

Solvency Games

N. Berger¹, N. Kapur², L. J. Schulman², V. V. Vazirani³

¹Department of Mathematics
Hebrew University of Jerusalem, Israel
berger@math.ucla.edu

²Department of Computer Science
California Institute of Technology
{kapur, schulman}@caltech.edu

³College of Computing
Georgia Institute of Technology
vazirani@cc.gatech.edu

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 $\pi : \mathbb{Z}_+ \rightarrow \mathcal{A}$ specifying the action that is chosen at each possible value of wealth. Corresponding to strategy π , define

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

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:

© Berger, Kapur, Schulman and Vazirani; licensed under Creative Commons License-NC-ND

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:

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

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 [2]. 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.

Perhaps closest to our work is an interesting paper that first appeared in June 2007 [5]*. In this paper, Foster and Hart consider the question of measuring the *riskiness* of a gamble (their "gamble" has the same definition as our "action"). Surprisingly enough, they show that this can be boiled down to a single number – the *critical wealth*. If the wealth is strictly smaller than this number, it is risky to play the action, i.e., it will lead to bankruptcy with probability 1. Else, playing this action is guaranteed to not lead to bankruptcy, again with probability 1.

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

*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.

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, terminology and structure of the paper

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 $l_A := -\min\{j < 0 : q_j^A > 0\}$ and $r_A := \max\{j > 0 : q_j^A > 0\}$. The action is said to have *positive drift* if $\sum_{j=-l_A}^{r_A} j q_j^A > 0$. The action is said to be *irreducible* if $\gcd(\{j : q_j^A > 0\}) = 1$. 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 $\pi : \mathbb{Z}_+ \rightarrow \mathcal{A}$, 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 $X_t \leq 0$ 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, \dots, X_t . The *failure probability* at a positive integer w (i.e, the probability of ever going broke) corresponding to π is defined as $p^\pi(w) := \Pr[\exists m > 0 : X_m \leq 0 \mid X_1 = w]$. A strategy is said to be *pure* if $\pi(w) = \pi(1)$ for all $w \geq 1$. It is said to have a *pure tail* if there is a $w' \geq 1$ such that $\pi(w) = \pi(w')$ for all $w \geq w'$.

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. Missing proofs and the Appendix can be found in the full paper, available online at ECCG [1].

2 Pure strategies

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

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

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

$$q(z) \equiv q^A(z) := -1 + \sum_{j=-l}^r q_j z^j \quad (2)$$

LEMMA 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 . For $\epsilon > 0$ define $A_\epsilon(z) := f_\epsilon(z) + h(z)$, where

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

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 p , 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 $|z| = 1$, 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 p , 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 $\zeta_2 = 1$.

Remarks: The positive root of q in the open unit disk will be called the *Perron root*, for reasons explained in Appendix A in [1]. 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 1 reveals that this zero is unique.

Corollary 2 *If a pure strategy π^A has positive drift then its failure probabilities are*

$$p(w) \equiv p^{\pi^A}(w) = \sum_{j=1}^d \lambda_j^w \sum_{k=0}^{m_j-1} c_{j,k} w^k, \quad (3)$$

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

PROOF. Let λ be a zero of the characteristic rational function (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) that $p(w) \rightarrow 0$ as $w \rightarrow \infty$, there cannot be any contribution from zeros with modulus at least 1. Since the pure strategy has positive drift, we have $(q^A)'(1) > 0$, so by Lemma 1, q^A has exactly l zeros in the unit disc and the result follows.

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

3 Optimal strategies

Let $\mathcal{A} = \{A_1, \dots, A_k\}$ be a finite set of actions with positive drifts. We consider strategies $\pi : \mathbb{Z}_+ \rightarrow \mathcal{A}$. We start with a simple fact.

Fact 3 For every strategy π , $p^\pi(w) \rightarrow 0$ as $w \rightarrow \infty$.

PROOF. For $j = 1, \dots, k$, 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 $\sum_{j=1}^k \sum_{i=1}^{m_j} Y_i^{(j)}$, 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,

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

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

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

$$E_w^A(p) := \sum_{j=-l_A}^{r_A} q_j^A p(w+j). \quad (4)$$

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

$$E_w^\pi(p) := E_w^{\pi(w)}(p) \quad (5)$$

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

$$p(w) = E_w^\pi(p) \quad (6)$$

for every $w \geq 1$. Equation (6) determines p in the following sense:

LEMMA 4. Fix a strategy π and initial conditions $b(w)$, $w \leq 0$. There exists a unique solution to (6) satisfying $p(w) = b(w)$ for all $w \leq 0$ and $\lim_{w \rightarrow \infty} p(w) = 0$.

PROOF. This proof follows a conventional outline. Existence follows from Fact 3, Remark 2 and the fact that the probabilities satisfy (6). To see uniqueness, assume that p and q both satisfy the conditions. Then, $h = p - q$ also satisfies (6), $h(w) = 0$ for all $w \leq 0$

and $\lim_{w \rightarrow \infty} h(w) = 0$. Assume that there exists w' such that $h(w') \neq 0$. Without loss of generality, $h(w') > 0$. Since $h(w) \rightarrow 0$, there exists w_0 so that $h(w) < h(w')$ for all $w > w_0$. Therefore, $\max_w h(w) = \max\{h(w) : w \leq w_0\}$ 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 (6), $h(\tilde{w})$ 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 5 We say that p is harmonic with respect to π if (6) holds for every $w \geq 1$. We say that p is subharmonic with respect to π if

$$p(w) \leq E_w^\pi(p) \quad (7)$$

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

$$p(w) \geq E_w^\pi(p) \quad \text{for every } w \geq 1. \quad (8)$$

The usefulness of Definition 5 is expressed in the following lemma:

LEMMA 6. Let π be a strategy and p the unique solution to (6) 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. $\lim_{w \rightarrow \infty} v(w) = 0$.
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 .

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 7 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 6, $p(w) \geq s(w)$ 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

$$\sum_{w=1}^{\infty} p^\pi(w). \quad (9)$$

in this topology (using the positive-drifts assumption), there exists a strategy σ that minimizes expression 9. 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 (9), and therefore $\sigma \preceq \pi$ and σ is locally optimal.

For an action A , let $\lambda_A^{(1)} > 0, \lambda_A^{(2)}, \dots, \lambda_A^{(l_A)}$ be the roots of its characteristic rational function [recall (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 8. *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 8 may seem disappointing. But this is not the case: strikingly, such conditions are necessary, as demonstrated in:

THEOREM 9. *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 $\pi^*(w'') = B$.*

Remarks: Theorem 9 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 8: For convenience of notation, let $\lambda := \lambda_A^{(1)}$. Let π be a (fixed) strategy such that for every M there exists $w > M$ with $\pi(w) \neq A$. 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 $a(w) \rightarrow 0$, and let a^{π^A} be the unique solution of $a(w) = E_w^{\pi^A}(a)$ with $a(w) \rightarrow 0$. Let p^π and p^{π^A} be the failure probabilities for π and π^A .

It is sufficient to show that there exists w so that $p^{\pi^A}(w) < p^\pi(w)$. Let l be the absolute value of the minimal number on the support of any of the actions in \mathcal{A} , i.e., $l := \max_{B \in \mathcal{A}} l_B$. Then by monotonicity (recall Remark 2), for every w ,

$$\begin{aligned} a^\pi(w) &\geq p^\pi(w) \geq \lambda^l a^\pi(w) \\ a^{\pi^A}(w) &\geq p^{\pi^A}(w) \geq \lambda^l a^{\pi^A}(w) \end{aligned}$$

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

$$a^{\pi^A}(w) < \lambda^l a^\pi(w). \quad (10)$$

In fact, we prove

$$\lim_{w \rightarrow \infty} \frac{a^\pi(w)}{a^{\pi^A}(w)} = \infty. \quad (11)$$

To see (11), first note that (see Remark 2)

$$a^{\pi^A}(w) = \lambda^w. \quad (12)$$

LEMMA 10. *a^{π^A} is subharmonic with respect to π .*

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 $l := \max\{l_B : B \in \mathcal{A}\}$ and $r := \max\{r_B : B \in \mathcal{A}\}$.

4.1 MDP formulations

For our purposes, a *Markov decision process* is a collection of objects $\{S, A_s, p(\cdot | s, a), r(s, a)\}$. 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 $p(\cdot | s, a)$, 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 8, 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^* , $\pi^*(w) = A$ for every $w > M$. Here A is the action with the smallest Perron root. In Appendix B [1] we show how to explicitly bound M , with a method that extends the arguments of Theorem 8. To find an optimal strategy we need only consider strategies that have a pure- A tail starting at M . Let $S = \{-l + 1, \dots, M + r, \infty\}$. (The state ∞ represents the possibility of never returning to $\{-l + 1, \dots, M + r\}$.) The actions for $s \in \{1, \dots, M\}$ are the original actions of the system. For $s \in \{M + 1, \dots, M + r\}$, the only action available is the action A' with the following transition probability function:

$$p(j | s, A') = \alpha_{s-j}^A; \quad j = s - 1, \dots, s - l_A; \quad p(\infty | s, A') = 1 - \sum_{j=1}^{l_A} \alpha_j^A =: \alpha_\infty^A.$$

Here the values $\{\alpha_j^A\}$ are the coefficients of the linear functional giving p_w as a function of p_{w-1}, \dots, p_{w-l} in the pure A strategy; see Appendix A [1] for further details. The action set for the state ∞ as well as for any state in $\{-l + 1, \dots, 0\}$, consists only of the trivial action that leaves the state unchanged. The reward function is given by (13):

$$r(s, a) := - \sum_{j \in S} \mathbf{1}\{s > 0 \text{ and } j < 0\} p(j | s, a), \quad s \in S; a \in A_s. \quad (13)$$

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

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

and increment n .

The sequences converge monotonically to the optimal expected total reward v^* [10]. We show next that the order of convergence 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

$$u^{n+1}(s) = r(s, d^*(s)) + \sum_{j \in S} p(j | s, d^*(s)) u^n(j). \quad (14)$$

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

Writing (14) in matrix notation we have $u^{n+1} = Pu^n + \alpha$, where $u^n, \alpha \in \mathbb{R}^{M+r}$ and $P \equiv P_{ij}$ is the $M+r \times M+r$ matrix with $P_{ij} := p(j | i, d^*(i))$.

LEMMA 11. *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 $p \in \{1, \dots, M\}$. Since P is the transition matrix for an optimal strategy there must be positive probability of reaching a state in $\{M+1, \dots, M+r\}$ starting from the state p . In other words, there exist a sequence of distinct integers $k_1 = p, k_2, \dots, k_m = q$ with $q \in \{M+1, \dots, M+r\}$ such that all of the matrix entries $P_{k_1 k_2}, \dots, P_{k_{m-1} k_m}$ are nonzero. But then [7, Lemma 6.2.3(b)], $|x_{k_i}| = |x_p|$ for each $i = 1, \dots, m$. In particular $|x_q| = |x_p|$, so that again [7, Lemma 6.2.3(a)],

$$1 = |1 - P_{qq}| = \sum_{\substack{j=1 \\ j \neq q}}^{M+r} P_{qj} = \sum_{j=1}^l \alpha_j < 1, \quad \text{which is a contradiction.}$$

Remarks: 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 $(\lambda + \epsilon, (\lambda + \epsilon)^2, \dots, (\lambda + \epsilon)^{M+r})$, 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 $\|D^{-1}PD\|_\infty \leq \delta < 1$. Indeed for $i \in \{1, \dots, M\}$, the i th row sum of $D^{-1}PD$ is given by

$$\sum_{j=1}^{M+r} P_{ij} (\lambda + \epsilon)^{j-i} \leq q^i (\lambda + \epsilon) + 1 := \delta_i, \quad (15)$$

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 (15) is bounded by $\delta_i < 1$.

On the other hand for $i \in \{M+1, \dots, M+r\}$, the i th row-sum of $D^{-1}PD$ is given by $\sum_{j=1}^l \alpha_j^A (\lambda + \epsilon)^{-j} := \delta_i < 1$, the last strict inequality following from the fact that if $\lambda_A < \lambda + \epsilon$ is the Perron root of the pure-A tail, then

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

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

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

THEOREM 12. *Let v^* denote the optimal total expected value and v^n the n th iterate of value iteration. Then $v^n \geq u^n$, where for some vector norm $\|\cdot\|$ and $n \geq 1$,*

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

and $c < 1$ satisfies

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

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

5 Discussion

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 in [1].

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 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 variant to study. (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 [2].

References

- [1] N. Berger, N. Kapur, L. J. Schulman, and V. V. Vazirani. Solvency games. Available at ECCC (Electronic Colloquium on Computational Complexity) Report TR08-089, 2008.
- [2] 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.
- [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.