SOLVENCY GAMES

NOAM BERGER, NEVIN KAPUR, LEONARD J. SCHULMAN, AND VIJAY V. VAZIRANI

Abstract. We study the decision theory of a maximally risk-averse investor — one whose objective, in the face of stochastic uncertainties, is to minimize the probability of ever going broke. With a view to developing the mathematical basics of such a theory, we start with a very simple model and obtain the following results: a characterization of best play by investors; an explanation of why poor and rich players may have different best strategies; an explanation of why expectation-maximization is not necessarily the best strategy even for rich players. For computation of optimal play, we show how to apply the Value Iteration method, and prove a bound on its convergence rate.

1. Introduction

A key concern in computer science and operations research is *decision-making under uncertainty*. We define a very simple game that helps us study the issue of solvency, or indefinite survival, in the presence of stochastic uncertainties. In Section 1.1 below we provide some motivating reasons for studying this issue.

We start by defining the model. A state of the game is an integer, which we call the wealth of the player. An *action* (representing, say, an investment choice) is a finitely supported probability distribution on the integers; this distribution specifies the probabilities with which various payoffs are received, if this action is chosen. Let w be the wealth of the player at time t. Let $\mathcal{A}$ be a set of actions. Suppose that after choosing a particular action from $\mathcal{A}$, the random variable sampled from that action is a. Then at time $t+1$ the wealth of the player is $w+a$. The game terminates if the player goes broke (wealth becomes ≤ 0). A *strategy* π for the set $\mathcal{A}$ of actions is a function $\pi : \mathbb{Z}_+ \to \mathcal{A}$ specifying the action that is chosen at each possible value of wealth. Corresponding to strategy π, define

$$p^\pi(w) = \mathbf{Pr}[\text{ever going broke, starting from wealth } w],$$

for each $w \in \mathbb{Z}_+$. The object of interest is a strategy that minimizes $p^\pi(w)$ for each value of $w \in \mathbb{Z}_+$. In this notation there are two implicit assumptions regarding an optimal strategy: that the action depends only on current wealth (not past history), and that the action is deterministic. Both assumptions can be made without loss of generality.

This model, which is a certain kind of infinite-state Markov Decision Process (MDP), is a natural and elementary one to consider both from the point of view of probability theory, and that of mathematical finance. As far as we have been able to determine it has not previously been studied.

Before going into detail we pause for a simple illustration. Suppose two actions are available, called A and B; let q_i^{A} denote the probability of winning i dollars with action A:

$$\text{Action A:}\quad q_{-1}^{\mathrm{A}} = 0.5,\ \ q_{15}^{\mathrm{A}} = 0.5 \qquad\qquad \text{Action B:}\quad q_{-10}^{\mathrm{B}} = 0.5,\ \ q_{150}^{\mathrm{B}} = 0.5$$

Expected profit is ten times greater in action B, but it is easy to see that an investor with, say \$10, has probability of survival less than $1/2$ if he plays B, and close to 1 if he chooses and sticks to A. This illustrates how maximizing the likelihood of solvency can be quite different from maximizing expected profit. The problem, of course, is to determine proper strategy in less obvious situations.

1.1. Motivation.

There are a couple of reasons to focus on maximization of the likelihood of indefinite survival. The first concerns investment strategies of individual, "middle class" investors. Economic decision theory concerns itself largely (though not solely) with maximization of utility

as expressed by expected profit (or log profit). This framework may be appropriate to the decision theory of a shareholder-owned firm, whose bankruptcy creates an unpleasant but bounded effect on a balanced portfolio. But it is ill suited to the decision theory of an individual investor, whose goal is often not maximization of wealth for its own sake, but financial stability. For such a typical investor, bankruptcy, and its consequences for self and family, are dearly to be avoided.

The second reason concerns investment (loan) strategies of banks, which are unlike other corporations in that they are supposed to provide their depositors with a strong assurance of preservation of capital. The incompatibility between doing so and acting competitively in the loan marketplace has led to banking crises which have been addressed in part through government intervention including, in the USA, both federal deposit insurance and mandatory holding requirements. These restrict the extent to which banks can pursue purely profit-maximizing strategies (although we do not suggest that banks conversely act to maximize probability of indefinite solvency).

We return to the clash between optimizing for profit or survival. Naturally, a good way to avoid bankruptcy is to make a lot of money! But investment decisions entail a trade-off between risk and reward. The most secure investments typically provide returns below or only marginally above the inflation rate. So even a decision-maker whose sole purpose is to avoid bankruptcy cannot escape risk entirely, and must weigh the alternatives. The purpose of this paper is to develop some basic ingredients relevant to these decisions. In defining our model, simplicity is a key criterion. As a result, the model does not capture complications that always accompany realistic situations. On the other hand, this simplicity leads to clean mathematics and a basis from which more elaborate models can be considered.

1.2. **Related work.** As noted, the simple model defined above has apparently not been studied before. However, our motivation is very similar to that of previous authors, especially Ferguson [4], Truelove [11] and Browne [1]. The models are different enough to make the conclusions incomparable; some main differences are that in the previous work (a) The player has only one investment choice at any one time, and is simply deciding how much to invest, (b) That amount is unbounded except by the player's wealth. In some of the results, even the last restriction is dropped, and the player is permitted to borrow unlimited funds, sometimes with and sometimes without interest. Put simply, these authors' models are more general in allowing for investment scaling, and more special in not posing choices between dissimilar types of investments. The latter issue is the grist of our work.

An early book in the area, more relevant to Ferguson, Truelove and Browne's work than ours, is Dubins and Savage [3].

Slightly less related, but still relevant in terms of the motivation, is work in mathematical finance, in which risk (volatility) vs. reward is often measured with the Sharpe or Sterling ratios: see, e.g., [8]. Optimal investing by these criteria is less risk-averse than by ours.

Shifting attention from the finance aspect to the decision theory, our work is more closely related to the large literature on the MDP model [10], a broad formalization of the study of decision-making under stochastic uncertainty. Specifically, the "multi-arm bandit" problem concerns maximizing profit from a collection of actions, where optimal play is characterized by the well-known Gittins index [6, 12]. Our problem does not seem to fit into this model.

Another related (but independent) paper is that of Foster and Hart [5] in which they define the risk level of a gamble (if the gamble is additively invariant w.r.t. the player's wealth) as the minimum wealth at which the logarithm of the player's wealth has nonnegative drift.

1.3. **Results.** The specific questions we address include:

1. In a set $\mathcal{A}$ of actions, is there a *rich man's strategy* — an investment that is always the best choice once one's wealth is above some threshold? Put another way, does the optimal strategy have a "pure tail"?

Besides its obvious role in the decision theory of our model, this question gets at a real phenomenon which we feel should be reflected in any good model of risk-averse investing: that the poor do disproportionately worse than the rich because they can not afford to make certain investments that are by-and-large profitable, yet risky.

2. If there is a "rich man's strategy," what characterizes it, and is there a bound on the threshold where it takes over? If there isn't one, then what *does* the tail of the optimal strategy look like?

3. Can the optimal strategy be computed "efficiently"?

In Sections 3 and 4 we provide answers to these questions. We show that under certain technical conditions there does exist a rich man's strategy, and we provide a bound on where the pure tail begins. We also show that in general there is no such strategy — an interesting phenomenon, since it says that optimal play in a small-stakes game can depend, say, on the low-order bit of your bank balance. The MDP literature suggests three possible algorithms for computing the optimal strategy in the pure tail case (where this strategy has a finite description). For one of these algorithms, Value Iteration, we prove "linear convergence" (i.e., exponentially decreasing relative error) to the failure function of the optimal strategy.

1.4. **Notation and terminology.** An *action* is represented by a probability mass function on a finite set of integers. For an action A, let q_j^{A} be the probability that the payoff is j. For an action A, define $l_{\mathrm{A}} := -\min\{j < 0 : \ q_j^{\mathrm{A}} > 0\}$ and $r_{\mathrm{A}} := \max\{j > 0 : \ q_j^{\mathrm{A}} > 0\}$. The action is said to have *positive drift* if $\sum_{j=-l_{\mathrm{A}}}^{r_{\mathrm{A}}} j q_j^{\mathrm{A}} > 0$. The action is said to be *irreducible* if $\gcd\Big(\{j : q_j^{\mathrm{A}} > 0\}\Big) = 1$. In this paper all actions will be assumed to be irreducible and positive drift, though some of our statements hold more generally.

A *strategy* (sometimes also referred to as *policy* or *decision rule*) is a function $\pi : \mathbb{Z}_+ \to \mathcal{A}$, where $\mathcal{A}$ is a set of actions. For a strategy π, we define the following Markov chain. $X_{t+1} = X_t + Y_t$ where Y_t is defined as follows: If $X_t \le 0$ then $Y_t = 0$, whereas if $X_t > 0$ then Y_t is sampled according to $\pi(X_t)$, but otherwise independently of $X_0, \ldots, X_t$.

The *failure probability* at a positive integer w (i.e, the probability of ever going broke) corresponding to π is defined as $p^\pi(w) := \mathbf{Pr}[\exists_{m>0} : X_m \le 0 \mid X_1 = w]$.

A strategy is said to be *pure* if $\pi(w) = \pi(1)$ for all $w \ge 1$. It is said to have a *pure tail* if there is a $w' \ge 1$ such that $\pi(w) = \pi(w')$ for all $w \ge w'$.

1.5. **Structure of the paper.** In Section 2 we develop the fairly simple theory of the behavior of the game under a pure strategy. Being basically a random walk, our results are mostly known. However, those results serve as necessary tools for the study of optimal strategy. In Section 3 we prove the main results of the paper - namely conditions for the existence of a rich man's strategy. In Section 4 we discuss algorithms for determining the optimal strategy.

2. Pure strategies

Consider a pure strategy π^{A} consisting only of the action A with $l \equiv l_{\mathrm{A}}$ and $r \equiv r_{\mathrm{A}}$. Then, the failure probability $p(w) \equiv p^{\pi^{\mathrm{A}}}(w)$ satisfies the linear recurrence

$$p(w) = \sum_{j=-l}^{r} q_j^{\mathrm{A}} p(w+j), \quad w \ge 1, \tag{2.1}$$

where $q_j \equiv q_j^{\mathrm{A}}$ and with $p(w) = 1$ for all $w \le 0$. The *characteristic rational function* of A is defined as

$$q(z) \equiv q^{\mathrm{A}}(z) := -1 + \sum_{j=-l}^{r} q_j z^j \tag{2.2}$$

Lemma 2.1. *If $q'(1) > 0$ then q has exactly l roots in the open unit disk. Furthermore q has a unique positive root in the open unit disk.*

Note that the condition is equivalent to positive drift of the action.

Proof. It suffices to consider instead the roots of the polynomial

$$A(z) := z^l q(z) := \sum_{\substack{j=0 \\ j\neq l}}^{l+r} q_{j-l} z^j - [1 - q_0] z^l.$$

Note that $A'(1) = q'(1) > 0$ and $A(1) = 0$ so that $A(1-) < 0$. Furthermore $A(0) = q_{-l} > 0$ so by continuity, A has a root in $(0, 1)$, call it ζ. Also because of $A(0) \neq 0$, A has the same number of roots in any compact region as q. For $\epsilon > 0$ define $A_\epsilon(z) := f_\epsilon(z) + h(z)$, where

$$f_\epsilon(z) := -(1+\epsilon)(1-q_0)z^l, \qquad h(z) := \sum_{\substack{j=0 \\ j\neq l}}^{l+r} q_{j-l} z^j.$$

Consider the circle $|z| = \zeta$. There,

$$|f_\epsilon(z)| = (1+\epsilon)(1-q_0)\zeta^l > (1-q_0)\zeta^l \quad \text{and} \quad |h(z)| \leq \sum_{\substack{j=0 \\ j\neq l}}^{l+r} q_{j-l}\zeta^j.$$

Since ζ is a zero of A, we have $|f_\epsilon(z)| > |h(z)|$ for all z with $|z| = \zeta$. Hence by Rouché's theorem (see, e.g., [9]), f_ϵ and A_ϵ have the same number of zeros inside $|z| = \zeta$. But f_ϵ has exactly l zeros inside $|z| = \zeta$, and hence so does A_ϵ. Similarly A_ϵ has exactly l zeros inside $|z| = 1$, so that there are no zeros of A_ϵ in $\zeta < |z| < 1$. Now letting $\epsilon \downarrow 0$ yields that A has exactly l zeros in the closed disk $|z| \leq \zeta$ and none in the annulus $\zeta < |z| < 1$ so that the first claim of the lemma follows.

For the second claim note that if ζ_1 and ζ_2 are distinct positive zeros of A, an argument similar to the one above yields that there are no zeros of p in the interval (ζ_1, ζ_2). The claim then follows by letting $\zeta_1 = \zeta$ and $\zeta_2 = 1$. □

The positive root of q in the open unit disk will be called the *Perron root*, for reasons explained in Appendix A.

Remark 2.2. Since $q(1) = 0$, if $q'(1) < 0$ then there exists $z > 1$ such $q(z) < 0$. Also $q(z) > 0$ for large enough z. It follows then that q has a positive zero outside the closed unit disk and the proof of Lemma 2.1 reveals that this zero is unique.

Corollary 2.3. *If a pure strategy π^A has positive drift then its failure probabilities are*

$$p(w) \equiv p^{\pi^A}(w) = \sum_{j=1}^{d} \lambda_j^w \sum_{k=0}^{m_j - 1} c_{j,k} w^k, \tag{2.3}$$

where $\lambda_1, \ldots, \lambda_d$ are the distinct zeros of q in the interior of the unit disk in decreasing order of norm, with multiplicities $m_1, \ldots, m_d$ such that $m_1 + \cdots + m_d = l$, and $(c_{j,k})$ are constants.

Proof. Let λ be a zero of the characteristic rational function (2.2) with multiplicity m. Such a zero contributes a linear combination of $(w^j \lambda^w)_{j=0}^{m-1}$ to $p(w)$. Furthermore since we know a priori (see Fact 3.1) that $p(w) \to 0$ as $w \to \infty$, there cannot be any contribution from zeros with modulus at least 1. Since the pure strategy has positive drift, we have $(q^{\mathrm{A}})'(1) > 0$, so by Lemma 2.1, q^{A} has exactly l zeros in the unit disc and the result follows. □

Remark 2.4. Observe that the recurrence (2.1) defines a linear transformation mapping the initial conditions $p(w)_{w\le 0}$ monotonically to $p(w)_{w\ge 1}$. In particular, if λ is a zero of q, then $(\lambda^w)_{w\le 0}$ is mapped to $(\lambda^w)_{w\ge 1}$.

3. Optimal strategies

Let $\mathcal{A} = \{\mathrm{A}_1, \ldots, \mathrm{A}_k\}$ be a finite set of actions with positive drifts. We consider strategies $\pi : \mathbb{Z}_+ \to \mathcal{A}$. We start with a simple fact.

Fact 3.1. *For every strategy π, $p^\pi(w) \to 0$ as $w \to \infty$.*

Proof. For $j = 1, \ldots, k$, let $\{Y_n^{(j)}\}_{n=1}^\infty$ be i.i.d. samples of A_j, and assume that for different values of j, the sequences $\{Y_n^{(j)}\}$ are independent. The displacement at any time n is of the form $\sum_{j=1}^k \sum_{i=1}^{n_j} Y_i^{(j)}$, where the $\{n_j\}$ sum to n and are (arbitrarily dependent) random variables. Fix ϵ. Due to the positive drifts, for all N large enough,

$$\mathbf{Pr}\left[\forall_{n,j} \sum_{i=1}^{n} Y_i^{(j)} > -N/k\right] > 1 - \epsilon.$$

But this shows that for all N large enough, $p^\pi(w) < \epsilon$. □

For $w \ge 1$, an action A and a sequence p, we define

$$E_w^{\mathrm{A}}(p) := \sum_{j=-l_{\mathrm{A}}}^{r_{\mathrm{A}}} q_j^{\mathrm{A}} p(w+j). \tag{3.1}$$

For this to make sense, we need to have values for $p(w)$ for $k \le 0$. Unless otherwise mentioned, we take $p(w)$ to be 1 for all $w \le 0$. Similarly for a strategy π we define

$$E_w^{\pi}(p) := E_w^{\pi(w)}(p) \tag{3.2}$$

Clearly if p is the failure probability sequence of π, then

$$p(w) = E_w^{\pi}(p) \tag{3.3}$$

for every $w \ge 1$. Equation (3.3) determines p in the following sense:

Lemma 3.2. *Fix a strategy π and initial conditions $b(w)$, $w \le 0$. There exists a unique solution to (3.3) satisfying $p(w) = b(w)$ for all $w \le 0$ and $\lim_{w\to\infty} p(w) = 0$.*

Proof. This proof follows a conventional outline. Existence follows from Fact 3.1, Remark 2.4 and the fact that the probabilities satisfy (3.3). To see uniqueness, assume that p and q both satisfy the conditions. Then, $h = p - q$ also satisfies (3.3), $h(w) = 0$ for all $w \le 0$ and

$$\lim_{w\to\infty} h(w) = 0.$$

Assume that there exists w' such that $h(w') \ne 0$. Without loss of generality, $h(w') > 0$. Since $h(w) \to 0$, there exists w_0 so that $h(w) < h(w')$ for all $w > w_0$. Therefore, $\max_w h(w) = \max\{h(w) : w \le w_0\}$ and the maximum exists since it is taken over a finite set. Let H be this maximum, and let $\tilde{w} = \max\{w : h(w) = H\}$. By (3.3), $h(\tilde{w})$ is the average of numbers, all of which are no larger than H and some of which are strictly smaller than H. Therefore $h(\tilde{w}) < H$, in contradiction to its definition. Therefore, $h(w) \equiv 0$. □

Definition 3.3. We say that p is *harmonic* with respect to π if (3.3) holds for every $w \ge 1$. We say that p is *subharmonic* with respect to π if

$$p(w) \le E_w^{\pi}(p) \tag{3.4}$$

for every $w \geq 1$, and we say that p is *superharmonic* with respect to π if

$$p(w) \geq E_w^\pi(p) \tag{3.5}$$

for every $w \geq 1$.

The usefulness of Definition 3.3 is expressed in the following lemma:

Lemma 3.4. *Let π be a strategy and p the unique solution to (3.3) with given initial conditions $b(w)$, $w \leq 0$. Let v be a sequence that satisfies the following conditions:*

(1) $v(w) = b(w)$ for all $w \leq 0$.
(2) $\lim_{w\to\infty} v(w) = 0$.
(3) v is subharmonic with respect to π.

Then $v(w) \leq p(w)$ for every w. If instead v is superharmonic, then $v(w) \geq p(w)$ for every w.

Proof. The proof is similar to that of Lemma 3.2: Let $h = v - p$. Then h is subharmonic, $h(w) \to 0$ as $w \to \infty$, and $h(w) = 0$ for $w \leq 0$. Therefore, by the argument of Lemma 3.2, $\sup\{h(w) : w \geq 1\} = 0$, so that $v(w) \leq p(w)$. A similar proof applies for the superharmonic case by considering $\inf\{h(w) : w \geq 1\}$. □

3.1. **Structure of optimal strategies.** We can define a natural partial order between strategies: $\pi_1 \preceq \pi_2$ if for every w, $p^{\pi_1}(w) \leq p^{\pi_2}(w)$. We say that π^* is *optimal* if $\pi^* \preceq \pi$ for every strategy π. We say that σ is *locally optimal* if $\sigma \preceq \pi$ for every π satisfying $|\{w : \sigma(w) \neq \pi(w)\}| \leq 1$.

Proposition 3.5. *For every finite collection $\mathcal{A}$ of actions, there exists an optimal strategy. Furthermore, σ is optimal if and only if it is locally optimal.*

Proof. We will start with the "furthermore" part: Let σ be locally optimal, and let π be another strategy. Let s be the failure probability sequence for σ, and let p be the failure probability sequence for π. By local optimality of σ, for every w,

$$E_w^\pi(s) \geq s(w).$$

Therefore, s is subharmonic with respect to π, and by Lemma 3.4, $p(w) \geq s(w)$ for every w, i.e., $\sigma \preceq \pi$ and σ is optimal.

In order to prove the proposition, all we need is to find a locally optimal strategy. By compactness of the space of strategies (the product space of actions over all wealths), and continuity of

$$\sum_{w=1}^{\infty} p^\pi(w). \tag{3.6}$$

in this topology (using the positive-drifts assumption), there exists a strategy σ that minimizes expression 3.6. We claim that σ is locally optimal. Indeed, let π be so that $|\{w : \sigma(w) \neq \pi(w)\}| = 1$, and let w be the unique index such that $\sigma(w) \neq \pi(w)$. Since σ and π disagree at exactly one point, p^σ is either subharmonic or superharmonic with respect to π. It has to be subharmonic since p^σ minimizes (3.6), and therefore $\sigma \preceq \pi$ and σ is locally optimal. □

For an action A, let $\lambda_{\mathrm{A}}^{(1)} > 0, \lambda_{\mathrm{A}}^{(2)}, \ldots, \lambda_{\mathrm{A}}^{(l_{\mathrm{A}})}$ be the roots of its characteristic rational function [recall (2.2)] in the open unit disk arranged in decreasing order of modulus.

We now present a characterization of optimal strategies. The next theorem exhibits the existence of a "rich man's strategy," as indicated in the introductory section.

Theorem 3.6. *Let $\mathcal{A}$ be a finite set of actions and let $A \in \mathcal{A}$ be an action so that $\lambda_A^{(1)} < \lambda_B^{(1)}$ for every $B \neq A$ in $\mathcal{A}$. Let π^* be optimal for $\mathcal{A}$. Then there exists M such that $\pi^*(w) = A$ for every $w > M$.*

The existence of a "rich man's strategy" may seem natural, and if so, the imposition of technical hypotheses in Theorem 3.6 may seem disappointing. But this is not the case: strikingly, such conditions are necessary, as demonstrated in:

Theorem 3.7. *Let $\mathcal{A} = \{A, B\}$ with $l_A = l_B = 2$, $\lambda_A^{(1)} = \lambda_B^{(1)}$, and $\lambda_A^{(2)} \neq \lambda_B^{(2)}$. If π^* is optimal for $\mathcal{A}$, then for every W there exist $w', w'' > W$ such that $\pi^*(w') = A$ and $\pi^*(w'') = B$.*

Remark 3.8. Theorem 3.7 can be generalized to the case where l_g or l_f is greater than 2 under the assumption that the characteristic rational function of A has a root in the interior of the unit disk that is not shared by B and vice versa. The proof is omitted.

Proof of Theorem 3.6: For convenience of notation, let $\lambda := \lambda_{\mathrm{A}}^{(1)}$. Let π be a (fixed) strategy such that for every M there exists $w > M$ with $\pi(w) \neq \mathrm{A}$. We will show that π is not optimal. Let π^{A} be the pure-A strategy. Let $a(-w) = \lambda^{-w}$ and $p(-w) = 1$ for $w \geq 0$. Let a^{π} be the unique solution of $a(w) = E_w^{\pi}(a)$ with $a(w) \to 0$, and let $a^{\pi^{\mathrm{A}}}$ be the unique solution of $a(w) = E_w^{\pi^{\mathrm{A}}}(a)$ with $a(w) \to 0$. Let p^{π} and $p^{\pi^{\mathrm{A}}}$ be the failure probabilities for π and π^{A}.

It is sufficient to show that there exists w so that $p^{\pi^{\mathrm{A}}}(w) < p^{\pi}(w)$. Let l be the absolute value of the minimal number on the support of any of the actions in $\mathcal{A}$, i.e., $l := \max_{\mathrm{B} \in \mathcal{A}} l_{\mathrm{B}}$. Then by monotonicity (recall Remark 2.4), for every w,

$$a^{\pi}(w) \geq p^{\pi}(w) \geq \lambda^l a^{\pi}(w)$$

and

$$a^{\pi^{\mathrm{A}}}(w) \geq p^{\pi^{\mathrm{A}}}(w) \geq \lambda^l a^{\pi^{\mathrm{A}}}$$

Therefore it will suffice if we prove that there exist w so that

$$a^{\pi^{\mathrm{A}}}(w) < \lambda^l a^{\pi}(w). \tag{3.7}$$

In fact, we prove

$$\lim_{w \to \infty} \frac{a^{\pi}(w)}{a^{\pi^{\mathrm{A}}}(w)} = \infty. \tag{3.8}$$

To see (3.8), first note that (see Remark 2.4)

$$a^{\pi^{\mathrm{A}}}(w) = \lambda^w. \tag{3.9}$$

Next observe:

Lemma 3.9. *a^{π^A} is subharmonic with respect to π.*

Proof. Suppose not. Then there exists $w \geq 1$ such that

$$a^{\pi^{\mathrm{A}}}(w) > E_w^{\pi}(a^{\pi^{\mathrm{A}}}) = \sum_{j=-l_{\pi(w)}}^{r_{\pi(w)}} q_j^{\pi(w)} a^{\pi^{\mathrm{A}}}(w+j).$$

Clearly $\pi(w) \neq \mathrm{A}$ since λ is a root of the characteristic rational function of the pure-A strategy. Let $\pi(w) = \mathrm{B}$. Then, using (3.9),

$$\lambda^w > \sum_{j=-l_{\mathrm{B}}}^{r_{\mathrm{B}}} q_j^{\mathrm{B}} \lambda^{w+j},$$

so for $w' = w + k$,

$$a^{\pi_{\mathrm{A}}}(w') = \lambda^{w+k} > \sum_{j=-l_{\mathrm{B}}}^{r_{\mathrm{B}}} q_j^{\mathrm{B}} \lambda^{w+j+k} = \sum_{j=-l_{\mathrm{B}}}^{r_{\mathrm{B}}} q_j^{\mathrm{B}} a^{\pi^{\mathrm{A}}}(w'+j).$$

Thus $a^{\pi^{\mathrm{A}}}$ is superharmonic for the pure-B strategy. But this is a contradiction since λ is the unique smallest Perron root. □

Also note that if $\pi(w) \neq \mathrm{A}$, then we have strict subharmonicity at w. This fact will be used repeatedly in the rest of the proof.

Now, (3.8) will follow from Lemma 3.4 by constructing a sequence v so that

(1) $\lim_{w\to\infty} \frac{v(w)}{a^{\pi^{\mathrm{A}}}(w)} = \infty$,
(2) $v(w) = a^{\pi^{\mathrm{A}}}(w)$, $w \leq 0$,
(3) $v(w) \to 0$ as $w \to 0$, and
(4) v is subharmonic with respect to π.

We work iteratively: we define sequences $v^{(1)}, v^{(2)}, \ldots$ and take $v(w) = \lim_{k\to\infty} v^{(k)}(w)$.

We take $v^{(1)} = a^{\pi^{\mathrm{A}}}$ and $N_1 = 1$. For some positive integer χ and a constant $\rho > 0$, for every $k \geq 2$, there exists a positive integer N_k such that $v^{(k)}$ and N_k satisfy

(1) $v^{(k)}$ is subharmonic with respect to π.
(2) $N_k > N_{k-1}$, and
(3) $v^{(k)}(j) = v^{(k-1)}(j)$ for every $j < N_{k-1}$.
(4) $v^{(k)}(j) = (1+\rho)v^{(k-1)}(j)$ for every $j > N_{k-1} + \chi$.

If we find ρ, χ, $\{N_k\}$ and $\{\vec{p}^{(k)}\}$ satisfying (1)–(4) immediately above, we have proved the theorem.

Given $v^{(k)}$ we construct $v^{(k+1)}$. We construct sequences $b^{(1)}, \ldots, b^{(s)}$ with $b^{(1)} = v^{(k)}$ and $b^{(s)} = v^{(k+1)}$, with s to be described below. Let $n_1, \ldots, n_h$ be integers so that $|n_1| \geq |n_2| \cdots \geq |n_h|$, $q^{\mathrm{A}}_{n_i} > 0$ for every i and $n_1 + n_2 + \cdots + n_h = 1$. (These exist because of the assumption that A is irreducible.) Set

$$\chi := \max_{1\leq j\leq h} \left| \sum_{i=1}^{j} n_i \right|$$

Let N_k be the smallest integer larger than $N_{k-1} + \chi + l$ so that $\pi(N_k) \neq \mathrm{A}$. Take $b^{(1)} = v^{(k)}$, and

$$b^{(2)}(N_k) = E^{\pi}_{N_k}(v^{(k)}) > v^{(k)}(N_k) \quad ; \quad b^{(2)}(j) = b^{(1)}(j) \text{ for } j \neq N_k.$$

The increase above is due to strict subharmonicity of $v^{(k)}$ with respect to π where $\pi(w) \neq \mathrm{A}$. Then, for every $i = 1, \ldots, h$, we take $\alpha = N_k - (n_1 + \cdots + n_i)$ and

$$b^{(2+i)}(\alpha) = E^{\pi}_{\alpha}(b^{(2+i-1)}) > b^{(2+i-1)}(\alpha) \quad ; \quad b^{(2+i)}(j) = b^{(2+i-1)}(j) \text{ for } j \neq \alpha,$$

i.e., replace b_α with its $\pi(\alpha)$-average. (The argument that increase at each i occurs is omitted.) Continuing for every $u < l$ and $i = 1, \ldots, h$ we take

$$b^{(2+uh+i)}(\alpha+u) = E^{\pi}_{\alpha+u}(b^{(2+uh+i-1)}) > b^{(2+uh+i-1)}(\alpha+u)$$
$$b^{(2+uh+i)}(j) = b^{(2+uh+i-1)}(j) \text{ for } j \neq \alpha+u.$$

Let ρ be the infimal possible value among all k of

$$\min\left(\frac{b^{(s-1)}(N_k)}{v^{(k)}(N_k)}, \frac{b^{(s-1)}(N_k+1)}{v^{(k)}(N_k+1)}, \ldots, \frac{b^{(s-1)}(N_k+l)}{v^{(k)}(N_k+l)} \right) - 1. \tag{3.10}$$

Note that the value of the minimum in (3.10) is a function of the values taken by π on the interval $[N_k - \chi, N_k + \chi]$. Therefore, the infimum over all choices of k is in fact a minimum over the finite set of sequences of actions of length $2\chi + 1$. Therefore, $\rho > 0$. For $s - 1 = 2 + lh$, and we take

$$v^{(k+1)}(j) = b^{(s)}(j) = \begin{cases} b^{(s-1)}(j) & \text{if } j \leq N_k + l \\ \max\left\{ b^{(s-1)}(j), (1+\rho)v^{(k)}(j) \right\} & \text{if } j > N_k + l \end{cases}.$$

It is straightforward to check that (1)–(4) are satisfied. □

Proof of Theorem 3.7: For notational convenience, define $\lambda := \lambda^{(1)}_{\mathrm{A}} = \lambda^{(1)}_{\mathrm{B}}$, $\lambda_{\mathrm{A}} := \lambda^{(2)}_{\mathrm{A}}$, and $\lambda_{\mathrm{B}} := \lambda^{(2)}_{\mathrm{B}}$. Clearly λ_{A} and λ_{B} are real and negative. Without loss of generality, suppose that π^* has a pure-A tail. In light of Proposition 3.5, it is enough to show that π^* isn't locally optimal.

As usual, let p^{π^*} denote the failure probabilities corresponding to π^*. Since π^* has a pure-A tail, there exists w_0 such that for all $w > w_0$, we have $\pi^*(w) = \mathrm{A}$ and consequently,

$$p^{\pi^*}(w) = c_1\lambda^{w-w_0} + c_2\lambda_{\mathrm{A}}^{w-w_0}$$

with $c_1, c_2 \neq 0$. Choose $w_1 = w_0 + 3$. Then

$$E^{\mathrm{B}}_{w_1}(p^{\pi^*}) - p^{\pi^*}(w_1) = \sum_{j=-2}^{r_{\mathrm{B}}} q^{\mathrm{B}}_j[c_1\lambda^{w_1-w_0+j} + c_2\lambda_{\mathrm{A}}^{w_1-w_0+j}] - c_1\lambda^{w_1-w_0} - c_2\lambda_{\mathrm{A}}^{w_1-w_0} = c_2\lambda_{\mathrm{A}}^{w_1-w_0}q^{\mathrm{B}}(\lambda_{\mathrm{A}})$$

because λ is a root of q^{B}. Equivalently, for $w_2 = w_1 + 1$,

$$E^{\mathrm{B}}_{w_2}(p^{\pi^*}) - p^{\pi^*}(w_2) = c_2\lambda_{\mathrm{A}}^{w_2-w_0}q^{\mathrm{B}}(\lambda_{\mathrm{A}}) = \lambda_{\mathrm{A}}[E^{\mathrm{B}}_{w_1}(p^{\pi^*}) - p^{\pi^*}(w_1)]$$

Recall that $\lambda_{\mathrm{A}} < 0$, and therefore $E^{\mathrm{B}}_{w_1}(p^{\pi^*}) - p^{\pi^*}(w_1)$ and $E^{\mathrm{B}}_{w_2}(p^{\pi^*}) - p^{\pi^*}(w_2)$ are of opposite signs. Therefore, if we take $\pi^{(1)}$ to be π^* with the choice at w_1 changed to B and $\pi^{(2)}$ to be π^* with the choice at w_2 changed to B, then the sequence $p^{\pi*}$ is strictly superharmonic w.r.t. exactly one of $\pi^{(1)}$ and $\pi^{(2)}$. Therefore, π^* is not locally optimal. □

4. Algorithms for determining optimal strategies

We now turn our attention to the problem of determining the optimal strategy. To that end it will be useful to cast our problem in terms of *Markov decision processes* (MDPs). For background on MDPs, we refer the reader to the excellent book by Puterman [10].

Throughout, $\mathcal{A}$ is a finite set of *actions* with $l := \max\{l_{\mathrm{B}} : \mathrm{B} \in \mathcal{A}\}$ and $r := \max\{r_{\mathrm{B}} : \mathrm{B} \in \mathcal{A}\}$.

4.1. **MDP formulations.** For our purposes, a *Markov decision process* is a collection of objects $\{S, A_s, p(\cdot \mid s, a), r(s, a)\}$. Here S is a set of possible *states* the system can occupy. For each $s \in S$, the set of possible *actions* is denoted by A_s. The function $p(\cdot \mid s, a)$, called the *transition probability function* is a distribution on the set of states S and the *reward function* $r(s, a)$ is a real-valued function.

Under the assumptions of Theorem 3.6, we can modify our problem into an equivalent finite Markov decision problem, which makes determining an optimal strategy tractable. Let M be such that for some optimal strategy π^*, $\pi^*(w) = \mathrm{A}$ for every $w > M$. Here A is the action with the smallest Perron root. In Appendix B we show how to explicitly bound M, with a method that extends the arguments of Theorem 3.6. To find an optimal strategy we need only consider strategies that have a pure-A tail starting at M. Let $S = \{-l + 1, \ldots, M + r, \infty\}$. (The state ∞ represents the possibility of never returning to $\{-l + 1, \ldots, M + r\}$.) The actions for $s \in \{1, \ldots, M\}$ are the original actions of the system. For $s \in \{M + 1, \ldots, M + r\}$, the only action available is the action A' with the following transition probability function:

$$p(j \mid s, \mathrm{A}') = \alpha^{\mathrm{A}}_{s-j}; \quad j = s - 1, \ldots, s - l_{\mathrm{A}}; \quad p(\infty \mid s, \mathrm{A}') = 1 - \sum_{j=1}^{l_{\mathrm{A}}} \alpha^{\mathrm{A}}_j =: \alpha^{\mathrm{A}}_\infty.$$

Here the values $\{\alpha^{\mathrm{A}}_j\}$ are the coefficients of the linear functional giving p_w as a function of $p_{w-1}, \ldots, p_{w-l}$ in the pure A strategy; see Appendix A for further details. The action set for the state ∞ as well as for any state in $\{-l+1, \ldots, 0\}$, consists only of the trivial action that leaves the state unchanged. The reward function is given by (4.1):

$$r(s, a) := -\sum_{j \in S} \mathbf{1}\{s > 0 \text{ and } j < 0\}p(j \mid s, a), \quad s \in S;\ a \in A_s. \tag{4.1}$$

Clearly the expected total reward is the negative of the failure probability.

Next, we present an algorithm that can be used to determine optimal decision rules.

4.2. **Value iteration.** An iterative procedure known as value iteration produces a sequence that converges to the optimal expected total reward for each $s \in S$. The critical thing of course will be the runtime analysis.

(1) Set $v^0(s) = 0$ for each $s \in S$.
(2) For each $s \in S$, compute $v^{n+1}(s)$ using

$$v^{n+1}(s) = \max_{a \in A_s} \left\{ r(s,a) + \sum_{j \in S} p(j \mid s,a) v^n(j) \right\}$$

and increment n.

The sequences converge monotonically to the optimal expected total reward v^* [10]. Conceptually each iteration may be thought of as calculating the probabilities of ruin within that that many iterations of play.

We show next that the order of convergence of this algorithm is linear.

To that end, let d^* denote an optimal decision rule and consider the sequence defined iteratively by $u^0(s) = 0$ for each $s \in S$ and

$$u^{n+1}(s) = r(s, d^*(s)) + \sum_{j \in S} p(j \mid s, d^*(s)) u^n(j). \tag{4.2}$$

This is just the sequence produced by value iteration when the only action available at a state is the optimal action. Clearly $u^n(s) \to v^*(s)$ and a simple induction argument yields $v^n(s) \geq u^n(s)$ for each $s \in S$ and $n \geq 0$.

Writing (4.2) in matrix notation we have

$$u^{n+1} = P u^n + \alpha,$$

where $u^n, \alpha \in \mathbb{R}^{M+r}$ and $P \equiv P_{ij}$ is the $M + r \times M + r$ matrix with $P_{ij} := p(j \mid i, d^*(i))$.

Lemma 4.1. *Let $P \equiv P(d^*)$ denote the transition matrix for an optimal decision rule d^*. Then, $\rho(P)$, the spectral radius of P is strictly less than 1.*

Proof. If $\|P\|_\infty < 1$, then the claim is true. Suppose $\|P\|_\infty = 1$ so that $\rho(P) \leq 1$. Suppose $\rho(P) = 1$. Since P is nonnegative an eigenvalue of maximum modulus must be 1. Let $Px = x$, $x = [x_i] \neq 0$ and suppose p is an index such that $|x_p| = \|x\|_\infty \neq 0$. Now 1 lies on the boundary of $G(P)$, the Geršgorin region for the rows of P so that [7, Lemma 6.2.3(a)]

$$1 - P_{pp} = |1 - P_{pp}| = \sum_{\substack{j=1 \\ j \neq p}}^{M+r} P_{pj},$$

i.e., $\sum_{j=1}^{M+r} P_{pj} = 1$ so that $p \in \{1, \ldots, M\}$. Since P is the transition matrix for an optimal strategy there must be positive probability of reaching a state in $\{M+1, \ldots, M+r\}$ starting from the state p. In other words, there exist a sequence of distinct integers $k_1 = p, k_2, \ldots, k_m = q$ with $q \in \{M+1, \ldots, M+r\}$ such that all of the matrix entries $P_{k_1 k_2}, \ldots, P_{k_{m-1} k_m}$ are nonzero. But then [7, Lemma 6.2.3(b)], $|x_{k_i}| = |x_p|$ for each $i = 1, \ldots, m$. In particular $|x_q| = |x_p|$, so that again [7, Lemma 6.2.3(a)],

$$1 = |1 - P_{qq}| = \sum_{\substack{j=1 \\ j \neq q}}^{M+r} P_{qj} = \sum_{j=1}^{l} \alpha_j < 1,$$

which is a contradiction. □

Remark 4.2. Using the fact that all actions have positive drift, we can estimate the spectral radius as follows. Let D be the diagonal matrix with entries $(\lambda+\epsilon, (\lambda+\epsilon)^2, \ldots, (\lambda+\epsilon)^{M+r})$, where λ is the largest Perron root among all roots of the characteristic rational functions of the actions and $\epsilon > 0$ is arbitrarily small. We show that $\|\!|D^{-1}PD|\!\|_\infty \le \delta < 1$. Indeed for $i \in \{1, \ldots, M\}$, the ith row sum of $D^{-1}PD$ is given by

$$\sum_{j=1}^{M+r} P_{ij}(\lambda+\epsilon)^{j-i} \le q^i(\lambda+\epsilon) + 1 := \delta_i, \tag{4.3}$$

where $q^i(\cdot)$ is the characteristic function of the action employed at state i. If λ_i is the *unique* positive root of q^i inside the unit disk, then $q^i(\lambda_i) = q^i(1) = 0$ and q^i has no zero crossing in $(\lambda_i, 1)$. Since i has positive drift we have $(q^i)'(1) > 0$ so that $q^i(z) < 0$ for $z \in (\lambda_i, 1)$. Hence the row-sum in (4.3) is bounded by $\delta_i < 1$.

On the other hand for $i \in \{M+1, \ldots, M+r\}$, the ith row-sum of $D^{-1}PD$ is given by

$$\sum_{j=1}^{l} \alpha_j^{\mathrm{A}}(\lambda+\epsilon)^{-j} := \delta_i < 1,$$

the last strict inequality following from the fact that if $\lambda_{\mathrm{A}} < \lambda + \epsilon$ is the Perron root of the pure-A tail [recall (A.1)], then

$$\sum_{j=1}^{l} \alpha_j^{\mathrm{A}} \lambda_{\mathrm{A}}^{-j} = 1$$

Taking $\delta := \max_{1\le i\le M+r} \delta_i$ gives us $\rho(P) = \rho(D^{-1}PD) \le \|\!|D^{-1}PD|\!\|_\infty \le \delta < 1$.

The preceding lemma and remark lead directly to the following result.

Theorem 4.3. *Let v^* denote the optimal total expected value and v^n the nth iterate of value iteration. Then $v^n \ge u^n$, where for some vector norm $\|\cdot\|$ and $n \ge 1$,*

$$\|v^* - u^n\| \le c\|v^* - u^{n-1}\|$$

and $c < 1$ satisfies

$$c \le \max\{1 + \max_{\text{action } B} q^B(\lambda+\epsilon), \sum_{j=1}^{l} \alpha_j(\lambda+\epsilon)^{-j}\},$$

where λ is the largest of the Perron roots of the actions and $\epsilon > 0$ is arbitrarily small.

4.3. **Other algorithms.** In the MDP formulation, two other algorithms can be applied to computing the failure probabilities of the optimal strategy: policy iteration and linear programming. Their adaptation to our problem is discussed in Appendix C.

5. Discussion

It is clear that our results are at best a sketch of some elements of a larger theory. To begin with an equally well-motivated (and more general) model is one in which players are prohibited from taking actions that have nonzero probability of driving them immediately to a negative balance. (The player loses if no actions are available.) Our basic results carry over to this model. Another natural variant allows for the payoffs to be arbitrary real numbers. We have not explored this case.

It is important to note that since our results pertain purely to probability of ruin, rather than (as is common in finance) to expected returns, they apply equally to the additive or multiplicative perspective, and the latter is perhaps better motivated from the point of view of making investment decisions. That is to say, in the multiplicative perspective, each action takes the player's entire wealth, and randomly multiplies it by one of several possible factors. The ruin event occurs, say, when one's balance goes below the minimum dollar value of a trade in the market.

The multiplicative point of view captures the important feature that, due to inflation, there are (generally) no risk-free investments available.

In the additive perspective, it is of course natural to ask what happens if each available action can be scaled, at the player's discretion, by a positive constant. Allowing scaling by large constants is an interesting to study, although ultimately it seems very close to the multiplicative perspective. Allowing scaling by arbitrarily small constants trivializes the model: for any positive-drift action the probabilities of failure can be made to tend to 0. More importantly, it fails to match the motivating real-world scenarios. A bank deciding whether to issue a particular $\$200,000$ mortgage cannot change the associated risks by renaming it as $200,000$ separate \$1 mortgages. Ideally in this context one would like to address a common extension of our model and those treated by Ferguson [4], Truelove [11] and Browne [1].

NB: Most of our results date to 2004 and were presented at the 2004 AGATE Workshop on Algorithmic Game Theory in Bertinoro, Italy, and at a Plenary talk in RANDOM 2004.

References

[1] S. Browne. Optimal investment policies for a firm with a random risk process: exponential utility and minimizing the probability of ruin. *Math. Oper. Res.*, 20(4):937–958, 1995.

[2] C. Derman. *Finite state Markovian decision processes*. Mathematics in Science and Engineering, Vol. 67. Academic Press, New York, 1970.

[3] L. E. Dubins and L. J. Savage. *How to gamble if you must. Inequalities for stochastic processes*. McGraw-Hill Book Co., New York, 1965.

[4] T. S. Ferguson. Betting systems which minimize the probability of ruin. *J. Soc. Indust. Appl. Math.*, 13:795–818, 1965.

[5] D. P. Foster and S. Hart. An operational measure of riskiness. Unpublished manuscript 2008. Available at: http://www.ma.huji.ac.il/ hart/papers/risk.pdf.

[6] J. C. Gittins and D. M. Jones. A dynamic allocation index for the sequential design of experiments. In *Progress in statistics (European Meeting Statisticians, Budapest, 1972)*, pages 241–266. Colloq. Math. Soc. János Bolyai, Vol. 9. North-Holland, Amsterdam, 1974.

[7] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[8] M. Magdon-Ismail, A. F. Atiya, A. Pratap, and Y. S. Abu-Mostafa. On the maximum drawdown of a Brownian motion. *J. Appl. Probab.*, 41(1):147–161, 2004.

[9] A. I. Markushevich. *Theory of functions of a complex variable. Vol. I, II, III*. Chelsea Publishing Co., New York, english edition, 1977. Translated and edited by Richard A. Silverman.

[10] M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1994. , A Wiley-Interscience Publication.

[11] A. J. Truelove. Betting systems in favorable games. *Ann. Math. Statist.*, 41:551–556, 1970.

[12] J. N. Tsitsiklis. A short proof of the Gittins index theorem. *Ann. Appl. Probab.*, 4(1):194–199, 1994.

Appendix A. Alternative analysis of pure strategies

It is useful to have also the following alternate point of view on a pure strategy π^{A}. For each $j = 1, \ldots, l$, let $\alpha_j \equiv \alpha_j^{\mathrm{A}}$ denote the probability that the corresponding Markov chain ever enters a state to the left of w, and that the first such state is $w - j$. These probabilities do not depend on w (since π^{A} is pure). Then the failure probabilities for π^{A} are given by

$$p(w) = \sum_{j=1}^{l} \alpha_j p(w-j), \qquad n \geq 1, \tag{A.1}$$

with $p(w) = 1$, for $w \leq 0$. The *companion matrix* associated with the action A is the $l \times l$ matrix

$$\mathbf{M} \equiv \mathbf{M}_{\mathrm{A}} := \begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ \alpha_l & \alpha_{l-1} & & \cdots & \alpha_1 \end{bmatrix},$$

all other entries in the matrix being 0. Since the characteristic polynomial of $\mathbf{M}$ is the characteristic polynomial of the recurrence (A.1), the l eigenvalues of the companion matrix $\mathbf{M}$ are precisely the l roots $\lambda_1, \ldots, \lambda_l$ of (2.2) in the open unit disk.

Suppose that we further assume that A is irreducible. Then we have $\alpha_j > 0$ for $j = 1, \ldots, l$. Thus the directed graph of $\mathbf{M}$ is strongly connected which implies [7, Theorem 6.2.24] that $\mathbf{M}$ is irreducible. It follows from the nonnegativity of $\mathbf{M}$ [7, Theorem 8.4.4(d)] that λ_1 is an algebraically simple eigenvalue and furthermore since $\alpha_1 > 0$, we have that λ_1 is the unique eigenvalue of maximum modulus [7, Corollary 8.4.8]. The failure probabilities of the strategy π^{A} are then $\Theta(\lambda_1^w)$. The eigenvalue λ_1 is often called the *Perron root*.

Appendix B. Bound on M (beginning of the pure tail)

Theorem 3.6 shows that all optimal strategies have a pure-A tail, i.e, there exists M such that for any optimal strategy π^*, we have $\pi^*(w) = \mathrm{A}$ for all $w > M$. Clearly $M \leq M_1 M_2$ where M_1 is an upper bound on the number of non-A actions in an optimal strategy and M_2 is an upper bound on the distance between two consecutive non-A actions in an optimal strategy. We omit the details of the following claims. Let $\lambda_1 > 0, \lambda_2, \ldots, \lambda_{l_{\mathrm{A}}}$ be the roots of the characteristic rational function of A listed in decreasing order of magnitude.

(1) Examination of the proof of Theorem 3.6 reveals the following bound on M_1. Let $v_1, v_2, \ldots, v_h$ be elements of the support of A whose sum is 1. Define

$$\chi := \max \left| \sum_{i=1}^{j} v_i \right|$$

$$\rho := \left(\prod_{i=1}^{j} q_{v_i}^{\mathrm{A}} \right)^{l} \times \min_{\mathrm{B} \neq \mathrm{A}} q^{\mathrm{B}}(\lambda_1)$$

Then,

$$M_1 \leq (\chi + l) \frac{\log(1/\lambda)^l}{\log(1+\rho)}.$$

(2) Define

$$\epsilon := \frac{1}{2} \lambda_1^{l+r} \min_{\mathrm{B} \neq \mathrm{A}} q^{\mathrm{B}}(\lambda_1),$$

$$K := (\lambda_1 \lambda_2 \cdot \cdots \cdot \lambda_l)^{-1} \lambda_1^{l^2}.$$

Take

$$n_1 = \frac{\log(2K/\epsilon)}{\log \lambda_1 - \log \lambda_2} \text{ and } n_2 = \frac{\log(2n_1)/\epsilon)}{\log(1/\lambda_1)}.$$

Then $M_2 \leq n_1 + n_2$.

Appendix C. Policy iteration and linear programming

C.1. **Policy iteration.** The policy iteration algorithm for the finite formulations is the following.

(1) Set $n = 0$ and select an arbitrary decision rule d_0.
(2) [Policy evaluation] Compute the expected total reward $\{v^n(s)\}_{s\in S}$ for the rule d_{n+1} by solving the linear system of equations:

$$v(s) = r(s, d_n(s)) + \sum_{j\in S} p(j \mid s, d_n(s))v(j), \quad s \in S.$$

(3) [Policy improvement] For each $s \in S$, choose $d_{n+1}(s)$ such that

$$d_{n+1}(s) \in \arg\max_{a\in A_s} \left\{ r(s,a) + \sum_{j\in S} p(j \mid s, a)v^n(j) \right\},$$

choosing $d_{n+1}(s) = d_n(s)$ whenever possible.
(4) If $d_{n+1}(s) = d_n(s)$ for all $s \in S$, then stop, setting d^* to d_n. Otherwise increment n.

Since local optimality implies optimality (Proposition 3.5), clearly policy iteration produces an optimal strategy in a finite number of iterations.

C.2. **Linear programming.** For the finite positive bounded formulation the following linear program determines the optimal expected total reward [10, §7.2.7]. Here $(\beta_j)_{j\in S}$ denote positive scalars that sum to 1.

$$\text{minimize} \sum_{j\in S} \beta_j v(j)$$

subject to

$$v(s) - \sum_{j\in S} p(j \mid s, a)v(j) \geq r(s,a); \qquad a \in A_s, s \in S,$$
$$v(s) \geq 0; \qquad s \in S.$$

To determine an optimal decision rule one considers the dual:

$$\text{maximize} \sum_{s\in S}\sum_{a\in A_s} r(s,a)x(s,a)$$

subject to

$$\sum_{a\in A_j} x(j,a) - \sum_{s\in S}\sum_{a\in A_s} p(j \mid s, a)x(s,a) \leq \beta_j; \qquad j \in S,$$
$$x(s,a) \geq 0; \qquad a \in A_s, s \in S.$$

Given an optimal basic feasible solution x^* to the dual an optimal decision rule can be determined as

$$d^*(s) = \begin{cases} a & \text{if } x^*(s,a) > 0 \text{ and } s \in S^* \\ \text{arbitrary} & \text{otherwise.} \end{cases}$$

Here $S^* := \{s \in S^* : \sum_{a\in A_s} x^*(s,a) > 0\}$. That d^* is well-defined follows from Theorem 7.2.18 of [10].

In general a negative MDP need not afford a solution by linear programming [10, §7.3.6]. However our formulation can be seen to be an instance of a *first-passage problem* [2]. (Briefly, one seeks to maximize the reward until a state in $\{-l+1, \ldots, 0\}$ is reached.) We aggregate the states $-l+1, \ldots, 0$ into another state $\mathbf{0}$. The transition probabilities and reward function are updated

to reflect this aggregation. The following linear program will then compute the optimal expected total cost.

$$\text{minimize} \sum_{j \in S \setminus \{\mathbf{0}\}} \beta_j v(j)$$

subject to

$$v(s) \geq r(s,a) + \sum_{j \in S \setminus \{\mathbf{0}\}} p(j \mid s,a) v(j); \qquad a \in A_s, s \in S \setminus \{\mathbf{0}\},$$

where $(\beta_j)_{j \in S \setminus \{\mathbf{0}\}}$ are positive scalars that sum to 1. An optimal decision rule can be determined again from the optimal basic feasible solution of the dual

$$\text{maximize} \sum_{s \in S \setminus \{\mathbf{0}\}} \sum_{a \in A_s} r(s,a) x(s,a)$$

subject to

$$\sum_{a \in A_j} x(j,a) - \sum_{s \in S \setminus \{\mathbf{0}\}} \sum_{a \in A_s} p(j \mid s,a) x(s,a) = \beta_j; \qquad j \in S \setminus \{\mathbf{0}\}$$

$$x(s,a) \geq 0; \qquad a \in A_s, s \in S \setminus \{\mathbf{0}\}.$$

(Noam Berger) Department of Mathematics, Hebrew University of Jerusalem, Israel

(Nevin Kapur, Leonard J. Schulman) Department of Computer Science, California Institute of Technology

(Vijay V. Vazirani) College of Computing, Georgia Institute of Technology
E-mail address, Noam Berger: berger@math.ucla.edu
E-mail address, Nevin Kapur, Leonard J. Schulman: {kapur,schulman}@caltech.edu
E-mail address, Vijay V. Vazirani: vazirani@cc.gatech.edu