{i,j}] \ .
\end{aligned}$$

Thus,

$$\begin{aligned}
&- \mathrm{E}[U \mid U < 0] \Pr[U < 0] \\
&= - \sum_{i=1}^{\infty} \sum_{j=1}^{i} \mathrm{E}[U_{i,j} \mid U_{i,j} < 0, A_{i,j}] \Pr[U_{i,j} < 0 \mid A_{i,j}] \Pr[A_{i,j}] \\
&\le \sum_{i=1}^{\infty} \sum_{j=1}^{i} \Big(\xi^3 (1-\xi)^{i-K+1} + \\
&\qquad (1-\xi)^{1-2K} \mathrm{E}[U_{i,j} \mid U_{i,j} > 0, A_{i,j}] \Pr[U_{i,j} > 0 \mid A_{i,j}] \Big) \Pr[A_{i,j}] \\
&\le \xi + (1-\xi)^{1-2K} \sum_{i=1}^{\infty} \sum_{j=1}^{i} \mathrm{E}[U_{i,j} \mid U_{i,j} > 0, A_{i,j}] \Pr[U_{i,j} > 0 \mid A_{i,j}] \Pr[A_{i,j}] \\
&= \xi + (1-\xi)^{1-2K} \mathrm{E}[U \mid U > 0] \Pr[U > 0]
\end{aligned}$$

Hence

$$-(1-\xi)^{2K-1} \mathrm{E}[U \mid U < 0] \Pr[U < 0] \le \xi (1-\xi)^{2K-1} + \mathrm{E}[U \mid U > 0] \Pr[U > 0] \ ,$$

and so

$$\begin{aligned}
-\mathrm{E}[U] &= -\mathrm{E}[U \mid U < 0] \Pr[U < 0] - \mathrm{E}[U \mid U > 0] \Pr[U > 0] \\
&\le -(1 - (1-\xi)^{2K-1}) \mathrm{E}[U \mid U < 0] \Pr[U < 0] + \xi (1-\xi)^{2K-1} \\
&\le K(1 - (1-\xi)^{2K-1}) + \xi (1-\xi)^{2K-1} \\
&\le K(2K-1)\xi + \xi \le \ 2K^2 \cdot \xi,
\end{aligned}$$

where we use the equation $1 - (1-x)^k \le kx$, which is valid for positive integer k and $0 \le x \le 1$. Hence the expected outcome conditioned on the game stopping can be estimated by

$$\mathrm{E}[U \mid S] = \frac{\mathrm{E}[U]}{\Pr[S]} \ge \frac{-2K^2 \xi}{K\sqrt{\xi}} = -2K\sqrt{\xi} \ ,$$

using the above estimate on $\mathrm{E}[U]$ and the assumption $\Pr[S] \geq K\sqrt{\xi}$. $\square$

7.1.3 Low density means play stops

Lemma 28. *Let $\varepsilon, \delta \in (0,1)$. Let σ be an arbitrary pure Markov strategy for Player 2. Let $a_{i,j}$ be some numbers. Consider the event Y where $t(i,j) = a_{i,j}$ for all i,j. If $\limsup_{T\to\infty} \mathrm{gdens}(\sigma^T) \leq -\delta$, then conditioned on Y, there is an infinite sequence of sub-epochs and epochs $(i_n, j_n)_n$ such that* $\mathrm{gdens}(\sigma, a_{i_n,j_n}, a_{i_n,j_n+1}) \leq -\delta/4$.

Proof. The proof follows that of Lemma 11 with small changes. Thus, let M be such that for every $T' \geq M$ we have $\mathrm{gdens}(\sigma^{T'}) \leq -\delta/2$. Let $(T_n)_n$ be a sequence such that $T_1 \geq M$ and for all $n \geq 1$ we have that $T_{n+1} \cdot \delta/4 \geq K \cdot T_n$ and $T_n = a_{i,j}$ for some i,j. Let $(i'_n, j'_n)_n$ be the sequence such that $T_n = a_{i'_n,j'_n}$. This means that even if $\mathrm{gdens}(\sigma^{T_n}) = -K$, the generalized density $\mathrm{gdens}(\sigma, T_n+1, T_{n+1})$ is at most $-\delta/4$, because $\mathrm{gdens}(\sigma^{T_{n+1}}) \leq -\delta/2$. But, we then get that there exists some sub-epoch j_n in epoch i_n, such that $j'_n \leq j_n \leq j'_{n+1}$ and such that $i'_n \leq i_n \leq i'_{n+1}$ for which the generalized density of that sub-epoch $\mathrm{dens}(\sigma, a_{i_n,j_n}+1, a_{i_n,(j_n+1)})$ is at most $1/2 - \delta/4$, because not all sub-epochs can have generalized density below that of the average sub-epoch. But then $(i_n, j_n)_n$ satisfies the lemma statement. $\square$

Lemma 29. *Let σ be an arbitrary pure Markov strategy for Player 2. If*

$$\limsup_{T\to\infty} \mathrm{gdens}(\sigma^T) < 0$$

then when played against σ_1^* the play stops with probability 1.

Proof. The proof follows that of Lemma 12 with small changes. Let $\delta > 0$ be such that $\limsup_{T\to\infty} \mathrm{gdens}(\sigma^T) \leq -\delta$. Consider arbitrary numbers $a_{i,j}$ and the event Y stating that $t(i,j) = a_{i,j}$ for all i,j. Let $(i_n, j_n)_n$ be the sequence of sub-epochs and epochs shown to exists by Lemma 28 with probability 1. That is, for each (i_n, j_n) we have that sub-epoch j_n of epoch i_n has generalized density at most $-\delta/4$. We see that, conditioned on Y that each sample are sampled uniformly at random in each sub-epoch j of each epoch i, except for the last sample.

Now consider some fixed n. By Hoeffding's inequality, Theorem 46 (setting $a_i = -K$ and $b_i = K$, and letting $c_i = \pi(\mathbf{L}, \sigma_{t(i_n,j_n,i)})$), the probability that the generalized density of the subsequence given by the first $(i_n)^2 - 1$ samples is more that $-\delta/8$ is bounded by

$$2\exp\left(-\frac{2(\frac{\delta}{8}((i_n)^2-1))^2}{((i_n)^2-1)(2K)^2}\right) = 2\exp\left(-\frac{\delta^2}{128\cdot K^2}((i_n)^2-1)\right) \quad .$$

For sufficiently large i_n, this is less than $\frac{1}{2}$. If on the other hand the generalized density of the subsequence is more than $-\delta/8$, we see that $(i_n)^2 - 1 \geq i_n \cdot 2K \cdot 8/\delta$ for large enough i_n, and in that case we have, by Lemma 26, that the game stops with probability at least $\xi^4 \cdot \omega$ in sub-epoch j_n of epoch i_n. Thus, for each of the infinitely many n's for which i_n is sufficiently high, we have a probability of at least $\frac{\xi^4\cdot\omega}{2}$ of stopping. Thus play must stop with probability 1.

The argument was conditioned on some fixed assignment of endpoints of sup-epochs and epochs, but since there is such a assignment with probability 1 (since they are finite with probability 1), we conclude that the proof works without the condition. $\square$

7.1.4 Proof of main result

Theorem 30. *The strategy σ_1^* is $2K^2\sqrt{\xi}$-supremum-optimal, and for all $\delta > 0$, with probability at least $1-\delta$ does it use space $O(\log f(T) + \log K)$.*

Proof. The space usage follows from Lemma 25. Let σ be a pure Markov strategy for Player 2. Let S be the event that the game stops. Let U be the random variable that denotes the outcome in case the game stops, and is equal to $\limsup_{T\to\infty} \text{gdens}(\sigma^T)$ otherwise. By Observation 3 we have $u_{\text{sup}}(\sigma_1^*, \sigma) = \text{E}[U]$. Let $s = \Pr[S]$ the the probability that the game stops. We now consider three cases, either (i) $s = 1$; or (ii) $K\sqrt{\xi} < s < 1$; or (iii) $s \le K\sqrt{\xi}$. In case (i), by Lemma 27 we have $\text{E}[U] = \text{E}[U \mid S] \ge -2K\sqrt{\xi}$.

In case (ii), by Lemma 27 $\text{E}[U \mid S] \ge -2K\sqrt{\xi}$, since $K\sqrt{\xi} < s$. And by Lemma 29 we have $\text{E}[U \mid \overline{S}] \ge 0$, since $s < 1$. Thus

$$\text{E}[U] \ge s \cdot (-2K\sqrt{\xi}) + (1-s)\cdot 0 \ge -2K\sqrt{\xi}\ .$$

In case (iii), again by Lemma 29 we have $\text{E}[U \mid \overline{S}] \ge 0$, since $s < 1$. Thus

$$\text{E}[U] \ge s \cdot (-K) + (1-s)\cdot 0 \ge -K^2\sqrt{\xi}\ .$$

□

7.2 An ε-optimal strategy that uses $\log\log T$ space for generalized Big Match games

In this section we give a ε-optimal strategy σ_1^* for Player 1 in any generalized Big Match game that for all $\delta > 0$ with probability $1-\delta$ uses $O(\log\log T)$ space. Similarly to how Section 4 showed that the strategy σ_1^*, if initialized correctly, from Section 3 was ε-optimal for Player 1 in the Big Match, we here show that the strategy σ_1^* from Section 7.1, if initialized correctly, is ε-optimal. Similarly to Section 4, the ε-optimal strategy for generalized Big Match games is simply an instantiation of the strategy σ_1^* from Section 7.1, setting $f(T) = \lceil \log T \rceil$. We can then let $\overline{f} = \log T$ and $F(T) = 2^T$. The proofs of the statements in this section is nearly identical to those of Section 4, but there are minor changes and they are thus given here in full.

The claim about the space usage of σ_1^* is already established in Section 7.1. To obtain the stronger property of ε-optimality rather than just ε-supremum-optimality, we just needs to establish a $\liminf$ version of Lemma 29, like how we in 4 gave a $\liminf$ version of Lemma 12.

First we show a technical lemma similarly to Lemma 14. Recall that for a pure Markov strategy σ and a sequence of integers $I = \{i_1, i_2, \ldots, i_m\}$, we have that σ_I is the sequence, $\sigma_{i_1}, \sigma_{i_2}, \ldots, \sigma_{i_m}$. Again, note that $\sigma^k = \sigma_{\{1,\ldots,k\}}$.

Lemma 31. *Let σ be a pure Markov strategy for Player 2, $\delta < 1/4$ be a positive real, and M be a positive integer. Let $\liminf_{T\to\infty} \text{gdens}(\sigma^T) \le -\delta$. Let $\ell_1, \ell_2, \ldots$ be such that for all $i \ge M$, we have that $\ell_i \in [(1-\delta)\cdot(2^{i+1}-1), (1+\delta)\cdot(2^{i+1}-1)]$. Then there exists a sequence $k_2, k_3, \ldots$ such that for infinitely many $i > M$, we have that $\ell_{i-1} + \delta/K2^{i-2} \le k_i \le \ell_i$ and that $\liminf_{i\to\infty} \text{gdens}(\sigma_{\{\ell_{i-1}+1,\ldots,k_i\}}) \le -\frac{\delta}{4}$.*

Proof. The proof follows Lemma 14 with small changes. Let ℓ_i be as required. If there are infinitely many i such that $\text{gdens}(\sigma_{\{\ell_{i-1}+1,\ldots,\ell_i\}}) \le -\frac{\delta}{4}$ then set $k_i = \ell_{i+1}$ and the lemma follows by observing

$(k_i - \ell_{i-1}) \geq (1-\delta)(2^{i+1}-1) - (1+\delta)(2^i - 1) = (1-3\delta)2^i \geq \delta/K2^{i-2}$, for $i > M$. So assume that only for finitely many i, $\text{gdens}(\sigma_{\{\ell_{i-1}+1,\ldots,\ell_i\}}) \leq -\frac{\delta}{4}$. Thus the following claim can be applied for arbitrary large i_0.

Claim 32. *Let $i_0 \geq M$ be given. If for every $i \geq i_0$, $\text{gdens}(\sigma_{\{\ell_{i-1}+1,\ldots,\ell_i\}}) > -\frac{\delta}{4}$ then there exist $j > i_0$ and k such that $\ell_{j-1} + (\delta/K)2^{j-2} \leq k \leq \ell_j$ and $\text{gdens}(\sigma_{\{\ell_{j-1}+1,\ldots,k\}}) \leq -\delta$.*

We can use the claim to find $k_2, k_3, \ldots$ inductively. Start with large enough $i_0 \geq M$ and set $k_i = \ell_i$ for all $i \leq i_0$. Then provided that we already inductively determined $k_2, k_3, \ldots, k_{i_0}$, we apply the above claim to obtain j and k, and we set $k_j = k$ and $k_i = \ell_i$, for all $i = i_0+1, \ldots, j-1$.

So it suffices to prove the claim. For any $d \geq 1$, $\ell_{i_0} \cdot 2^{d-1} \leq \ell_{i_0+d}$ and

$$\text{gdens}(\sigma^{\ell_{i_0+d}}) \geq \frac{-\frac{\delta}{4}(\ell_{i_0+d} - \ell_{i_0}) - K \cdot \ell_{i_0}}{\ell_{i_0+d}} \quad .$$

Furthermore, if $d \geq 1 + \log(4K/\delta)$ then $\ell_{i_0} \leq \frac{\delta}{4K}\ell_{i_0+d}$ and

$$\text{gdens}(\sigma^{\ell_{i_0+d}}) \geq -\frac{\delta}{4} - \frac{\delta}{4} = -\frac{\delta}{2} \quad .$$

Since $\liminf_{k\to\infty} \text{gdens}(\sigma^k) \leq -\delta$, there must be k and $d \geq 1 + \log(4/\delta)$ such that $\ell_{i_0+d-1} \leq k \leq \ell_{i_0+d}$ and $\text{gdens}(\sigma^k) \leq -\delta$. Set $j = i_0 + d$. Also

$$\text{gdens}(\sigma^k) = \frac{\text{gdens}(\sigma^{\ell_{j-1}})\ell_{j-1} + \text{gdens}(\sigma_{\{\ell_{j-1}+1,\ldots,k\}})(k - \ell_{j-1})}{\ell_{j-1} + (k - \ell_{j-1})} \quad ,$$

which means

$$\begin{aligned}\Big(\text{gdens}(\sigma^k) - \text{gdens}(\sigma_{\{\ell_{j-1}+1,\ldots,k\}})\Big)(k - \ell_{j-1}) &= \Big(\text{gdens}(\sigma^{\ell_{j-1}}) - \text{gdens}(\sigma^k)\Big)\ell_{j-1} \\ &\geq \left[-\frac{\delta}{2} - \delta\right]\ell_{j-1} = \frac{\delta}{2}\ell_{j-1} \quad .\end{aligned}$$

Thus $\text{gdens}(\sigma_{\{\ell_{j-1}+1,\ldots,k\}}) \leq \text{gdens}(\sigma^k)$ which in turn is less than $-\delta$. Furthermore, $k - \ell_{j-1} \geq \frac{\delta}{2K}\ell_{j-1} \geq \frac{\delta}{2K}(1-\delta)(2^j - 1) \geq \frac{\delta}{4K}2^j$, provided that $j \geq 2$. Hence, k and j have the desired properties. □

We are now ready to prove the lim inf version of Lemma 29.

Lemma 33. *Let σ be an arbitrary pure Markov strategy for Player 2. If*

$$\liminf_{t\to\infty} \text{gdens}(\sigma_1, \ldots, \sigma_t) < 0 \quad ,$$

then when played against σ_1^* the play stops with probability 1.

Proof. The proof is similarly to Lemma 16 with minor modifications. Let $0 < \delta < \frac{1}{4}$ be such that

$$\liminf_{t\to\infty} \text{gdens}(\sigma_1, \ldots, \sigma_t) \leq -\delta$$

Pick arbitrary $\gamma \in (0,1)$. We will show that with probability at least $1-\gamma$ the game stops, and this implies the statement. Let M be given by Lemma 24 applied for γ and $\delta/2$. Then we have that with probability at least $1-\gamma$, for all $i \geq M$ and $j \in \{1,\ldots,i\}$,

$$t(i,j,i^2) - t(i,j-1,i^2) \in [(1-\frac{\delta}{2})2^i/i, (1+\frac{\delta}{2})2^i/i] \ .$$

Pick $t_{i,j} \in \mathbb{N}$, for $i = 1,2,\ldots$ and $j \in \{1,\ldots,i\}$, so that $t_{i,j-1} < t_{i,j}$ where $t_{i,0}$ stands for $t_{i-1,i-1}$. Let $t_{i,j} - t_{i,j-1} \in [(1-\frac{\delta}{2})2^i/i, (1+\frac{\delta}{2})2^i/i]$, for all $i \geq M$ and $j \in \{1,\ldots,i\}$. Pick M' so that $\frac{\delta}{2}(2^{M'}-1) \geq \max\{t_{M,0}, (1-\frac{\delta}{2})2^M\}$. Define $\ell_i = t_{i,i}$ for all $i \geq 1$. Then for all $i \geq M'$, $\ell_i \in [(1-\delta)\cdot(2^{i+1}-1), (1+\delta)\cdot(2^{i+1}-1)]$ as

$$\begin{aligned}
\ell_i = t_{i,i} &= t_{M,0} + \sum_{M \leq i' \leq i, 1 \leq j \leq i'} t_{i',j} - t_{i',j-1} \\
&\leq \frac{\delta}{2}(2^{M'}-1) + \sum_{M \leq i' \leq i} i' \cdot (1+\frac{\delta}{2})2^{i'}/i' \\
&\leq \frac{\delta}{2}(2^{M'}-1) + (1+\frac{\delta}{2}) \cdot (2^{i+1}-1) \\
&\leq (1+\delta)\cdot(2^{i+1}-1),
\end{aligned}$$

and similarly for the lower bound: $\ell_i \geq \sum_{i',j} t_{i',j} - t_{i',j-1} \geq (1-\frac{\delta}{2})(2^{i+1}-2^M) \geq (1-\delta)\cdot(2^{i+1}-1)$. Thus Lemma 31 is applicable on ℓ_i with M set to M', and we obtain a sequence $k_2, k_3, \ldots$ such that $\mathrm{gdens}(\sigma_{\ell_{i-1}+1,\ldots,k_i}) \leq -\frac{\delta}{4}$ and $k_i - \ell_{i-1} \geq \delta 2^{i-2}/K$ for infinitely many i. Pick any of the infinitely many $i \geq \max\{M', 32(1+\delta)K/\delta\}$ for which $k_i - \ell_{i-1} \geq \delta 2^{i-2}/K$ and $\mathrm{gdens}(\sigma_{\ell_{i-1}+1,\ldots,k_i}) \leq -\frac{\delta}{4}$. Since $\delta 2^{i-3}/K \geq (1+\delta)2^i/i$, there is some $j \in \{1,\ldots,i\}$ such that $\ell_{i-1} + \delta/K 2^{i-3} \leq k_i - (1+\delta)2^i/i \leq t_{i,j} \leq k_i$. Fix such j. Since $k_i \leq t_{i,j} + (1+\delta)2^i/i$, we have

$$\begin{aligned}
\mathrm{gdens}(\sigma_{\ell_{i-1}+1,\ldots,t_{i,j}}) &= \frac{\mathrm{gdens}(\sigma_{\ell_{i-1}+1,\ldots,k_i})(k_i - \ell_{i-1})}{t_{i,j} + \ell_{i-1}} \\
&\leq \frac{\mathrm{gdens}(\sigma_{\ell_{i-1}+1,\ldots,k_i})((1+\delta)2^i/i + t_{i,j} - \ell_{i-1})}{t_{i,j} + \ell_{i-1}} \\
&\leq -\frac{\delta}{4} \cdot \left(1 + \frac{8(1+\delta)}{i}\right) \\
&\leq -\frac{\delta}{4} + \frac{4(1+\delta)}{i} \leq -\frac{\delta}{8}.
\end{aligned}$$

Hence, $\mathrm{gdens}(\sigma_{\ell_{i-1}+1,\ldots,t_{i,j}}) \leq -\frac{\delta}{8}$. So for some $j' \in \{1,\ldots,j\}$, $\mathrm{gdens}(\sigma_{t_{i,j'-1}+1,\ldots,t_{i,j'}}) \leq -\frac{\delta}{8}$. We can state the following claim.

Claim 34. *For i large enough, conditioned on $t(a,b,a^2) = t_{a,b}$, for all $a \geq M$ and all b, and conditioned on that the game did not stop before the time $t_{i,j'-1}+1$, the game stops during times $t_{i,j'-1}+1,\ldots,t_{i,j'}$ with probability at least $\xi^4\omega/2$.*

Conditioned on $t(a,b,a^2) = t_{a,b}$, for all a, b, the claim implies that the game stops with probability 1. Note that the condition is true for some valid choice of $t_{a,b}$ with probability $1-\gamma$. This is because the claim can be invoked for infinitely many i's and for each such i we will have $\xi^4\omega/2$ chance of stopping.

It remains to prove the claim. Assume $t(i, j'-1, i^2) = t_{i,j'-1}$ and $t(i, j', i^2) = t_{i,j'}$. Clearly,

$$\mathrm{gdens}(\sigma_{t_{i,j'-1}+1,\ldots,t_{i,j'}-1}) \leq \mathrm{gdens}(\sigma_{t_{i,j'-1}+1,\ldots,t_{i,j'}}) \cdot (2^{i-1}/(2^{i-1}-1)) \leq -\frac{\delta}{16} \ ,$$

for i large enough. So if we sample i^2-1 times from $\sigma_{t_{i,j'-1}+1}, \ldots, \sigma_{t_{i,j'}-1}$ the generalized density is at most $-\frac{\delta}{16}$ in expectation. By Hoeffding's inequality, Theorem 46 (setting $a_i = -K$ and $b_i = K$, and letting $c_i = \pi(\mathbf{L}, \sigma_{t(i_n,j_n,i)})$), the probability that the generalized density of the subsequence given by the first i^2-1 samples is more than $-\delta/32$ (i.e. $\delta/32$ greater than the expectation) is bounded by

$$2\exp\left(-\frac{2(\delta/32(i^2-1))^2}{(i^2-1)(2K)^2}\right) = 2\exp\left(-\frac{\delta^2}{2048\cdot K^2}(i^2-1)\right) \ .$$

The probability is taken over the possible choices of $t(i, j', 1) < t(i, j', 2) < \cdots < t(i, j', i^2-1)$ assuming $t(i, j'-1, i^2) = t_{i,j'-1}$ and $t(i, j', i^2) = t_{i,j'}$. For i sufficiently large, $2\exp\left(-\frac{\delta^2}{2048\cdot K^2}(i^2-1)\right) \leq 1/2$. Also, whenever

$$\mathrm{dens}(\sigma_{t(i,j',1)}, \sigma_{t(i,j',2)}, \ldots, \sigma_{t(i,j',i^2-1)}) \leq -\frac{\delta}{32}$$

we have at least $\xi^4 \cdot \omega$ chance of stopping by Lemma 26, as Player 1 plays $\tau_1^{i,\varepsilon}$ against

$$\sigma_{t(i,j',1)}, \sigma_{t(i,j',2)}, \ldots, \sigma_{t(i,j',i^2-1)}$$

and $\frac{-\delta}{32} \leq -i \cdot 2K/(i^2-1)$ for sufficiently large i.

Hence, the game stops with probability at least $(1-1/2)\cdot\xi^4\cdot\omega = \xi^4\cdot\omega/2$. The claim, and thus the lemma, follows. □

We can now conclude with the main result of this section.

Theorem 35. *The strategy σ_1^* is $2K^2\sqrt{\xi}$-optimal, and for all $\delta > 0$, with probability at least $1-\delta$ does it use space $O(\log\log T + \log K)$.*

Proof. This is proved just like Theorem 30, except that Lemma 33 is used in place of Lemma 29. □

8 Reduction of repeated games with absorbing states to generalized Big Match games

As explained in Section 7, for defining strategies for repeated games with absorbing states, Kohlberg reduced such games in general form to the special case of generalized Big Match games. The actual terminology, "generalized Big-Match games", is due to Coulomb [3].

Performing the reduction of Kohlberg requires two things. The first thing is to determine the value of the repeated game. Kohlberg showed that the value is the same as the limit of the value of the associated n-stage game as n goes to infinity. The other thing is finding two optimal strategies in an associated parametrized matrix game with certain closeness properties. Here Kohlberg appealed just to semi-continuity of the mapping from the parameter to an optimal strategy of the matrix game. In this section we show how to make these two ingredients efficient, namely by describing polynomial time algorithms for them.

Hansen et.al. [7] recently showed the existence of a polynomial time algorithm for computing the value of any undiscounted stochastic game with a constant number of non-absorbing states. We present below a much simpler algorithm for the case of repeated games with absorbing states based on a characterization of the value of those given by Kohlberg [11]. This algorithm is based only on bisection together with solving linear programs.

The algorithm of Hansen et.al. is in fact similar in spirit, based on bisection and linear programming as well, but is applied to a discounted version of the game with a discount factor for which no explicit expression is readily available.

Additional definitions The *bit-size* of an integer n is the smallest non-zero integer τ such that $|n| < 2^\tau$. Thus $\tau = \lfloor \log_2 |n| \rfloor + 1$ for non-zero n. For a polynomial $p \in \mathbb{Z}[x]$, we denote by $\|p\|_\infty$ the maximum magnitude of a coefficient of p.

8.1 Marginal value of matrix games and value of repeated games

A matrix game is given by a $m \times n$ real matrix $A = (a_{ij})$. The game is played by the two players simultaneously choosing a pure strategy, where Player 1 chooses action i among the m rows and Player 2 chooses action j among the n columns. Hereafter Player 1 receives payoff a_{ij}. A strategy of a player is a probability distribution over the player's actions. Let Δ^n denote the strategies of Player 1 and Δ^m denote the strategies of Player 2. Given $x \in \Delta^m$ and $y \in \Delta^n$, the expected payoff to Player 1 when Player 1 uses strategy x and Player 2 uses strategy y is then $x^\mathsf{T} Ay$. As shown by von Neumann [17] every matrix game A has a value $\mathrm{v}(A)$ in mixed strategies, namely

$$\mathrm{v}(A) = \max_{x \in \Delta^m} \min_{y \in \Delta^n} x^\mathsf{T} Ay = \min_{y \in \Delta^n} \max_{x \in \Delta^m} x^\mathsf{T} Ay \ .$$

Let $O_1(A) \subseteq \Delta^m$ and $O_2(A) \subseteq \Delta^n$ denote the set of optimal strategies for Player 1 and Player 2, respectively. That is, $O_1(A) = \{x \in \Delta^m \mid \forall y \in \Delta^n : x^\mathsf{T} Ay \geq \mathrm{v}(A)\}$ and $O_2(A) = \{y \in \Delta^n \mid \forall x \in \Delta^m : x^\mathsf{T} Ay \geq \mathrm{v}(A)\}$.

Let B be another $m \times n$ real matrix. Mills [12] showed that the limit

$$\frac{\partial\, \mathrm{v}(A)}{\partial B} := \lim_{\alpha \to 0^+} \frac{\mathrm{v}(A + \alpha B) - \mathrm{v}(A)}{\alpha}$$

exists and characterized the limit as the value of the game B when the strategies of Player 1 and Player 2 are restricted to be optimal in A.

Theorem 36 (Mills)**.**

$$\frac{\partial\, \mathrm{v}(A)}{\partial B} = \max_{x \in O_1(A)} \min_{y \in O_2(A)} x^\mathsf{T} By \tag{1}$$

The limit $\frac{\partial\, \mathrm{v}(A)}{\partial B}$ is called the *marginal value* of A with respect to B. It is not hard to see that Equation (1) implies that $\frac{\partial\, \mathrm{v}(A)}{\partial B}$ may be computed using linear programming. Indeed, we may express that simultaneously $x \in O_1(A)$ and $y \in O_2(A)$ by linear equalities and inequalities with auxiliary variable v as[2]:

$$\begin{aligned} f_n v - A^\mathsf{T} x &\leq 0 \qquad & f_m v - Ay &\geq 0 \\ x &\geq 0 & y &\geq 0 \\ f_m^\mathsf{T} x &= 1 & f_n^\mathsf{T} y &= 1 \end{aligned}$$

Thus for fixed x, the quantity $\min_{y \in O_2(A)} x^\mathsf{T} By$ may be computed by the linear program

$$\begin{array}{lrcl} \min & x^\mathsf{T} By & & \\ s.t. & A^\mathsf{T} x' - f_n v & \geq & 0 \\ & f_m v - Ay & \geq & 0 \\ & f_m^\mathsf{T} x' & = & 1 \\ & f_n^\mathsf{T} y & = & 1 \\ & x', y & \geq & 0 \end{array}$$

with auxiliary variables x' and v. Taking the dual we obtain

$$\begin{array}{lrcl} \max & r + s & & \\ s.t. & Ap + f_m r & \leq & 0 \\ & f_n s - A^\mathsf{T} q & \leq & x^\mathsf{T} B \\ & f_m^\mathsf{T} q - f_n^\mathsf{T} p & = & 0 \\ & p, q & \geq & 0 \end{array}$$

in variables p, q, r, s, and then by reintroducing $x \in O_1(A)$ as variables we obtain the following linear program for computing $\frac{\partial \, \mathrm{v}(A)}{\partial B}$

$$\begin{array}{lrcl} \max & r + s & & \\ s.t. & Ap + f_m r & \leq & 0 \\ & f_n s - A^\mathsf{T} q - x^\mathsf{T} B & \leq & 0 \\ & f_m^\mathsf{T} q - f_n^\mathsf{T} p & = & 0 \\ & f_n v - A^\mathsf{T} x & \leq & 0 \\ & Ay' - f_m v & \leq & 0 \\ & f_m^\mathsf{T} x & = & 1 \\ & f_n^\mathsf{T} y' & = & 1 \\ & p, q, x, y' & \geq & 0 \end{array}$$

Appealing to the existence of polynomial time algorithms for linear programming we get:

Corollary 37. *The marginal value $\frac{\partial \, \mathrm{v}(A)}{\partial B}$ can be computed in polynomial time in the bit-size of A and B.*

In this section it will be useful to introduce an alternative notation for repeated games with absorbing states, their *matrix form*. Consider such a game given by action sets A_1 and A_2, the stage payoff function $\pi : A_1 \times A_2 \to \mathbb{R}$, and the absorption probability function $\omega : A_1 \times A_2 \to \mathbb{R}$. We shall now assume $A_1 = \{1, 2, \ldots, m\}$ and $A_2 = \{1, 2, \ldots, n\}$. We then let $b_{ij} = \pi(i, j)$ and $\omega_{ij} = \omega(i, j)$. The game will now be identified by a $m \times n$ matrix $A = (a_{ij})$, populated by the symbolic entries a_{ij} defined by letting $a_{ij} = \omega_{ij} {b_{ij}}^*$ if $\omega_{ij} > 0$ and $a_{ij} = b_{ij}$ if $\omega_{ij} = 0$.

[2]By f_n we mean the vector $(1, \ldots, 1)^\mathsf{T}$ of dimension n.

Let $A = (a_{ij})$ be a $m \times n$ repeated game with absorbing states. The notion of the derived matrix game $\widetilde{A} = (\widetilde{a}_{ij})$ obtained from A is generalized from the definition of Section 7 to be given by

$$\widetilde{a}_{ij} = \begin{cases} \omega_{ij} b_{ij} & \text{if } a_{ij} = \omega_j b_{ij}{}^* \\ b_{ij} & \text{if } a_{ij} = b_{ij} \end{cases}$$

Also, given reals u and t, we define an associated $m \times n$ matrix game denoted $A(u,t)$, letting entry (i,j) be

$$\begin{aligned} A(u,t)_{ij} &= \omega_{ij} b_{ij} + (1-\omega_{ij})(t b_{ij} + (1-t)u) \\ &= \omega_{ij} b_{ij} + (1-\omega_{ij})u + t((1-\omega_{ij})(b_{ij}-u)) \end{aligned}$$

Let A_1 be the $m \times n$ matrix with entries $\omega_{ij} b_{ij} + (1-\omega_{ij})u$, and A_2 be the $m \times n$ matrix with entries $(1-\omega_{ij})(b_{ij}-u)$. In other words, we write $A(u,t) = A_1 + tA_2$. Then the limit

$$\frac{\partial\, \mathrm{v}(A(u,t))}{\partial t^+} := \lim_{t \to 0^+} \frac{\mathrm{v}(A(u,t)) - \mathrm{v}(A(u,0))}{t} = \frac{\partial\, \mathrm{v}(A_1)}{\partial A_2}$$

exists and may be computed in polynomial time by Theorem 36 and Corollary 37. Define the extended real number $\Delta_A(u)$ by the limit

$$\Delta_A(u) := \lim_{t \to 0^+} \frac{\mathrm{v}(A(u,t)) - u}{t} \quad .$$

Clearly, $\Delta_A(u) = \frac{\partial\, \mathrm{v}(A(u,t))}{\partial t^+}$ when $\mathrm{v}(A(u,0)) = u$, and otherwise $\Delta_A(u)$ is ∞ or $-\infty$ depending on whether $\mathrm{v}(A(u,0)) > u$ or $\mathrm{v}(A(u,0)) < u$. Kohlberg showed that the value of A can be characterized by $\Delta_A(u)$.

Theorem 38 (Kohlberg)**.** *Let A be a repeated game with absorbing states. The value of A is the unique point u_0 for which*

$$u < u_0 \Rightarrow \Delta_A(u) > 0$$

and

$$u > u_0 \Rightarrow \Delta_A(u) < 0 \ .$$

Using this characterization and bisection together with Corollary 37 yields a very simple algorithm for approximating the value of a repeated game with absorbing states.

Proposition 39. *There is an algorithm that given a repeated game with absorbing states A and $\varepsilon > 0$ computes the value of A to within an additive error ε in polynomial time in the bit-size of A and* $\log(1/\varepsilon)$*.*

8.2 Parametrized Matrix Games

The value of a $m \times n$ matrix game A as well as an optimal strategy for Player 1 may be computed by the following linear program in variables (x, v).

$$\begin{array}{lrcl} \max & v & & \\ s.t. & f_n v - A^{\mathsf{T}} x & \leq & 0 \\ & x & \geq & 0 \\ & f_m^T x & = & 1 \end{array} \tag{2}$$

A *basic solution* to LP (2) is obtained by selecting $m+1$ constraints indexed by B, that includes the equality constraint. This gives the $(m+1)\times(m+1)$ matrix M_B^A, consisting of the coefficients of these constraints, appropriately ordered (we shall assume the equality constraint is ordered last). A basic solution is determined by B if M_B^A is non-singular, and in that case it is $(x,v)^{\mathsf{T}} = (M_B^A)^{-1}e_{m+1}$[3]. By Cramer's rule $x_i = \det((M_B^A)_i)/\det(M_B^A)$ and $v = \det((M_B^A)_{m+1})/\det(M_B^A)$, where $(M_B^A)_i$ is the matrix obtained from M_B^A by replacing column i with e_{m+1}. The basic solution $(x,v)^{\mathsf{T}}$ is a *basic feasible solution* (bfs) if also $x \geq 0$ and $f_n v - A^{\mathsf{T}} x \leq 0$.

We consider the setting where each entry of A is a linear function in a variable t. Let $A(t)$ denote this matrix game. When $t_0 > 0$ is sufficiently small, then if B defines an optimal basic feasible solution for $A(t_0)$, then B also defines an optimal basic feasible solution for any $0 < t \leq t_0$. We give an explicit bound for this using the following fundamental root bound.

Lemma 40. *[18, Chapter 6, equations (4) and (5)] Let $f \in \mathbb{Z}[x]$ be a non-zero integer polynomial. Then for any non-zero root γ of f it holds $(2\|f\|_\infty)^{-1} < |\gamma| < 2\|f\|_\infty$.*

Using this we have the following precise statement.

Proposition 41. *Let $A(t)$ be a $m\times n$ matrix game parametrized by t, where each entry is a linear function in t with integer coefficients of bit-size at most τ. Let $t_0 = (4((m+1)2^{\tau+1})^{2(m+1)})^{-1}$. If B defines an optimal bfs for $A(t_0)$ then B also defines an optimal bfs for $A(t)$ for all $0 < t \leq t_0$.*

Proof. Let $P_i^B(t) = \det((M_B^A)_i)$ and $Q^B(t) = \det(M_B^A)$. These are polynomials of degree at most $(m+1)$ having coefficients of magnitude at most $(m+1)!2^{m+1}(2^\tau)^{m+1} \leq ((m+1)2^{\tau+1})^{m+1}$. By Lemma 40 we then have that $\operatorname{sgn}(Q^B(t)) = \operatorname{sgn}(Q(t_0)$ and $\operatorname{sgn}(P_i^B(t)) = \operatorname{sgn}(P_i^B(t_0))$ for all $0 < t \leq t_0$. This means that if B defines a bfs for t_0, then B also defines a bfs for all $0 < t < t_0$. To ensure that the bfs defined by B is optimal we shall compare it with any other bfs defined by a different set B'. We then need to ensure that $\frac{P_{m+1}^B(t)}{Q^B(t)} \geq \frac{P_{m+1}^{B'}(t)}{Q^{B'}(t)}$. For this we consider the polynomial $H(t) = P_{m+1}^B(t)Q^{B'}(t) - P_{m+1}^{B'}Q^B(t)$. Note that H is a polynomial of degree at most $2(m+1)$ having coefficients of magnitude at most $2((m+1)2^{\tau+1})^{2(m+1)}$. Then by Lemma 40 again, also $\operatorname{sgn}(H(t)) = \operatorname{sgn}(H(t_0))$ for all $0 < t < t_0$, which means that if B defines a bfs that is also optimal for t_0 the bfs it defines for all $0 < t \leq t_0$ is optimal as well. $\square$

8.3 Reduction to generalized Big Match games

We give here an effective version of the reduction of Kohlberg of repeated games with absorbing states to the special case of generalized big match games [11, Lemma 2.8 and Theorem 2.1]. We additionally make the (rather simple) extension to repeated games with generalized absorbing states.

We shall need the following lemma.

Lemma 42. *Let P and Q be integer polynomials such that $\lim_{t\to 0^+}\frac{P(t)}{Q(t)}$ exists. Let $\eta = 1/k$ for a positive integer k and suppose $\|P\|_\infty \leq M$ as well as $\|Q\|_\infty \leq M$. Then*

$$\left|\frac{P(t)}{Q(t)} - \lim_{t\to 0^+}\frac{P(t)}{Q(t)}\right| < \eta \tag{3}$$

whenever $0 < t \leq t_0 = (6kM^2)^{-1}$.

[3] By e_n we mean the standard nth unit vector of appropriate dimension.

Proof. Note that $\lim_{t \to 0^+} \frac{P(t)}{Q(t)} = \frac{a}{b}$, where b is the non-zero coefficient of Q of lowest degree, and a is the coefficient of P of the same degree. Let $H_1(t) = k(bP(t) - aQ(t)) - bQ(t)$ and $H_2(t) = k(aQ(t) - bP(t)) - bQ(t)$. Then Equation (3) holds if and only if $H_1(t) < 0$ and $H_2(t) < 0$. Noting that $\|H_1\|_\infty \leq 3kM^2$ as well as $\|H_2\|_\infty \leq 3kM^2$ the conclusion follows from Lemma 40. $\square$

Let A' be a $m \times n$ repeated game with absorbing states with stage payoffs b'_{ij} and stopping probabilities ω_{ij}. For an optimal strategy x in $A'(0,0)$, define:

$$\begin{aligned}
\omega_j &= \sum_{i=1}^{m} x_i \omega_{ij} \\
b_j &= \begin{cases} \frac{1}{\omega_j} \sum_{i=1}^m x_i \omega_{ij} b'_{ij} & \text{if } \omega_j > 0 \\ 0 & \text{if } \omega_j = 0 \end{cases} \\
e_j &= \begin{cases} \frac{1}{1-\omega_j} \sum_{i=1}^m x_i (1-\omega_{ij}) b'_{ij} & \text{if } \omega_j < 1 \\ 0 & \text{if } \omega_j = 1 \end{cases}
\end{aligned} \tag{4}$$

and similarly for an optimal strategy $x(t)$ in $A'(0,t)$ for some given $t > 0$, define:

$$\begin{aligned}
\omega(t)_j &= \sum_{i=1}^{m} x(t)_i \omega_{ij} \\
b(t)_j &= \begin{cases} \frac{1}{\omega(t)_j} \sum_{i=1}^m x(t)_i \omega_{ij} b'_{ij} & \text{if } \omega(t)_j > 0 \\ 0 & \text{if } \omega(t)_j = 0 \end{cases} \\
e(t)_j &= \begin{cases} \frac{1}{1-\omega(t)_j} \sum_{i=1}^m x(t)_i (1-\omega_{ij}) b'_{ij} & \text{if } \omega(t)_j < 1 \\ 0 & \text{if } \omega(t)_j = 1 \end{cases}
\end{aligned} \tag{5}$$

Suppose that the ω_{ij}'s are rational numbers with common denominator β_1 and the nominators and β_1 are of bit-size τ_1. Similarly suppose that the b'_{ij}'s are rational numbers with common denominator β_2 and the nominators and β_2 are of bit-size τ_2. By definition $A'(0,t)_{ij} = \omega_{ij} b'_{ij} + t(1-\omega_{ij}) b'_{ij}$. Thus the entries of $A'(0,t)$ are linear functions in t where the coefficients are rational numbers with common denominator $\beta = \beta_1 \beta_2$ and the bit-sizes of the nominators and denominators are at most $\tau = \tau_1 + \tau_2$. Multiplying each entry of $A'(0,t)$ by β only scales every bfs, so setting $t_0 = (4((m+1)2^{\tau+1})^{2(m+1)})^{-1}$, whenever B defines an optimal bfs in $\beta A'(0,t_0)$ it also defines an optimal bfs for $A'(0,t)$ for all $0 < t \leq t_0$, by Proposition 41. So let B define an optimal bfs for $A'(0,t_0)$. Let now $P_i(t) = \det((M_B^{\beta A'(0,t)})_i)$ and $Q(t) = \det(M_B^{\beta A'(0,t)})$, and define $x_i(t) = P_i(t)/Q(t)$. $P_i(t)$ and $Q(t)$ are polynomials of degree at most $m+1$ having integer coefficients of magnitude at most $((m+1)2^{\tau+1})^{m+1}$. Furthermore is $x(t)$ an optimal strategy in $A'(0,t)$ for all $0 < t \leq t_0$. Let $x = \lim_{t \to 0^+} x(t)$. Then x is an optimal strategy in $A'(0,0)$. Each coordinate x_i is the ratio between two coefficients from $P_i(t)$ and $Q(t)$ and is therefore a rational number with nominator and denominator of magnitude at most $((m+1)2^{\tau+1})^{m+1}$.

We have always

$$\omega(t)_j = \sum_{i=1}^{m} \frac{P_i(t)}{Q(t)} \omega_{ij} = \frac{\sum_{i=1}^m P_i(t) \omega_{ij}}{Q(t)}$$

and

$$1-\omega(t)_j = \sum_{i=1}^m x(t)_i(1-\omega_{ij}) = \frac{\sum_{i=1}^m P_i(t)(1-\omega_{ij})}{Q(t)}$$

Consider now the case of $\omega(t)_j > 0$. Then

$$b(t)_j = \frac{Q(t)}{\sum_{i=1}^m P_i(t)\omega_{ij}} \sum_{i=1}^m \frac{P_i(t)}{Q(t)}\omega_{ij}b'_{ij} = \frac{\sum_{i=1}^m P_i(t)\omega_{ij}b'_{ij}}{\sum_{i=1}^m P_i(t)\omega_{ij}}$$

Consider now the case of $\omega(t)_j < 1$. Then

$$e(t)_j = \frac{Q(t)}{\sum_{i=1}^m P_i(t)(1-\omega_{ij})} \sum_{i=1}^m \frac{P_i(t)}{Q(t)}(1-\omega_{ij})b'_{ij} = \frac{\sum_{i=1}^m P_i(t)(1-\omega_{ij})b'_{ij}}{\sum_{i=1}^m P_i(t)(1-\omega_{ij})}$$

Since $x = \lim_{t\to 0^+} x(t)$ we also have $\lim_{t\to 0^+} \omega_j(t) = \omega_j$. Furthermore, in case that $\omega_j > 0$, we have $\lim_{t\to 0^+} b_j(t) = b_j$, and in case $\omega_j = 0$, we have $\lim_{t\to 0^+} e_j(t) = e_j$.

Let $\eta = 1/k$ for a positive integer k. Define $t_1 = (2k((m+1)2^{\tau+1})^{m+2})^{-1}$ and $t_2 = (6k((m+1)2^{\tau+1})^{2(m+1)})^{-1}$. Note that $t_2 \le \min(t_0, t_1)$.

Suppose that $\omega_j > 0$. Then $\omega(t)_j > 0$ if $\sum_{i=1}^m P_i(t)\omega_{ij} > 0$. Since $\omega_j > 0$ there exists i such that $P_i(t)$ is a non-zero polynomial and $\omega_{ij} > 0$. Then by Lemma 40 it follows that $\omega(t)_j > 0$ for all $0 < t \le t_0$. Also from Lemma 42, we have $|b_j - b(t)_j| < \eta$ whenever $0 < t \le t_1$. Suppose now that $\omega_j = 0$. Consider the polynomial $H(t) = \frac{\beta_1}{\eta}\sum_{i=1}^m P_i(t)\omega_{ij} - \beta_1 Q(t)$. We then have $\omega(t)_j < \eta$ if and only if $H(t) < 0$, for $0 < t \le t_0$. The polynomial H is of degree at most $m+1$ and has integer coefficients of magnitude at most $k((m+1)2^{\tau+1})^{m+2}$. By Lemma 40 we have $\omega(t)_j < \eta$ whenever $0 < t \le \min(t_0, t_2)$. Also from Lemma 42 we have $|e_j - e(t)_j| < \eta$, whenever $0 < t \le t_1$.

Putting all these observations together gives us the following effective version of [11, Lemma 2.8].

Lemma 43. *There is an algorithm that given a $m \times n$ repeated game A' with absorbing states as above and $\eta = 1/k$ computes in polynomial time strategies x and $x(t)$ that are optimal for Player 1 in $A(0,0)$ and $A(0,t_1)$, respectively, such that*

1. *If $\omega_j > 0$ then $\omega(t_2)_j > 0$ and $|b_j - b(t_2)_j| < \eta$.*
2. *If $\omega_j = 0$ then $\omega(t_2)_j < \eta$ and $|e_j - e(t_2)_j| < \eta$.*

where $t_2 = (6k((m+1)2^{\tau+1})^{2(m+1)})^{-1}$.

Proof. By Proposition 41 if B defines an optimal bfs for the matrix game $A'(0,t_2)$, then B also defines an optimal bfs for $A'(0,t)$ for all $0 < t' < t$. We may find such a B simply by solving the LP (2) for $A'(0,t_2)$. Let now $P_i(t) = \det((M_B^{A'(0,t)})_i)$ and $Q(t) = \det(M_B^{A'(0,t)})$ as above. We can compute these polynomials, which are of degree at most $m+1$ by evaluating them on $m+2$ distinct points from the interval $(0,t_2)$ and interpolating. We can furthermore do this evaluation using LP (2) since we already have found B. This further gives us $x_i(t) = P_i(t)/Q(t)$, and $x(t)$ is an optimal strategy in the matrix game $A(0,t)$ for all $0 < t \le t_2$. Let $x = \lim_{t\to 0^+} x(t)$ be the optimal limit strategy in $A'(0,0)$. As in the proof of Lemma 42 we may compute each coordinate x_i by considering the non-zero coefficients of the lowest degree of $P_i(t)$ and $Q(t)$. We conclude that $x(t_2)$ and x together satisfy the required properties. $\square$

We will now show how to reduce an arbitrary repeated games with absorbing states to generalized Big Match games. By reduction we mean that a generalized Big Match game D is computed from a repeated game with absorbing states A, such that a strategy σ_1 for D can be extended to a strategy τ_1 for A. In case σ_1 is ε' optimal for D then τ_1 is ε-optimal for A, and likewise, in case σ_1 is ε'-supremum-optimal for D then τ_1 is ε-supremum-optimal for A, where ε' depends on ε and A.

Theorem 44. *Let A be a $m \times n$ repeated game with absorbing states and let $\varepsilon = 2^{-\ell}$. Assume the stage payoff b_{ij} are rational numbers such that $|b_{ij}| \leq 1$. Assume the stopping probabilities ω_{ij} are rational numbers with common denominator β_1 and the nominator and β_1 are of bit-size at most τ_1. Then A can be reduced in polynomial time to a generalized Big Match game satisfying Assumption 1 with integer entries of magnitude at most $(24(m+2)2^{\ell+\tau_1+1})^{20(m+2)^2(2n+1)}$.*

Proof. We will have 4 sources of error: In approximating the value of A, in rounding the entries of A, from the strategy for the generalized Big Match to which we reduce, and finally from additional strategies of Player 2 that are not part of this. We shall allow $\varepsilon/4$ to all these.

First we use the Algorithm of Proposition 39 to compute u such that

$$u + \frac{\varepsilon}{2} \leq \mathrm{v}(A) < u + \frac{3\varepsilon}{4} \quad . \tag{6}$$

Using u we translate and round the entries of A to obtain another repeated game with absorbing states A' with the same stopping probabilities but with stage payoff b'_{ij} given by

$$b'_{ij} = \left\lfloor \frac{b_{ij} - u}{\varepsilon/4} \right\rfloor \frac{\varepsilon}{4} \quad .$$

Using Equation (6) we have that $\varepsilon/4 \leq \mathrm{v}(A') < 3\varepsilon/4$. Also the rounded and translated stage payoffs b'_{ij} satisfy $-2 \leq b'_{ij} \leq 2$ and are rational numbers with common denominator $\beta_2 = 4/\varepsilon = 2^{\ell+2}$ and nominators of bit-size at most $\tau_2 = \ell + 3$.

Since $\mathrm{v}(A') > 0$ from Theorem 38 we have that $\Delta_{A'}(0) > 0$, and this means that $\mathrm{v}(A'(0,t))$ can be bounded below by a linear function in an interval to the right of 0. We shall make this explicit below, providing constants δ and t_1 such that

$$\mathrm{v}(A'(0,t)) \geq \delta t \tag{7}$$

whenever $0 \leq t \leq t_1$.

So $\Delta_{A'}(0) = \lim_{t \to 0^+} \frac{\mathrm{v}(A'(0,t))}{t} > 0$. We first fix $\delta > 0$ and then determine a corresponding t_1. If $\mathrm{v}(A'(0,0)) > 0$ we may choose any $\delta > 0$. If $\mathrm{v}(A'(0,0)) = 0$ we should choose δ such that $\delta < \Delta_{A'}(0)$.

Let $\tau = \tau_1 + \tau_2$ and $\beta = \beta_1 \beta_2$. Scaling the entries of $A'(0,t)$ by β and setting $t_0 = (4((m+1)2^{\tau+1})^{2(m+1)})^{-1}$, whenever B defines an optimal bfs for $\beta A'(0,t_0)$ it also defines an optimal bfs for $A'(0,t)$ for all $0 < t \leq t_0$ by Proposition 41. So let B define an optimal bfs for $A'(0,t_0)$. Let now $P(t) = \det((M_B^{\beta A'(0,t)})_{m+1})$ and $Q(t) = \det(M_B^{\beta A'(0,t)})$. Then when $0 < t \leq t_0$ we have $\mathrm{v}(A'(0,t)) = \frac{P(t)}{\beta Q(t)}$. The polynomials P and Q are of degree at most $m+1$ and having integer coefficients of magnitude at most $((m+1)2^{\tau+1})^{m+1}$.

Suppose that $\mathrm{v}(A'(0,0)) = 0$. Then

$$\Delta_{A'}(0) = \frac{\partial\, \mathrm{v}(A'(0,t))}{\partial t^+} = \lim_{t \to 0^+} \frac{\frac{d}{dt}P(t)Q(t) - P(t)\frac{d}{dt}Q(t)}{\beta(Q(t))^2} \quad .$$

Thus $\Delta_{A'}(0)$ is the ratio between the coefficients of integers polynomial where the denominator has coefficients of maximum magnitude $2^\tau((m+1)2^{\tau+1})^{2(m+1)}$. It follows that

$$\frac{1}{2}\Delta_{A'}(0) \geq (((m+1)2^{\tau+1})^{2m+3})^{-1}$$

so we let $\delta = (((m+1)2^{\tau+1})^{2m+3})^{-1}$. To determine t_1 we need to ensure that $\mathrm{v}(A'(0,t)) \geq \delta t$. To this end, define the polynomial $H(t) = P(t)/\delta - \beta t Q(t)$. This is an integer polynomial of degree at most $m+2$ and $\|H\|_\infty \leq 2((m+1)2^{\tau+1})^{3m+4}$. By Lemma 40, letting $t_1 = (4((m+1)2^{\tau+1})^{3m+4})^{-1}$ we obtain the desired Equation (7).

Let $\eta = \delta/4$ and use the algorithm from Lemma 43 to compute strategies $x(t_2)$ and x, where $t_2 = (24((m+1)2^{\tau+1})^{4(m+1)})^{-1}$. Note that $t_2 \leq t_1$. Also note for later that $\delta \leq 2^{-\tau_2} = \varepsilon/8$.

We may now proceed as in [11, Theorem 2.1]. Player 1 will commit to at every stage playing either the strategy x or the strategy $x(t_2)$. In this way Player 1 becomes restricted to the $2 \times n$ repeated game with absorbing states $C = (c_{ij})$, where

$$c_{1j} = \begin{cases} \omega_j b_j{}^* & \text{if } \omega_j > 0 \\ e_j & \text{if } \omega_j = 0 \end{cases} \tag{8}$$

and similarly

$$c_{2j} = \begin{cases} \omega(t_2)_j b(t_2)_j{}^* & \text{if } \omega(t_2)_j > 0 \\ e(t_2)_j & \text{if } \omega(t_2)_j = 0 \end{cases} \tag{9}$$

Since x is optimal in $A'(0,0)$ Equation (7) gives for all j,

$$\omega_j b_j = \sum_{i=1}^m x_i \omega_{ij} b'_{ij} \geq 0 \tag{10}$$

and similarly since $x(t_2)$ is optimal in $A'(0,t_2)$ Equation (7) gives for all j,

$$\omega(t_2)_j b(t_2)_j + (1-\omega(t_2)_j) t_2 e(t_2)_j = \sum_{i=1}^m x(t_2)_i \left(\omega_{ij} b'_{ij} + t_2(1-\omega_{ij} b'_{ij})\right) \geq \delta t_2 \tag{11}$$

Let $J = \{j \in \{1,\ldots,n\} \mid \omega_j = 0 \text{ and } \omega(t_2)_j > 0\}$, and consider any $j \in J$. Since $\omega_j = 0$, Lemma 43 gives $\omega(t_2)_j < \eta = \delta/4$. Since $|b'_{ij}| \leq 2$, we then get $\omega(t_2)_j e(t_2)_j \leq \delta/2$, and Equation (11) gives

$$\omega(t_2)_j b(t)_j + t_2 e(t_2)_j \geq \delta t_2/2 \ .$$

Also from Lemma 43 we have $|e_j - e(t_2)_j| < \eta \leq \delta/2$, which means $e(t_2)_j \geq e_j - \delta/2$, which in turn means we have

$$\omega(t_2)_j b(t_2)_j + t_2 e_j \geq 0 \ .$$

Let $\widetilde{C} = (\widetilde{c}_{ij})$ be the derived matrix game from C. For $j \in J$, $\widetilde{c}_{1j} = e_j$ and $\widetilde{c}_{2j} = \omega(t_2)_j b(t_2)_j$. Thus, dividing by $1+t_2$ we get

$$\frac{1}{1+t_2}\widetilde{c}_{2j} + \frac{t_2}{1+t_2}\widetilde{c}_{1j} \geq 0 \ ,$$

which means that the value of the matrix game $\widetilde{C}$ restricted to the columns of J is at least 0. We define a $2 \times |J|$ repeated game with absorbing states $C' = (c'_{ij})$ by restricting C to the columns J

and subtracting a value from each entry such the value of the derived matrix game $\widetilde{C'}$ is 0. More precisely, let v be the value of $\widetilde{C}$ restricted to the columns of J, and for $j \in J$ we let $c'_{1j} = e_j - v$ and $c'_{2j} = \omega(t_2)_j(b(t_2) - v/\omega(t_2)_j)^*$.

Using the expressions previously obtained for $\omega(t)_j$, $b(t)_j$, and $e(t)_j$, we have that for $0 < t \leq t_0$, each of $\omega(t)_j$, $b(t)_j$, $e(t)_j$, and $\omega(t)_j b(t)_j$ can be expressed as rational functions of integer polynomials of degree at most $m+1$ and integer coefficients of magnitude at most $((m+1)2^{\tau+1})^{m+2}$. This means that $e_j = \lim_{t\to 0^+} e(t)_j$ is a rational number with nominator and denominator of magnitude at most $((m+1)2^{\tau+1})^{m+2}$ as well. We can bound the nominator and denominator of the numbers $\omega(t_2)_j$, $b(t_2)_j$, and $\omega(t_2)_j b(t_2)_j$ by estimating the magnitudes after substitution of t_2 in the corresponding rational functions. This yield that they are rational numbers with nominator and denominator of magnitude at most $((m+1)2^{\tau+1})^{m+1}(24((m+1)2^{\tau+1})^{4(m+1)})^{m+2} \leq (24(m+1)2^{\tau+1})^{4(m+2)^2}$. Now the value v is given by the value of a 2×2 sub-game of the matrix game $\widetilde{C}$ restricted to the columns of J. This in turn means that v has nominator and denominator of magnitude at most $4(24(m+1)2^{\tau+1})^{16(m+2)^2}$.

We can now estimate the entries of $\widetilde{C'} = (\widetilde{c'}_{ij})$. These are just the entries from $\widetilde{C}$ subtracted v. Hence they all have nominator and denominators of magnitude at most $8(24(m+1)2^{\tau+1})^{20(m+2)^2} \leq (24(m+2)2^{\tau+1})^{20(m+2)^2}$.

We now scale the entries of C' obtaining another repeated game with absorbing states D such that the entries of $\widetilde{D}$ are integers. We simply do this by multiplying by least common multiple M of all the denominators of the entries of $\widetilde{C'}$. Note that $M \leq (24(m+2)2^{\tau+1})^{40(m+2)^2 n}$, which makes the entries of $\widetilde{D}$ integers of magnitude at most $K = (24(m+2)2^{\tau+1})^{20(m+2)^2(2n+1)}$.

In case $\widetilde{D}$ does not have a pure optimal strategy, then D satisfies Assumption 1, and we let σ_1 be a memory based strategy for Player 1 for D with action map σ_1^a and update map σ_1^u that is either $\varepsilon/(4M)$-optimal or $\varepsilon/(4M)$-supremum optimal. In case that $\widetilde{D}$ has a pure optimal strategy we simple take σ_1 to be the strategy that plays this pure action always.

From σ_1 we now construct a strategy τ_1 for A. The action map τ_1^a will sample an action from σ_1^a. In case of a **L** sample, τ_1^a will sample the final action from x and in case of a **R** sample, τ_1^a will sample the final action from $x(t_2)$. The update map will be a simple filtering map τ_1^u given as follows. Let (m, j) be a pair of a memory state m and an action j of Player 2. In case $j \in J$ we let $\tau_1^u(m, j) = \sigma_1^u(m, j)$. But if $j \notin J$ we let $\tau_1^u(m, j)$ stay in the memory state m, that is we let the next state be m with probability 1.

Looking at the rounds where $j \in J$, the strategy τ_1 inherits the performance of σ_1. Consider now $j \notin J$. Then we have either (a) $\omega_j = 0$ and $\omega(t_2)_j = 0$ or (b) $\omega_j > 0$. In case (a) Equation (11) gives $e(t_2) \geq \delta$ and from Lemma 43 follows $e_j \geq 0$. In case (b) Equation (10) gives $b_j \geq 0$. From Lemma 43 follows $\omega(t_2)_j > 0$ as well as $b(t_2)_j \geq -\eta = -\delta/4 \geq -\varepsilon/32$. Thus in each case the expected stage payoff is at least $u - \varepsilon/32 \geq \mathrm{v}(A) - \epsilon$. □

Note that $\log K = O(m^2 n(\tau + \log m)) = O(m^2 n(\log 1/\varepsilon + \tau_1 + \log m))$, which means that for each $\delta > 0$ with probability at least $1 - \delta$ the resulting strategy ε-supremum optimal strategy will use space $O(f(T) + m^2 n(\log 1/\varepsilon + \tau_1 + \log m))$ and the resulting ε-optimal strategy will use space $O(\log\log T + m^2 n(\log 1/\varepsilon + \tau_1 + \log m))$.

References

[1] R. J. Aumann. Survey of repeated games. In V. Bohm, editor, *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, volume 4 of *Gesellschaft, Recht, Wirtschaft*, pages 11–42. Bibliographisches Institut, Mannheim, 1981.

[2] D. Blackwell and T. S. Ferguson. The big match. *The Annals of Mathematical Statistics*, 39(1):159–163, 1968.

[3] J. M. Coulomb. Generalized "big-match". *Mathematics of Operations Research*, 24(4):795–816, 1999.

[4] P. Flajolet. Approximate counting: A detailed analysis. *BIT*, 25(1):113–134, 1985.

[5] D. Gillette. Stochastic games with zero stop probabilities. In *Contributions to the Theory of Games III*, volume 39 of *Ann. Math. Studies*, pages 179–187. Princeton University Press, 1957.

[6] K. Hansen, M. Koucký, and P. Miltersen. Winning concurrent reachability games requires doubly exponential patience. In *Proc. of IEEE Symp. on Logic in Comp. Sci., LICS*, pages 332–341, 2009.

[7] K. A. Hansen, M. Koucký, N. Lauritzen, P. B. Miltersen, and E. P. Tsigaridas. Exact algorithms for solving stochastic games. In *STOC 2011*, pages 205–214. ACM, 2011.

[8] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[9] R. Ibsen-Jensen. *Strategy complexity of two-player, zero-sum games.* PhD thesis, Aarhus University, 2013.

[10] E. Kalai. Bounded rationality and strategic complexity in repeated games. In T. Ichiishi, A. Neyman, and Y. Tauman, editors, *Game Theory and Applications*, pages 131–157. Academic Press, 1990.

[11] E. Kohlberg. Repeated games with absorbing states. *The Annals of Statistics*, 2(4):724–738, 1974.

[12] H. D. Mills. Marginal values of matrix games and linear programs. In H. W. Kuhn and A. W. Tucker, editors, *Linear Inequalities and Related Systems*, volume 38 of *Annals of Mathematics Studies*, pages 183–193. Princeton University Press, 1956.

[13] R. Morris. Counting large numbers of events in small registers. *Communications of the ACM*, 21(10):840–842, 1978.

[14] L. Shapley. Stochastic games. *Proc. Natl. Acad. Sci. U. S. A.*, 39:1095–1100, 1953.

[15] S. Sorin. Repeated games with complete information. In *Handbook of Game Theory with Economic Applications*, volume 1, chapter 4, pages 71–107. Elsevier, 1 edition, 1992.

[16] S. Sorin. *A First Course on Zero Sum Repeated Games.* Springer, 2002.

[17] J. von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, 1928.

[18] C. K. Yap. *Fundamental Problems of Algorithmic Algebra.* Oxford University Press, New York, 2000.

A Tail inequalities

Theorem 45 (Multiplicative Chernoff bound)**.** *Let $X = \sum_{i=1}^{n} X_i$ where $X_1, \ldots, X_n$ are random variables independently distributed in $[0,1]$. Then for any $\varepsilon > 0$*

$$\Pr[X \geq (1+\varepsilon)\mathrm{E}[X]] \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{(1+\varepsilon)}}\right)^{\mathrm{E}[X]} \leq \exp\left(-\frac{\varepsilon^2}{2+\varepsilon}\mathrm{E}[X]\right),$$

and

$$\Pr[X \leq (1-\varepsilon)\mathrm{E}[X]] \leq \left(\frac{e^{-\varepsilon}}{(1-\varepsilon)^{(1-\varepsilon)}}\right)^{\mathrm{E}[X]} \leq \exp\left(-\frac{\varepsilon^2}{2}\mathrm{E}[X]\right) \quad .$$

Hoeffding [8] gave the following bound for sampling without replacement.

Theorem 46 (Hoeffding)**.** *Let a population C consist of N values $c_1, \ldots, c_N$, where $a_i \leq c_i \leq b_i$. Let $X_1, \ldots, X_n$ denote a random sample without replacement from C and $X = \sum_{i=1}^{n} X_i$. Then*

$$\Pr[|X - \mathrm{E}[X]| \geq t] \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) \quad .$$