The Gambler's Problem and Beyond

Anonymous authors
Paper under double-blind review

Abstract

We analyze the Gambler's problem, a simple reinforcement learning problem where the gambler has the chance to double or lose their bets until the target is reached. This is an early example introduced in the reinforcement learning textbook by Sutton & Barto (2018), where they mention an interesting pattern of the optimal value function with high-frequency components and repeating non-smooth points but without further investigation. We provide the exact formula for the optimal value function for both the discrete and the continuous case. Though simple as it might seem, the value function is pathological: fractal, self-similar, non-smooth on any interval, zero derivative almost everywhere, and not written as elementary functions. Sharing these properties with the Cantor function, it holds a complexity that has been uncharted thus far. With the analysis, our work could lead insights on improving value function approximation, Q-learning, and gradient-based algorithms in real applications and implementations.

1 Introduction

We analytically investigate a deceptively simple problem, the Gambler's problem, introduced in the reinforcement learning textbook by Sutton & Barto (2018), on Example 4.3, Chapter 4 page 84 and is described as below. The problem presents a natural and simple setting, which would hide its attractiveness. A close inspection will show that the problem, as an representative of the entire family of Markov decision process (MDP), involves a level of complexity and curiosity uncharted in years of reinforcement learning research.

The problem discusses a gambler's casino game, where they conducts multiple rounds of betting. The gambler doubles up the bet if they wins a round or loses the bet if they loses the round. The game ends when either the gambler reaches their goal of N or running out of money. On each round, the gambler must decide what portion of the capital to stake. In the discrete setting this bet must be an integer but it can also be a real number in the continuous setting. To formulate it as an MDP, let state s be the current capital and action a the amount of bet. The reward is zero on all transitions but $+1$ on $s = N$. Let $p \geq 0.5$ be the probability that the gambler loses a round of bet.

The state-value function then gives the probability of winning from each state. Our goal is to solve the optimal value function of the problem. As a preliminary, the state-value function of an MDP with respect to policy π is defined as

$$f^\pi(s) = \mathbb{E}\left[\sum_t \gamma^t r_t \middle| s_0 = s, a_t \sim \pi(s_t), s_{t+1} \sim \mathcal{T}(s_t, a_t), r_t \sim r(s,a)\right],$$

where $\pi(\cdot)$, s_t, a_t, r_t, $\mathcal{T}$, and γ are the policy, state, action, reward, transition kernel, and the discount factor, respectively. The state-value function and the action-state value function can induce each other so we will focus on the former for the rest of the discussion. Let π^* be one of the optimal policies, $f^{\pi^*}(s)$ is then the optimal state-value function. Note this optimal value function is unique by (Sutton & Barto, 2018), despite the possible existence of multiple optimal policies.

In this paper, we first give the solution to the discrete Gambler's problem. Denote N as the target capital, n as the starting capital (n denotes the state in the discrete setting), $p \geq 0.5$ as the probability of losing a bet, and γ as the discount factor. The special case of $N = 100$, $\gamma = 1$ corresponds to the original setting in Sutton and Barto's book.

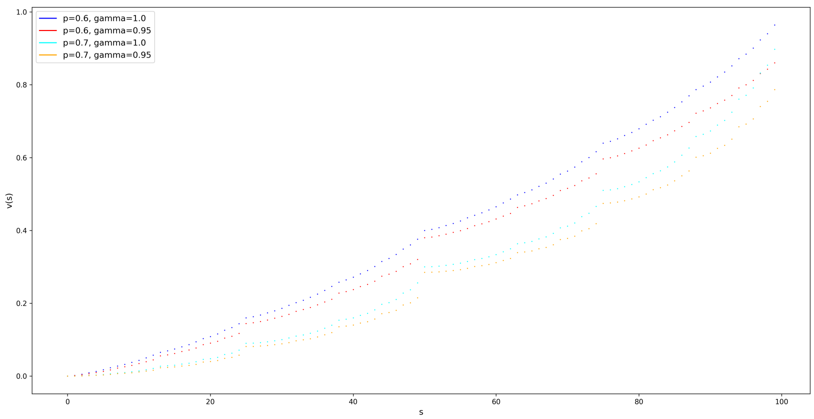

Figure 1: The optimal state-value function of the discrete Gambler's problem.

Proposition 1. *Let $0 \leq \gamma \leq 1$ and $p > 0.5$. The optimal value function $b(n)$ is $v(n/N)$ in the discrete setting of the Gambler's problem, where $v(\cdot)$ is the optimal value function under the continuous case defined in Theorem 11.*

The above statement is depending on our main theorem, which states the solution of the more general, continuous setting of the problem. In the continuous setting the target capital is 1, the state space is $[0, 1]$, and the action space is $0 < a \leq \min(s, 1-s)]$ at state s, meaning that the bet can be any fraction of the current capital as long as the capital after winning does not exceed 1:

Theorem 11. *Let $0 \leq \gamma \leq 1$ and $p > 0.5$. Under the continuous setting of the Gambler's problem, the optimal value function is $v(1) = 1$ and*

$$v(s) = \sum_{i=1}^{\infty} (1-p)\gamma^i b_i \prod_{j=1}^{i-1} ((1-p) + (2p-1)b_j) \tag{1}$$

on $0 \leq s < 1$, where $s = 0.b_1 b_2 \ldots b_l \ldots_{(2)}$ is the binary representation of the state s.

Next, we solve the Bellman equation of the continuous gambler's problem. In the strictly discounted setting $0 \leq \gamma < 1$, the solution of the Bellman equation $f(s) = \max_{0 < a \leq \min(s, 1-s)} (1-p)\gamma f(s + a) + p\gamma f(s-a)$, $f(0) = 0$, $f(1) = 1$ is uniquely $f(s) = v(s)$ the optimal value function.

This uniqueness does not hold in general. If the rewards are not discounted, the solution of the Bellman equation is either the value function, or a constant function larger than 1:

Theorem 18. *Let $\gamma = 1$ and $p > 0.5$. The solution of the Bellman equation $f(s) = \max_{0 < a \leq \min(s, 1-s)} (1-p) f(s+a) + p f(s-a)$, $f(0) = 0$, $f(1) = 1$, is either of*

- *$f(s)$ is $v(s)$ defined in Theorem 11, or*
- *$f(0) = 0$, $f(1) = 1$, and $f(s) = C$ for all $0 < s < 1$, for some constant $C \geq 1$.*

Under the corner case of $\gamma = 1$, $p = 0.5$ (where the gambler do not lose capital in bets in expectation), the problem involves midpoint concavity (Sierpiński, 1920a;b) and Cauchy's functional equation. The measurable function that solves the Bellman equation is uniquely $f(s) = C's + B'$ on $s \in (0, 1)$, for some constants $C' + B' \geq 1$. Additionally, Under Axiom of Choice, $f(s)$ can also be some non-constructive, non Lebesgue measurable function described by the Hamel basis.

Though the description of the Gambler's problem seems natural and simple, Theorem 11 shows that its simpleness is deceptive. The optimal value function presents its self-similar, fractal and non-rectifiable form, which cannot be described by any simple analytic formula. At any level of zooming-in, the value function keeps showing the same texture as itself. This can be observed in

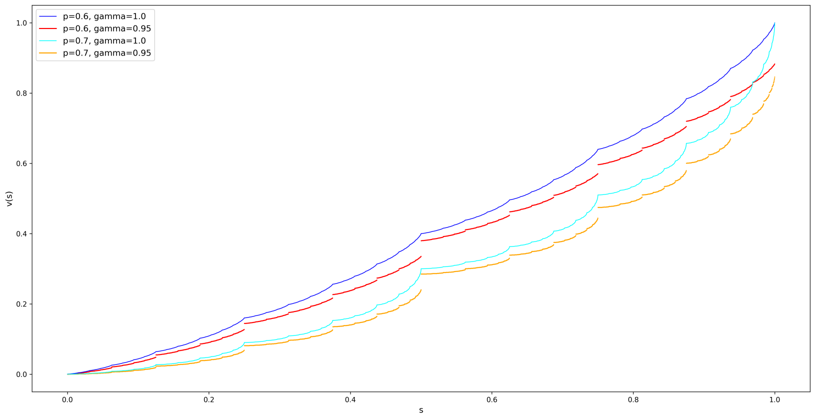

Figure 2: The optimal state-value function of the continuous Gambler's problem.

Figure 1 and 2. With the fractal nature, the value function does not possess many of the desired properties for algorithms and analysis. Namely, the function is not continuous under $\gamma < 1$; not differentiable on the dyadic rational, where any point on the dyadic rational has a left-derivative of zero and a right derivative of infinity; no local linear or Taylor expansion; cannot be approximated efficiently in polynomial many bins to the error. These properties are not desired and not expected by the recent line of reinforcement learning studies, who commonly use a neural network approximate the value function. These properties are likely to be extended to a wider range of MDPs, consider the simplicity of the Gambler's problem and the similar fractal patterns observed empirically in other reinforcement learning tasks.

Intuitive description of $v(s)$. All the statements above requires the definition of $v(s)$. In fact, in this paper, $v(s)$ is important enough such that its definition will not change with the context. The function cannot be written as a combination of the elementary functions. Nevertheless, we give a intuitive way to understand the function. The function can be regarded as generated by the following iterative process: First we fix $v(0) = 0$ and $v(1) = 1$, and have

$$v(\frac{1}{2}) = \gamma(pv(0) + (1-p)v(1)) = (1-p)\gamma.$$

Here, $v(\frac{1}{2})$ is γ times the weighted average of the two "neighbors" $v(0)$ and $v(1)$ that have been already evaluated. Further, the same operation applies to $v(\frac{1}{4})$ and $v(\frac{3}{4})$, where

$$v(\frac{1}{4}) = \gamma(pv(0) + (1-p)v(\frac{1}{2})) = (1-p)^2\gamma^2,$$
$$v(\frac{3}{4}) = \gamma(pv(\frac{1}{2}) + (1-p)v(1) = (1-p)\gamma + p(1-p)\gamma^2.$$

Repeatedly, we have $v(\frac{1}{8}) = (1-p)^3\gamma^3$, $v(\frac{3}{8}) = (1-p)^2\gamma^2 + p(1-p)^2\gamma^3$, $v(\frac{5}{8}) = p(1-p)\gamma^2 + (1-p)^2\gamma^2 + p(1-p)^2\gamma^3$, and $v(\frac{7}{8}) = (1-p)\gamma + p(1-p)\gamma^2 + p^2(1-p)\gamma^3$, and so forth. This process will gives the evaluation of $v(s)$ on the dense and compact dyadic rational $\bigcup_{\ell \geq 1} G_\ell$, where $G_\ell = \{k2^{-\ell} \mid k \in \{1, \dots, 2^\ell - 1\}\}$. With the fact that $v(s)$ is monotonically strictly increasing, this dyadic rationals determines the function $v(s)$ uniquely.

It can also be explained from the analytical definition of $v(s)$ this iterative process. Starting with the first bit, a bit of 0 will not change the value, while a bit of 1 will add $(1-p)\gamma^i \prod_{j=1}^{i-1}((1-p) + (2p-1)b_j)$ to the value. This term can also be written as $(1-p)\gamma^i((1-p)^{\#\mathbf{0}\text{ bits}} + p^{\#\mathbf{1}\text{ bits}})$, where the number of bits is counted over all previous bits. The value $(1-p)^{\#\mathbf{0}\text{ bits}} + p^{\#\mathbf{1}\text{ bits}}$ decides the gap between two neighbor existing points in the above process, when we insert a new point in the middle. This insertion corresponds to the iteration on G_ℓ over ℓ

Illustrative description of $v(s)$. We provide high resolution plots of $b(n)$ and $v(s)$ in Figure 1 and Figure 2, respectively. The non-smoothness and the self-similar fractal patterns can be clearly observed from the figures. In principle, these two function cannot be completely illustrated as their non-smooth patterns continue indefinitely when we zoom in the figure. We have though tried to draw them at a fine enough grain where the human vision does not distinguish the context. Both the figures are by dot-plot, where the dots in then second figure is extreme dense so as it looks like a curve.

As observed in the figure, $v(s)$ is continuous when $\gamma = 1$ while $v(s)$ is not continuous on infinity many points when $\gamma < 1$. In fact, when $\gamma < 1$, the function is discontinuous on the dyadic rationals $\bigcup_{\ell \geq 1} G_\ell$ while continuous on its complement, as we will rigorously show later.

Self similarity. The function on $[0, \frac{1}{2}]$ and on $[\frac{1}{2}, 1]$ is similar with the function itself on $[0, 1]$. This similarity repeats to $[0, \frac{1}{4}]$, $[\frac{1}{4}, \frac{1}{2}]$ and so forth. The set of fractal functions have a higher level of complexity of the elementary functions. It can lead to chaos as well. Functions described by an combination of elementary functions on $\mathbb{R}$ has a dimension of 1. But the plotted curve of $v(s)$ has a dimension of 1.64, according to our simulation of the box counting method.

Optimal policies. It is immediate by Theorem 11 and its lemmas that $\pi(s) = \min(s, 1-s)$ is one of the Blackwell optimal policies. Here Blackwell optimal is defined as the uniform optimality under any $0 \leq \gamma \leq 1$. This agrees with the intuition that under a game that is in favor of the casino ($p > 0.5$), the gambler desire to bet the maximum to finish the game in as few rounds as possible. This Blackwell optimality is not unique, for example, $\pi(\frac{15}{32}) = \frac{1}{32}$ is also optimal for any γ. Under $\gamma = 1$ and when s can be written in finite many bits $s = b_1 ... b_{\ell(2)}$ in binary (assume $b_\ell = 1$), $\pi(s) = 2^{-l}$ is also an optimal policy. This policy is by repeatedly rounding the capital to carryover the bits, keeping the game within at most ℓ rounds of bets.

Implications. Our results indicates hardness on reinforcement learning. The hardness on value function approximation: as the value function can fall on the set of fractal functions, it will not be possible to approximate the function with a piece-wise constant function (discretization) or a Lipschitz-continuous function (including a neural network) by an ϵ accuracy with poly(ϵ) complexity. The hardness on derivative: The value function's derivative cannot be estimated properly, as $v(s)$ has a derivative of 0 almost everywhere, except on G_ℓ, where it has a left derivative of infinity and a right derivative of 0. Algorithms relying on $\frac{\partial v(s)}{\partial s}$ and $\frac{\partial Q(s,a)}{\partial a}$ (Lillicrap et al., 2015; Gu et al., 2017) can suffer from the error estimation, where $Q(s, a)$ is the action-state-value function. In practice the boolean implementation of float numbers can further increase this error, as all points evaluated are on G_ℓ. The hardness on Q-learning (Mnih et al., 2015; Watkins & Dayan, 1992; Baird, 1995): When $\gamma = 1$, the solution to the Bellman equation is not necessarily the value function. A large constant function can also be a solution, who may have a even small Bellman error than the optimal value function. This challenges Q-learning, whose goal is to find a solution of the Bellman equation and then treats the solution as the value function. Though the artificial γ is originally introduced to prevent the return from diverging, it can be necessary to prevent the algorithm from converging to a large constant in Q-learning.

2 Discrete Case

The analysis of the discrete case of The Gambler's problem will give an exact solution. It will also explain the reason the plot on the book has a strange pattern of repeating spurious points.

The discrete case can be described by the following MDP: The state space is $\{0, \ldots, N\}$; the action space at n is $\mathcal{A}(n) = \{0 < a \leq \min(n, N-n)\}$; the transition from state n and action a is $n-a$ and $n+a$ with probability p and $1-p$, respectively; the reward function is $r(N) = 1$ and $r(n) = 0$ for $0 \leq n \leq N-1$; The MDP terminates at $n \in \{0, N\}$; We use a time-discount factor of $0 \leq \gamma \leq 1$, where the agent receives $\gamma^T r(N)$ rewards if the agents reaches the state $n = N$ at time T.

Let $b(n)$, $n \in \mathbb{N}, 0 \leq n \leq N$, be the value function. The exact solution below of the discrete case is relying on Theorem 11, our main theorem which describes the exact solution of the continuous case. This theorem will be discussed and proved later in Section 4.1.

Proposition 1. *Let $0 \leq \gamma \leq 1$ and $p > 0.5$. The optimal value function $b(n)$ is $v(n/N)$ in the discrete setting of the Gambler's problem, where $v(\cdot)$ is the optimal value function under the continuous case defined in Theorem 11.*

Proposition 1 indicates the discretization of the problems yields the discrete, exact evaluation of the continuous value function at $0, 1/N, \ldots, 1$. If we omit the learning error, the plots on the book and by the open source implementation (Zhang, 2019) are the evaluation of the fractal $v(s)$ at $0, 1/N, \ldots, 1$. This explains the strange appearance of curve in the figure.

3 SETTING

We formulate the continuous Gambler's problem as a Markov decision process (MDP) with $\mathcal{S} = [0, 1]$ and $\mathcal{A}(s) = (0, \min(s, 1-s)]$, $s \in (0, 1)$ to be the state space and the action space, respectively. Here $s \in \mathcal{S}$ represents the capital the gambler currently possesses and the action $a \in \mathcal{A}(s)$ denotes the amount of bet. Without loss of generality we have assumed that the bet amount should be less or equal to $1 - s$ to avoid the total capital to be more than 1. The consecutive state s' transits to $s - a$ and $s + a$ with probability $p \geq 0.5$ and $1 - p$ respectively. The process terminates if $s \in \{0, 1\}$ and the agent receives an episodic reward $r = s$ at the terminal state. Let $0 \leq \gamma \leq 1$ be the discount factor and $f(\cdot)$ be the value function.

From the Bellman equation of the above described MDP, the properties for $f(\cdot)$ are

$$f(s) = \max_{a \in \mathcal{A}(s)} p\gamma\, f(s-a) + (1-p)\gamma\, f(s+a) \text{ for any } s \in (0, 1), \tag{\textbf{A}}$$

and

$$f(0) = 0, \;\; f(1) = 1. \tag{\textbf{B}}$$

It can be shown (in Lemma 2 and Lemma 3 later) that a function satisfying (**AB**) must be lower bounded by 0. A reasonable upper bound is 1, as the value function is the probability of the gambler eventually reaching the target, which must be between 0 and 1. It is also reasonable to assume the continuity of the value function at least at $s = 0$, Otherwise an arbitrary small amount of will have a fixed probability of reaching the target 1. The bounded version of the problem leads to the optimal value function:

$$0 \leq \gamma \leq 1, \; p > 0.5, \; f(s) \leq 1 \text{ for all } s, \; f(s) \text{ is continuous on } s = 0. \tag{\textbf{X}}$$

Respectively, the unbounded version of the problem leads to the solutions of the Bellman equation:

$$0 \leq \gamma \leq 1, \; p > 0.5. \tag{\textbf{Y}}$$

The results hold for $p = 0.5$ as well, except an extreme corner case of $\gamma = 1$, $p = 0.5$, where the monotonicity in Lemma 3 will not apply. This case (**Z**) involves arguments over measurability and the assumption of Axiom of Choice, which we will discuss in the end of Section 4:

$$\gamma = 1, \; p = 0.5, \; f(s) \text{ is unbounded.} \tag{\textbf{Z}}$$

We are mostly interested in two settings: the first setting (**ABX**) and its solution Theorem 11, describes a set of necessary conditions of $f(s)$ being the optimal value function of the gamblers problem. As we show later the solution of (**ABX**) is unique, this solution must be the value function. The second setting (**ABY**) and its solution Proposition 17 and Theorem 18, describes all the functions that satisfy the Bellman equation. These functions are the optimal points that value iteration and Q-learning algorithms may converge to. (**ABY**) is discussed in Theorem 23.

4 ANALYSIS

4.1 ANALYSIS OF THE GAMBLER'S PROBLEM

In this section we show that $v(s)$ defined below is a unique solution of the system (**ABX**). Since the optimal state-value function must satisfies the system (**ABX**), $v(s)$ is the optimal state-value function of the Gambler's problem. This statement is rigorously proved in **Theorem 11**.

Let $0 \leq \gamma \leq 1$ and $p > 0.5$. We define

$$v(s) = \sum_{i=1}^{\infty} (1-p)\gamma^i b_i \prod_{j=1}^{i-1} ((1-p) + (2p-1)b_j) \quad (1)$$

for $0 \leq s < 1$, where $s = 0.b_1 b_2 \dots b_l \dots_{(2)}$ is the binary representation of s. It is obvious that the series converge for any $0 \leq s < 1$.

The notation $v(s)$ will always refer to the definition above in this paper and will not change with the context. We show later that this $v(s)$ is the optimal value function of the problem. We use the notation $f(s)$ to denote a general solution of a system, which varies according to the required properties.

Let the dyadic rationals

$$G_\ell = \{k2^{-\ell} \mid k \in \{1, \dots, 2^\ell - 1\}\} \quad (2)$$

such that G_ℓ is the set of numbers that can be represented by at most ℓ binary bits. The general idea to verify the Bellman equation is to prove

$$v(s) = \max_{a \in G_\ell \cap \mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a) \text{ for any } s \in G_\ell$$

by induction of $\ell = 1, 2, \dots$, and generalize this optimality to the entire interval $s \in (0, 1)$. Then we show the uniqueness of $v(s)$ that solves the system (**ABX**). For presentation purposes, the uniqueness is discussed first, in Lemma 2, though it is depending on other lemmas.

As an overview, Lemma 2, 3, and 4 describe the system (**ABX**). Lemma 5, 7, and 8 describe the properties of $v(s)$. All the proofs are deferred to the appendix. Among them Lemma 2 carries the main idea leading to the theorem.

Lemma 2 (Uniqueness under existence)**.** *If $v(s)$ and $f(s)$ both satisfy (**ABX**), then $v(s) = f(s)$ for all $0 \leq s \leq 1$.*

Lemma 3 (Monotonicity)**.** *Let $\gamma = 1$ and $p > 0.5$. If $f(\cdot)$ satisfies (**AB**) then $f(\cdot)$ is monotonically increasing on $[0, 1)$.*

Lemma 4 (Continuity)**.** *Let $\gamma = 1$ and $p \geq 0.5$. If $f(s)$ is monotonically increasing on $(0, 1]$ and it satisfies (**AB**), then $f(s)$ is continuous on $(0, 1]$.*

When $f(s)$ is only required to be monotonically increasing on $(0, 1)$, the continuity still holds but only on $(0, 1)$.

Lemma 5. *Let $\ell \geq 1$. For any $s \in G_\ell$,*

$$\max_{a \in (G_{\ell+1} \setminus G_\ell) \cap \mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a) < \max_{a \in G_\ell \cap \mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a).$$

The arguments in the proof that either $N_b + k \geq N_c + k' + 1$ or $N_c + k' \geq N_b + k$ must hold is tight for integers N_b and N_c. This is the case for $a \in G_{\ell+1} \setminus G_\ell$. When $a \notin G_{\ell+1}$, this sufficient condition becomes even looser. This makes G_ℓ to be the only set of possible optimal actions, given $s \in G_\ell$.

Corollary 6. *Let $\ell \geq 1$. For any $s \in G_\ell$,*

$$\arg\max_{a \in \mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a) \subseteq G_\ell.$$

Now we verify the Bellman property on $\bigcup_{\ell \geq 1} G_\ell$.

Lemma 7. *Let $\ell \geq 1$. For any $s \in G_{\ell+1}$,*

$$\min(s, 1-s) \in \arg\max_{a \in G_{\ell+1} \cap \mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a).$$

Lemma 8. *Both $v(s)$ and $v'(s) = \max_{a \in \mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a)$ are continuous at s if there does not exist an ℓ such that $s \in G_\ell$.*

The continuity of $v(s)$ extends to the dyadic rationals $\bigcup_{\ell \geq 1} G_\ell$ when $\gamma = 1$, which means that $v(s)$ is *continuous everywhere* on $[0, 1]$ under $\gamma = 1$. It worth note that similar to the Cantor function, $v(s)$ is *not absolutely continuous*. In fact, $v(s)$ shares more common properties with the Cantor function, as they both have *a derivative of zero almost everywhere*, while having their value goes from 0 to 1, and their range is every value in between of 0 and 1.

The continuity of $v'(s) = \max_{a \in \mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a)$ indicates the optimal action to the uniquely $\min(s, 1-s)$ on $s \notin G_\ell$. This optimal action agrees with the optimal action we specified in Lemma 7, which makes $\pi(s) = \min(s, 1-s)$ an optimal policy for every state.

Corollary 9. *If $s \notin G_\ell$ for any $\ell \geq 1$,*

$$\arg\max_{a \in \mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a) = \{\min(s, 1-s)\}.$$

Lemma 10. *$v(s)$ is the unique solution of the system (**ABX**).*

Theorem 11. *Let $0 \leq \gamma \leq 1$ and $p > 0.5$. Under the continuous setting of the Gambler's problem, the optimal value function is $v(1) = 1$ and $v(s) = \sum_{i=1}^{\infty} (1-p)\gamma^i b_i \prod_{j=1}^{i-1} ((1-p) + (2p-1)b_j)$ on $0 \leq s < 1$, where $s = 0.b_1 b_2 \dots b_l \dots_{(2)}$ is the binary representation of the state s.*

Proof. As the optimal state-value function must satisfy the system (**ABX**) and $v(s)$ is the unique solution to the system, $v(s)$ is the optimal value function. ☐

Lemma 7 and Corollary 9 together induce one of the optimal deterministic policies as below. As the arguments hold uniformly for any $0 \leq \gamma \leq 1$, this optimality is also Blackwell optimal.

Corollary 12. *The policy $\pi(s) = \min(s, 1-s)$ is Blackwell optimal, meaning it is optimal under any γ.*

It is notably that when $\gamma = 1$ and $s \in G_\ell \setminus G_{\ell-1}$ for some ℓ, then $\pi'(s) = 2^{-\ell}$ is also an optimal policy at s.

Lemma 7 and Theorem 11 also induce the following corollary that the optimal value function $v(s)$ is fractal and self-similar.

Corollary 13. *The curve of the value function $v(s)$ on the interval $[k2^{-\ell}, (k+1)2^{-\ell}]$ is similar (in geometry) to the curve of $v(s)$ itself on $[0, 1]$, for any integer $\ell \geq 0$ and $0 \leq k \leq 2^\ell - 1$.*

Some other notable facts about $v(s)$ are as below:

Fact 14. *The expectation*

$$\int_0^1 v(s) ds = (1-p)\gamma = v(\frac{1}{2}).$$

Fact 15. *The derivative*

$$\lim_{\Delta s \to 0^+} \frac{v(s + \Delta s)}{\Delta s} = 0, \quad \lim_{\Delta s \to 0^-} \frac{v(s + \Delta s)}{\Delta s} = \begin{cases} +\infty, & \text{if } s = 0 \text{ or } s \in \bigcup_{\ell \geq 1} G_\ell, \\ 0, & \text{otherwise}. \end{cases} \tag{3}$$

Fact 16.

$$\arg\min_{0 \leq s \leq 1} v(s) - s = \{\frac{2}{3}\}.$$

4.2 Analysis of the Bellman Equation

We have proved that $v(s)$ is the optimal value function in Theorem 11, by showing the uniqueness of the solution of the system (**ABX**). However, the bounds (**X**) is derived from the context of the Gambler's problem by hand. It is rigorous enough to prove the optimal value function, but we are also interested in the solutions purely derived by the MDP setting. Also, algorithmic approaches such as Q-learning (Watkins & Dayan, 1992; Baird, 1995; Mnih et al., 2015) solves the MDP by finding the solution of the Bellman equation (**AB**), without eliciting the context of the problem. The solution will be treated as the optimal value function without further arguments. In this section, we will inspect the system of Bellman equation (**AB**) of the Gambler's problem. We first discuss a more general case (**ABY**) where $p > 0.5$.

The value function $v(s)$ is obviously still a solution of the system (**ABY**) without the $f(s) \leq 1$ condition. The natural question is if there exist any other solutions. The answer is two-fold: When $\gamma < 1$, $f(s) = v(s)$ is unique. However, when $\gamma = 1$, the solution is either $v(s)$ or a constant function at least 1. This indicates that algorithms like Q-learning has constant functions as their set of converging points. As $v(s)$ itself is hard to approximate due to the non-smoothness, a constant function in fact induces a smaller approximation error and thus has a better optimality for Q-learning with function approximation.

It is immediate to generate this result to general MDPs, as function of a large constant solves MDPs with episodic rewards. This indicates that Q-learning may have more than one converging points and may diverge from the optimal value function under $\gamma = 1$. This leads to the need of γ, which is artificially introduced and biases the learning objective. More generally, the Bellman equation may have a continuum of finite solutions in an infinite state space, even with $\gamma < 1$. Some studies exist on the necessary and sufficient conditions for a solution of the Bellman equation to be the value function (Kamihigashi & Le Van, 2015; Latham, 2008; Harmon & Baird III, 1996). Though, the majority of this topic remain open.

The following proposition show that when the discount factor is strictly less than 1, the solution toward the Bellman equation is uniquely the value function.

Proposition 17. *When $\gamma < 1$, $v(s)$ is the unique solution of the system (**ABY**).*

Proof. The uniqueness has been shown in Lemma 2 for the system (**ABY**). When $\gamma < 1$ it corresponds to case (II), where the upper bound $f(s) \leq 1$ in condition (**X**) is not used. Therefore Lemma 2 holds for (**ABY**) under $\gamma < 1$, so follows Lemma 10 the uniqueness as desired. □

This uniqueness no longer holds under $\gamma = 1$.

Theorem 18. *Let $\gamma = 1$ and $p > 0.5$. A function $f(s)$ satisfies (**ABY**) if and only if either*

- *$f(s)$ is $v(s)$ defined in Theorem 11, or*
- *$f(0) = 0$, $f(1) = 1$, and $f(s) = C$ for all $s \in (0, 1)$, for some constant $C \geq 1$.*

The fact that a large constant function can also be a solution to the Bellman equation can be extended to wide range of MDP settings. The below proposition list one of the sufficient conditions but even without this condition it holds in practice most likely.

Proposition 19. *For an arbitrary MDP with episodic rewards where every state has an action to transit to a non-terminal state almost surely, $f(s) = C$ for all non-terminal states s is a solution of the Bellman equation system for any C greater or equal to the maximum one-step reward.*

Proof. The statement is immediate by verifying the Bellman equation. □

The rest of the section discusses the Gambler's problem under $p = 0.5$. In this case, the optimal value function is still $v(s)$ by the same proof of Theorem 11. Proposition 17 also holds so $v(s)$ is the only solution given $\gamma < 1$. When $\gamma = 1$, Theorem 11 still holds. Interestingly, when $\gamma = 1$ and $p = 0.5$, $v(s) = s$. This agrees with the intuition that the gambler does not lose its capital by placing bets in expectation, therefore the optimal value function should be linear to s. The problem that remains is the solution to the Bellman equation, under $\gamma = 1$ and $p = 0.5$. This corresponds to the system (**ABZ**).

When $p = 0.5$, condition (**A**) indicates midpoint concavity

$$f(s) \geq \frac{1}{2} f(s - a) + \frac{1}{2} f(s + a), \tag{4}$$

where the equality must holds for some $a \in \mathcal{A}(s)$. As Lemma 3 no longer holds, a solution $f(s)$ may have negative value for some s. Though if it does not have a negative value, it is not hard to show that the function must be linear. By condition (**A**) we need $f(s) \geq s$ for any s. Therefore the solution is $f(0) = 0$, $f(1) = 1$, and $f(s) = C's + B'$ on $0 < s < 1$ for some constants $C' + B' \geq 1$.

If $f(s)$ does have a negative value on some s, then the midpoint concavity does not imply concavity. By recursively applying Equation 4 we see that the set $\{(s, f(s)) \mid s \in [0,1]\}$ is dense and compact on $[0,1] \times \mathbb{R}$. The function becomes pathological, if it exists. Despite this, the following lemma shows that $f(s)$ needs to be positive on the rationals $\mathbb{Q}$.

Lemma 20. *Let $f(s)$ satisfies (**ABZ**). If there exists $0 \leq s^- < s^+ \leq 1$ and a constant C such that $f(s^-), f(s^+) \geq C$, then $f(s) \geq C$ for all $s \in \{s^- + q(s^+ - s^-) \mid q \in \mathbb{Q}, 0 \leq q \leq 1\}$.*

Lemma 20 agrees with the intuition that midpoint concavity indicates rational concavity. The below statements then gives some insight on the irrational points.

Lemma 21. *Let $f(s)$ satisfies (**ABZ**). If there exists an $\bar{s} \in \mathbb{R} \setminus \mathbb{Q}$ such that $f(\bar{s}) \geq 0$, then $f(s) \geq 0$ for all $s \in \{q\bar{s} + r \mid q, r \in \mathbb{Q}, 0 \leq q, r, \leq 1, q + r \leq 1\}$.*

Corollary 22. *Let $f(s)$ satisfies (**ABZ**). If there exists an $\bar{s} \in \mathbb{R} \setminus \mathbb{Q}$ such that $f(\bar{s}) < 0$, then $f(q\bar{s})$ is monotonically decreasing with respect to q for $q \in \mathbb{Q}$, $1 \leq q < 1/\bar{s}$.*

Lemma 21 and Corollary 22 indicate that when there exists a negative or positive value, infinity many other points (that are not necessarily in its neighbor) must be negative or positive as well. It is sufficient to observe the complexity of the problem with these statements. In fact, it is shown that $f(s)$ is not Lebesgue measurable and non-constructive (Sierpiński, 1920b), just by being midpoint concave but not concave.

Such an $f(s)$ exists if and only if we assume Axiom of Choice (Sierpiński, 1920b;a). With the axiom we consider the field extension $\mathbb{R}/\mathbb{Q}$ and specify a set of basis $\mathbb{B} = \{b_i\}_{i \in \mathcal{I}}$, known as the Hamel bases. With this basis $\mathbb{B}$ every real number can be written uniquely as a combination of the elements in the $\mathbb{B} \cup \{0\}$ with rational coefficients. Now denote every real number s as a unique vector $(q, q_i)_{i \in \mathcal{I}}$ such that $s = q + \sum_{i \in \mathcal{I}} q_i b_i$.

One of the solution can be shown by defining $r(s) = q, s \in \mathbb{R}$, where q is the rational component in the Hamel basis representation. As per there is only one rational number in the basis, the function $r(s)$ is additive, namely,

$$r(s_1) + r(s_2) = r(s_1 + s_2).$$

Let $\beta(s)$, $s \in \mathbb{R}$ be an arbitrary concave function. It is immediate to verify that

$$f(s) = \beta(s - r(s)) + r(s) \tag{5}$$

is a solution for the system (**ABZ**).

More generally, $f(s)$ is any function in the form $\beta(q, \{q_i\}_{i \neq i_0}) + z(q_{i_0})$, where $\beta(\cdot)$ is rational concave and $z(\cdot)$ is linear and it satisfies the boundary conditions.

Theorem 23. *Let $\gamma = 1$ and $p = 0.5$. A function $f(s)$ satisfies (**ABZ**) if and only if either*

- *$f(s) = C's + B'$ on $s \in (0,1)$, for some constants $C' + B' \geq 1$, or*
- *$f(s)$ is some non-constructive, non Lebesgue measurable function under Axiom of Choice.*

REFERENCES

Leemon Baird. Residual algorithms: Reinforcement learning with function approximation. In *Machine Learning Proceedings 1995*, pp. 30–37. Elsevier, 1995.

Shixiang Shane Gu, Timothy Lillicrap, Richard E Turner, Zoubin Ghahramani, Bernhard Schölkopf, and Sergey Levine. Interpolated policy gradient: Merging on-policy and off-policy gradient estimation for deep reinforcement learning. In *Advances in neural information processing systems*, pp. 3846–3855, 2017.

Mance E Harmon and Leemon C Baird III. Spurious solutions to the bellman equation. 1996.

Takashi Kamihigashi and Cuong Le Van. Necessary and sufficient conditions for a solution of the bellman equation to be the value function: A general principle. *Documents de travail du Centre dÉconomie de la Sorbonne 2015.07*, 2015. ISSN: 1955-611X.

Peter Latham. Bellman's equation has a unique solution. 2008.

Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

Wacław Sierpiński. Sur l'équation fonctionnelle f (x+ y)= f (x)+ f (y). *Fundamenta Mathematicae*, 1(1):116–122, 1920a.

Wacław Sierpiński. Sur les fonctions convexes mesurables. *Fundamenta Mathematicae*, 1(1):125–128, 1920b.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

Shangtong Zhang. Python implementation of Reinforcement learning: An introduction, 2019. URL `https://github.com/ShangtongZhang/reinforcement-learning-an-introduction`.

A Proofs

Proposition 1. *Let $0 \leq \gamma \leq 1$ and $p > 0.5$. The optimal value function $b(n)$ is $v(n/N)$ in the discrete setting of the Gambler's problem, where $v(\cdot)$ is the optimal value function under the continuous case defined in Theorem 11.*

Proof. We first verify the Bellman equation. By the definition of $v(\cdot)$ we have

$$\begin{aligned} b(n) &= v(n/N) \\ &= \max_{0<a\leq\min(n/N,1-n/N)} p\gamma\, v(n/N-a) + (1-p)\gamma\, v(n/N+a) \\ &\geq \max_{0<a\leq\min(n/N,1-n/N),Na\in\mathbb{N}} p\gamma\, v(n/N-a) + (1-p)\gamma\, v(n/N+a) \\ &= \max_{0<a\leq\min(n,N-n),a\in\mathbb{N}} p\gamma\, b(n-a) + (1-p)\gamma\, b(n+a). \end{aligned}$$

Meanwhile let $a^* = \min(n, N-n)$, Corollary 12 suggests that

$$\begin{aligned} b(n) &= v(n/N) \\ &= p\gamma\, v((n-a^*)/N) + (1-p)\gamma\, v((n+a^*)/N) \\ &= p\gamma\, b(n-a^*) + (1-p)\gamma\, v(n+a^*) \\ &\leq \max_{0<a\leq\min(n,N-n),a\in\mathbb{N}} p\gamma\, b(n-a) + (1-p)\gamma\, b(n+a). \end{aligned}$$

Therefore $b(n) = \max_{0<a\leq\min(n,N-n),a\in\mathbb{N}}\ p\gamma\, b(n-a) + (1-p)\gamma\, b(n+a)$ as desired.

We then show that $b(n) = v(n/N)$ is the unique function that satisfies the Bellman equation. The proof is similar to the proof of Lemma 2, but as both the state space and the action space are discrete the arguments will be relatively easier. Let $f(n)$ also satisfies the Bellman equation, we desire to prove that $f(n)$ is identical to $b(n)$.

Define $\delta = \max_{1\leq n\leq N-1}\ f(n) - b(n)$. This maximum must exists as there are finite many states. Then define the non-empty set $S = \{n \mid f(n) - b(n) = \delta, 1 \leq n \leq N-1\}$. For any $n' \in S$ and $a' \in \arg\max_{1\leq n\leq\min(n',N-n')} p\gamma\, f(n'-a) + (1-p)\gamma\, f(n'+a)$, we have

$$\begin{aligned} f(n') &= p\gamma\, f(n'-a') + (1-p)\gamma\, f(n'+a') \\ &\overset{(\heartsuit)}{\leq} p\gamma\, (b(n'-a')+\delta) + (1-p)\gamma\, (b(n'+a')+\delta) \\ &\leq p\gamma\, b(n'-a') + (1-p)\gamma\, b(n'+a') + \delta \end{aligned}$$

$$\begin{aligned}&\leq b(n') + \delta \\ &= f(n').\end{aligned}$$

As the equality holds, by the equality of (♥) we have $n' - a' \in S$ and $n' + a' \in S$.

Now we specify some $n_0 \in S$ and $a_0 \in \arg\max_{1 \leq n \leq \min(n_0, N-n_0)} p\gamma\ f(n_0 - a) + (1 - p)\gamma\ f(n_0 + a)$. Then, we have $n_0 - a_0 \in S$. Denote $n_1 = n_0 - a_0$ and, recursively, $a_t \in \arg\max_{1 \leq n \leq \min(n_t, N-n_t)} p\gamma\, f(n_t - a) + (1-p)\gamma\, f(n_t + a)$ and $n_{t+1} = n_t - a_t, t = 1, 2, \ldots$; Since $a_t \geq 1$ and $n_t \in \mathbb{N}$, there must exist a T such that $n_T = 0$. Therefore, $\delta = f(n_T) - b(n_T) = 0$.

By the same argument $\bar{\delta} = \max_{1 \leq n \leq N-1}\ b(n) - f(n) = 0$. Therefore, $b(n)$ and $f(n)$ are identical, as desired.

As $b(n)$ is the unique function that satisfies the Bellman equation, it is the optimal value function of the problem. □

Lemma 2. *If $v(s)$ and $f(s)$ both satisfy (**ABX**), then $v(s) = f(s)$ for all $0 \leq s \leq 1$.*

Proof. We proof the lemma by contradiction. Assume that $f(s)$ is also a solution of the system such that $f(s)$ is not identical with $v(s)$ at some s. Define $\delta = \sup_{0<s<1} f(s) - v(s)$. As $f(2^{-1}) \geq (1-p)\gamma\, f(1) + p\gamma\, f(0) = (1-p)\gamma = v(2^{-1})$, we have $\delta \geq 0$.

We show that δ cannot be zero by contradiction. If δ is zero, as $v(s)$ and $f(s)$ are not identical, there exists an s such that $f(s) < v(s)$. In this case, let $\bar{\delta} = \sup_{0<s<1} v(s) - f(s)$. Then we choose $\bar{\epsilon} = (1 - p\gamma)\bar{\delta}$ and specify s_0 such that $v(s_0) - f(s_0) > \bar{\delta} - \bar{\epsilon}$. Let $a_0 = \min(s_0, 1 - s_0)$, we have

$$\begin{aligned}v(s_0) &= (1-p)\gamma\, v(s_0 - a_0) + p\gamma\, v(s_0 + a_0) \\ &\leq (1-p)\gamma\, f(s_0 - a_0) + p\gamma\, f(s_0 + a_0) + p\gamma\bar{\delta} \\ &\leq f(s_0) + p\gamma\bar{\delta}.\end{aligned}$$

The above inequality is due to at least one of $s_0 - a_0 = 0$ and $s_0 + a_0 = 1$ must hold. Thus at least one of $v(s_0 - a_0) - f(s_0 - a_0)$ and $v(s_0 + a_0) - f(s_0 + a_0)$ must be zero. The inequality contradicts with $v(s_0) - f(s_0) > \bar{\delta} - \bar{\epsilon}$. Hence, δ cannot be zero. We discuss under $\delta > 0$ for the rest of the proof.

Case (I): $\gamma < 1$. In this case, we choose $\epsilon = (1-\gamma)\delta$. By the definition of δ we specify s_0 such that $f(s_0) > v(s_0) + \delta - \epsilon$. In fact, the existence of s_0 is by the condition $\gamma < 1$. Let $a_0 \in \arg\max_{a \in \mathcal{A}(s_0)} p\gamma\, f(s_0 - a) + (1-p)\gamma\, f(s_0 + a)$, we have

$$\begin{aligned}f(s_0) &= p\gamma\, f(s_0 - a_0) + (1-p)\gamma\, f(s_0 + a_0) \\ &\leq p\gamma\, (v(s_0 - a_0) + \delta) + (1-p)\gamma\, (v(s_0 + a_0) + \delta) \\ &= p\gamma\, v(s_0 - a_0) + (1-p)\gamma\, v(s_0 + a_0) + \gamma\delta \\ &\leq v(s_0) + \delta - \epsilon.\end{aligned}$$

The inequality $f(s_0) \leq v(s_0) + \delta - \epsilon$ contradicts with $f(s_0) > v(s_0) + \delta - \epsilon$. Hence, the lemma is proved for the case $\gamma < 1$.

Case (II): $\gamma = 1$. When there exists an s' such that $f(s') - v(s') = \delta$, we show the contradiction. Let $S = \{s | f(s) - v(s) = \delta, 0 < s < 1\} \neq \varnothing$. For any $s' \in S$ and $a' \in \arg\max_{a \in \mathcal{A}(s')} p\gamma\, f(s' - a) + (1-p)\gamma\, f(s' + a)$, we have

$$\begin{aligned}f(s') &= p\gamma\, f(s' - a') + (1-p)\gamma\, f(s' + a') \\ &= p\, f(s' - a') + (1-p)\, f(s' + a') \\ &\overset{(\heartsuit)}{\leq} p\, (v(s' - a') + \delta) + (1-p)\, (v(s' + a') + \delta) \\ &= p\, v(s' - a') + (1-p)\, v(s' + a') + \delta \\ &\overset{(\diamondsuit)}{\leq} v(s') + \delta.\end{aligned}$$

Thus, the equality in (♥) and (♦) must hold. We specify $s_0 \in S$, and by the equality of (♥) we have $f(s_0 - a_0) = v(s_0 - a_0) + \delta$, thus $s_0 - a_0 \in S$. Let $s_1 = s_0 - a_0$, and we recursively specify an arbitrary $a_i \in \arg\max_{a \in \mathcal{A}(s_i)} (1-p)\gamma\, f(s_i + a) + p\gamma\, f(s_i - a)$ and $s_{i+1} = s_i - a_i$, for $i = 1, 2, \ldots$ until $s_T = 0$ for some T, or indefinitely if such a s_T does not exist. If s_T exists and the sequence $\{s_t\}$ terminates at $s_T = 0$, then $f(s_T) = v(s_T) + \delta = \delta$ by (♥), which contradicts with the boundary condition $f(s_T) = f(0) = 0$.

We desire to show the existence of T. When there exists t and ℓ such that $s_t \in G_\ell$, by Corollary 6 we have $s_{t+1} \in G_\ell$ and inductively $s_{t'} \in G_\ell$ for all $t' \geq t$. Consider that $\{s_t\}$ is strictly decreasing and there are finite many elements in G_ℓ, $\{s_t\}$ cannot be infinite. Otherwise $s_t \notin G_\ell$ for any $t, \ell \geq 1$. Then by Corollary 9 the uniqueness of optimal action we have $s_{t+1} = 2s_t - 1$ if $s_t \geq \frac{1}{2}$, and $s_{t+1} = 0$ if $s_t \leq \frac{1}{2}$. After finite many steps of $s_{t+1} = 2s_t - 1$ we will have $s_t = 0$ for some t.

It amounts to show the existence of s' where $f(s') - v(s') = \delta$. By Lemma 8 we have the continuity of $v(s)$. Lemma 3 indicates the monotonicity of $f(s)$ on $[0, 1)$. The upper bound $f(s) \leq f(1)$ in (**X**) extends this monotonicity to the closed interval $[0, 1]$. Then by Lemma 4 we have the continuity of $f(s)$ on $(0, 1]$. By (**X**) this extends to $[0, 1]$. Thus we have the continuity of $f(s) - v(s)$, and consequently the existence of $\max_{0 \leq s' \leq 1} f(s') - v(s')$. As $f(0) - v(0) = f(1) - v(1) = 0$ and $\delta > 0$, this maximum must be attained at $0 < s' < 1$. Therefore we have the existence of $\max_{0 < s' < 1} f(s') - v(s')$, which concludes the lemma. □

Lemma 3. *Let $\gamma = 1$ and $p > 0.5$. If $f(\cdot)$ satisfies (**AB**) then $f(\cdot)$ is monotonically increasing on $[0, 1)$.*

Proof. We prove the claim by contradiction. Assume that there exists $s_1 < s_2$ where $f(s_1) > f(s_2)$. Denote $\Delta s = s_2 - s_1 > 0$ and $\Delta f = f(s_1) - f(s_2) > 0$. By induction we have

$$f(s_2 - 2^{-\ell}\Delta s) - f(s_2) \geq p^\ell \Delta f$$

for an arbitrary integer $\ell \geq 1$. Then when $s_2 + 2^{-\ell}\Delta s < 1$, by $f(s_2) \geq p f(s_2 - 2^{-\ell}\Delta s) + (1 - p) f(s_2 + 2^{-\ell}\Delta s)$,

$$\begin{aligned} f(s_2 + 2^{-\ell}\Delta s) &\leq \frac{1}{1-p} f(s_2) - \frac{p}{1-p} f(s_2 - 2^{-\ell}\Delta s) \\ &= f(s_2) + \frac{p}{1-p}(f(s_2) - f(s_2 - 2^{-\ell}\Delta s)) \\ &\leq f(s_2) + f(s_2) - f(s_2 - 2^{-\ell}\Delta s). \end{aligned}$$

This concludes $f(s_2 + 2^{-\ell}\Delta s) - f(s_2) \leq f(s_2) - f(s_2 - 2^{-\ell}\Delta s)$. By induction we have

$$f(s_2 + k2^{-\ell}\Delta s) - f(s_2 + (k-1)2^{-\ell}\Delta s) \leq f(s_2 + (k-1)2^{-\ell}\Delta s) - f(s_2 - (k-2)2^{-\ell}\Delta s)$$

for $k = 1, 2, \ldots$, when $s_2 + k2^{-\ell}\Delta s < 1$. We sum this inequality over k and get

$$\begin{aligned} f(s_2 + k2^{-\ell}\Delta s) - f(s_2) &\leq k(f(s_2) - f(s_2 - 2^{-\ell}\Delta s)) \\ &\leq -kp^\ell \Delta f. \end{aligned}$$

By letting $k = 2^n$, $\ell = n + n_0$, $s_2 + 2^{-n_0}\Delta s < 1$, and $n \to +\infty$, we have $s_2 + k2^{-\ell}\Delta s < 1$ and $-kp^\ell \Delta f \to -\infty$. The arbitrarity of n indicates the non-existence of $f(s_2 + k2^{-\ell}\Delta s)$, which contradicts with the existence of the solution $f(\cdot)$. □

Lemma 4. *Let $\gamma = 1$ and $p \geq 0.5$. If $f(s)$ is monotonically increasing on $(0, 1]$ and it satisfies (**AB**), then $f(s)$ is continuous on $(0, 1]$.*

Proof. We show the continuity by contradiction. Suppose that there exists a point $s' \in (0, 1)$ such that $f(s)$ is discontinuous at s', then there exists $\epsilon, \delta > 0$ where $f(s' + \epsilon_1) - f(s' - \epsilon_2) \geq \delta$ for any $\epsilon_1 + \epsilon_2 = \epsilon$. Then, by

$$f(s' - \frac{1}{4}\epsilon) \geq p\, f(s' - \epsilon) + (1 - p)\, f(s' + \frac{2}{4}\epsilon),$$

we have

$$f(s' - \frac{1}{4}\epsilon) - f(s' - \epsilon) \geq (1 - p)\delta / p.$$

Similarly, for $k = 1, 2, \dots,$

$$f(s' - \frac{1}{4^k}\epsilon) - f(s' - \frac{1}{4^{k-1}}\epsilon) \geq (1-p)\delta/p.$$

Let $k > ((1-p)\delta/p)^{-1}$, we have $f(s' - \frac{1}{4^k}\epsilon) - f(s' - \epsilon) \geq 1$. This contradict with the fact that $f(s)$ is bounded between 0 and 1. The continuity follows on $(0, 1)$.

If the function is discontinuous on $s = 1$, then there exists $\epsilon, \delta > 0$ where $f(1) - f(1-\epsilon_1) \geq \delta$ for any $\epsilon_1 \leq \epsilon$. The same argument holds by observing

$$f(1 - \frac{1}{2^{k-1}}\epsilon) \geq p\, f(1 - \frac{1}{2^k}\epsilon) \geq f(1).$$

The lemma follows. ☐

Lemma 5. *Let $\ell \geq 1$. For any $s \in G_\ell$,*

$$\max_{a \in (G_{\ell+1} \setminus G_\ell) \cap \mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a) < \max_{a \in G_\ell \cap \mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a).$$

Proof. It suffices to show that for any $s, a \in G_\ell$, $a \in \mathcal{A}(s)$, at least one of

$$(1-p)\gamma\, v(s+a) + p\gamma\, v(s-a) > (1-p)\gamma\, v(s+a-2^{-(\ell+1)}) + p\gamma\, v(s-a+2^{-(\ell+1)}) \tag{6}$$

and

$$(1-p)\gamma\, v(s+a-2^{-\ell}) + p\gamma\, v(s-a+2^{-\ell}) > (1-p)\gamma\, v(s+a-2^{-(\ell+1)}) + p\gamma\, v(s-a+2^{-(\ell+1)}) \tag{7}$$

is satisfied. We discuss the sufficient condition for both the inequalities respectively and show that they complement each other.

For inequality (6), according to the definition of $v(s)$,

$$\begin{aligned} v(s-a+2^{-(\ell+1)}) - v(s-a) &= \sum_{i=1}^{\ell+1} (1-p)\gamma^i b_i \prod_{j=1}^{i-1} ((1-p) + (2p-1)b_j) \\ &\quad - \sum_{i=1}^{\ell} (1-p)\gamma^i b_i \prod_{j=1}^{i-1} ((1-p) + (2p-1)b_j) \\ &= (1-p)\gamma^{\ell+1} b_{\ell+1} \prod_{j=1}^{\ell} ((1-p) + (2p-1)b_j) \\ &= (1-p)\gamma^{\ell+1} \prod_{j=1}^{\ell} ((1-p) + (2p-1)b_j), \end{aligned} \tag{8}$$

where $s - a = 0.b_1 b_2 \dots b_{\ell(2)}$ and $b_{\ell+1} = 1$. Also,

$$\begin{aligned} & v(s+a) - v(s+a-2^{-(\ell+1)}) \\ &= (1-p)c_k \gamma^k \prod_{j=1}^{k-1} ((1-p) + (2p-1)c_j) \\ &\quad - \sum_{i=k+1}^{\ell+1} (1-p)^2 \gamma^i c_i \prod_{j=1}^{k-1} ((1-p) + (2p-1)c_j) \prod_{j=k+1}^{i-1} ((1-p) + (2p-1)c_j) \\ &= (1-p)\gamma^k \prod_{j=1}^{k-1} ((1-p) + (2p-1)c_j) - \sum_{i=k+1}^{\ell+1} (1-p)^2 p^{i-k-1} \gamma^i \prod_{j=1}^{k-1} ((1-p) + (2p-1)c_j) \\ &= (1 - \sum_{i=k+1}^{\ell+1} (1-p)p^{i-k-1}\gamma^{i-k})(1-p)\gamma^k \prod_{j=1}^{k-1} ((1-p) + (2p-1)c_j) \end{aligned}$$

$$= (1-(1-p)\gamma\frac{1-(p\gamma)^{\ell-k+1}}{1-p\gamma})(1-p)\gamma^k\prod_{j=1}^{k-1}((1-p)+(2p-1)c_j)$$

$$\geq (1-(1-p)\gamma\frac{1-(p\gamma)^{\ell-k+1}}{(1-p)\gamma})(1-p)\gamma^k\prod_{j=1}^{k-1}((1-p)+(2p-1)c_j)$$

$$= p^{\ell-k+1}(1-p)\gamma^{\ell+1}\prod_{j=1}^{k-1}((1-p)+(2p-1)c_j), \tag{9}$$

where $s+a = 0.c_1c_2\dots c_{\ell(2)}$ and c_k is the last $\mathbf{1}$ bit of $s+a$. When $p^{\ell-k+1}\prod_{j=1}^{k-1}((1-p)+(2p-1)c_j) > \prod_{j=1}^{\ell}((1-p)+(2p-1)b_j)$, we have inequality (6) holds, which is one of the desired sufficient conditions.

For inequality (7), we have $s+a-2^{-\ell} = 0.c_1\dots c_{k-1}\mathbf{0}_k\mathbf{1}_{k+1}\dots\mathbf{1}_{\ell+1(2)}$ and $s-a+2^{-\ell} = 0.b_1\dots b_{k'-1}\mathbf{1}_{k'(2)}$, where k' is the last $\mathbf{0}$ bit of $s+a$. Therefore, expanding by definition yields

$$v(s+a-2^{-(\ell+1)}) - v(s+a-2^{-\ell}) = p^{\ell-k}(1-p)^2\gamma^{\ell+1}\prod_{j=1}^{k-1}((1-p)+(2p-1)c_j),$$

and

$$v(s-a+2^{-\ell}) - v(s-a+2^{-(\ell+1)}) \geq p^{\ell-k'}(1-p)^2\gamma^{\ell+1}\prod_{j=1}^{k'-1}((1-p)+(2p-1)b_j).$$

Inequality (7) thus amounts to

$$p^{\ell-k'}\prod_{j=1}^{k'-1}((1-p)+(2p-1)b_j) > p^{\ell-k}\prod_{j=1}^{k-1}((1-p)+(2p-1)c_j).$$

Recall that inequality (6) amounts to

$$p^{\ell-k+1}\prod_{j=1}^{k-1}((1-p)+(2p-1)c_j) > \prod_{j=1}^{\ell}((1-p)+(2p-1)b_j)$$

$$= (1-p)p^{\ell-k'}\prod_{j=1}^{k'-1}((1-p)+(2p-1)b_j).$$

Let N_b and N_c indicates the number of $\mathbf{1}$ bits in $b_1,\dots b_{k'-1}$ and $c_1,\dots c_{k-1}$, respectively. If $N_b+k \geq N_c+k'+1$, then inequality (6) holds. If $N_c+k' \geq N_b+k$ then inequality (7) holds. As N_b and N_c are integers, at least one of the inequality must hold, which concludes the lemma. □

Lemma 7. *Let $\ell \geq 1$. For any $s \in G_{\ell+1}$,*

$$\min(s, 1-s) \in \operatorname*{arg\,max}_{a\in G_{\ell+1}\cap\mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a).$$

Proof. We prove the lemma by induction on ℓ. The base case $\ell = 1$ is obvious since $a \in g_1$ has only one element. The base case $\ell = 2$ is also immediate by exhausting $a \in \{2^{-1}, 2^{-2}\}$ for $s = 2^{-1}$. Now we assume that for any $s \in G_\ell$,

$$\min(s, 1-s) \in \operatorname*{arg\,max}_{a\in G_{\ell}\cap\mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a).$$

We aim to prove this lemma for $\ell+1$.

As shown in Lemma 5, for $s \in G_\ell$, $\arg\max_{a\in G_{\ell+1}}(1-p)\gamma v(s+a) + p\gamma v(s-a) \in G_\ell$. By the induction assumption, $\min(s,1-s) \in \arg\max_{a\in G_\ell\cap\mathcal{A}(s)}(1-p)\gamma v(s+a) + p\gamma v(s-a) \subseteq \arg\max_{a\in G_{\ell+1}\cap\mathcal{A}(s)}(1-p)\gamma v(s+a) + p\gamma v(s-a)$. Hence, the lemma holds for $s \in G_\ell$. We assume $s \in G_{\ell+1} \setminus G_\ell$ for the rest of the proof.

For any $s \geq 2^{-1}$, that is, $s = 0.c_1c_2\dots c_{\ell+1(2)} \in G_{\ell+1}$ with $c_1 = 1$,

$$\begin{aligned}
v(s) &= \sum_{i=1}^{\ell}(1-p)\gamma^i c_i \prod_{j=1}^{i-1}((1-p)+(2p-1)c_j) \\
&= (1-p)\gamma + \sum_{i=2}^{\ell}(1-p)\gamma^i c_i \prod_{j=1}^{i-1}((1-p)+(2p-1)c_j) \\
&= (1-p)\gamma + \sum_{i=1}^{\ell-1}(1-p)\gamma^{i+1} c_{i+1}((1-p)+(2p-1)c_1) \prod_{j=1}^{i-1}((1-p)+(2p-1)c_{j+1}) \\
&= (1-p)\gamma + p\gamma\, v(0.c_2\dots c_{\ell+1(2)}) \\
&= (1-p)\gamma + p\gamma\, v(2s-1).
\end{aligned}$$

Similarly, for any $s < 2^{-1}$, that is, $s = 0.c_1c_2\dots c_{\ell+1(2)} \in G_{\ell+1}$ with $c_1 = 0$,

$$\begin{aligned}
v(s) &= \sum_{i=1}^{\ell-1}(1-p)^2\gamma^{i+1} c_{i+1} \prod_{j=1}^{i-1}((1-p)+(2p-1)c_{j+1}) \\
&= (1-p)\gamma v(2s).
\end{aligned}$$

We first discuss under $s \geq 2^{-1}+2^{-2}$. As $a \leq 1-s$, we have $s-a \geq 2^{-1}$ and $s+a \geq 2^{-1}$. Hence, the first bit after the decimal of s, $s-a$, and $s+a$ is a **1** bit. Hence,

$$\begin{aligned}
&(1-p)\gamma\, v(s+a) + p\gamma\, v(s-a) \\
&= (1-p)\gamma^2 + p\gamma\,((1-p)\gamma\, v(2s+2a-1) + p\gamma\, v(2s-2a-1)) \\
&= (1-p)\gamma^2 + p\gamma\,((1-p)\gamma\, v((2s-1)+2a) + p\gamma\, v((2s-1)-2a)).
\end{aligned}$$

We have both $2s-1 \in G_\ell$ and $2a \in G_\ell$. Hence, according to the induction assumption the maximum of $(1-p)\gamma\, v((2s-1)+2a) + p\gamma\, v((2s-1)-2a)$ is obtained at $2a = \min(2s-1, 1-(2s-1)) = 2-2s$, that is, $a = 1-s$. As $a = 1-s$ is a feasible point of $a \leq \min(s, 1-s), a \in G_{\ell+1}$, we have

$$1-s \in \operatorname*{arg\,max}_{a \in G_{\ell+1}\cap\mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a)$$

as desired.

We then discuss under the case $2^{-1} \leq s < 2^{-1}+2^{-2}$. If $s-a \geq 2^{-1}$, then we still the first bit of s, $s-a$, and $s+a$ as a **1** bit. The lemma follows the same argument as the last case. If $s-a < 2^{-1}$, we have

$$\begin{aligned}
&(1-p)\gamma\, v(s+a) + p\gamma\, v(s-a) \\
&= (1-p)^2\gamma^2 + p(1-p)\gamma^2\, v(2s+2a-1) + p(1-p)\gamma^2\, v(2s-2a) \\
&= (1-p)\gamma\,(p\gamma\, v((2s-2^{-1})-(2a-2^{-1})) + (1-p)\gamma\, v((2s-2^{-1})+(2a-2^{-1}))) \\
&\quad + (1-p)\gamma^2\,(2p-1)v((2s-2^{-1})+(2a-2^{-1})) + (1-p)^2\gamma^2.
\end{aligned}$$

We have both $2s-2^{-1} \in G_\ell$ and $2a-2^{-1} \in G_\ell$ whenever $l \geq 2$. Thus, according to the induction assumption $p\gamma\, v((2s-2^{-1})-(2a-2^{-1})) + (1-p)\gamma\, v((2s-2^{-1})+(2a-2^{-1}))$ obtains its maximum at $2a-2^{-1} = 1-(2s-2^{-1})$, that is, $a = 1-s$. We verify that $a = 1-s$ is a feasible point of $a \leq \min(s, 1-s), a \in G_{\ell+1}$. Meanwhile, according to Equation (8) and Equation (9), the function is monotonically increasing on G_ℓ for any ℓ. Hence, $v((2s-2^{-1})+(2a-2^{-1}))$ obtains the maximum at the maximum possible a, which is $a = 1-s$. Since both parts of the above equation takes their respective maximum at $a = 1-s$, we conclude that

$$1-s \in \operatorname*{arg\,max}_{a \in G_{\ell+1}\cap\mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a)$$

as desired.

In similar arguments we show that $a = s$ is a maxima when $s \leq 2^{-1}-2^{-2}$ and when $2^{-1}-2^{-2} < s < 2^{-1}$, respectively. The lemma follows. □

Lemma 8. *Both $v(s)$ and $v'(s) = \max_{a \in \mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a)$ are continuous at s if there does not exist an ℓ such that $s \in G_\ell$.*

Proof. We first proof the continuity of $v(s)$. For $s = b_1 b_2 \dots b_\ell \dots_{(2)}$, $s \notin G_\ell$ indicates that for any integer N there exists an $n_1 \geq N$ such that $b_{n_1} = 1$ and an $n_0 \geq N$ such that $b_{n_0} = 0$. The monotonicity of $v(s)$ is obvious from that flipping a **0** bit to a **1** bit will always yields a greater value. For any $s - 2^{-N} \leq s' \leq s + 2^{-N}$, we specify n_1 and n_0 such that $s - 2^{-n_1} \leq s' \leq s + 2^{-n_0}$. By the monotonicity of $v(s)$ we have

$$
\begin{aligned}
&v(s) - v(s') \leq v(s) - v(s - 2^{-n_1}) \\
&= (1-p)\gamma^{n_1} \prod_{j=1}^{n_1-1} ((1-p) + (2p-1)b_j) \cdot (1 + \sum_{i=n_1+1}^{\infty} \gamma^{i-n_1} b_i p \prod_{j=n_1+1}^{i-1} ((1-p) + (2p-1)b_j)) \\
&\quad - (1-p)\gamma^{n_1} \prod_{j=1}^{n_1-1} ((1-p) + (2p-1)b_j) \sum_{i=n_1+1}^{\infty} \gamma^{i-n_1} b_i (1-p) \prod_{j=n_1+1}^{i-1} ((1-p) + (2p-1)b_j) \\
&= (1-p)\gamma^{n_1} \prod_{j=1}^{n_1-1} ((1-p) + (2p-1)b_j) \cdot (1 + \sum_{i=n_1+1}^{\infty} \gamma^{i-n_1} b_i (2p-1) \prod_{j=n_1+1}^{i-1} ((1-p) + (2p-1)b_j)) \\
&\leq (1-p)\gamma^{n_1} p^{n_1-1} \cdot (1 + \sum_{i=n_1+1}^{\infty} \gamma^{i-n_1} (2p-1) p^{n_1-i-1}) \\
&\leq 2(1-p)\gamma^N p^{N-1}.
\end{aligned}
$$

And similarly,

$$
\begin{aligned}
v(s) - v(s') &\geq v(s) - v(s + 2^{-n_0}) \\
&\geq -(1-p)\gamma^{n_0} p^{n_0-1} \cdot (1 + \sum_{i=n_0+1}^{\infty} \gamma^{i-n_0} (2p-1) p^{n_0-i-1}) \\
&\geq -2(1-p)\gamma^N p^{N-1}.
\end{aligned}
$$

Hence, $|v(s) - v(s')|$ is bounded by $-2(1-p)\gamma^N p^{N-1}$ for $s - 2^{-N} \leq s' \leq s + 2^{-N}$. As $-2(1-p)\gamma^N p^{N-1}$ converges to zero when N approaches infinity, $v(s)$ is continuous as desired.

We then show the continuity of $v'(s) = \max_{a \in \mathcal{A}(s)} (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a)$. We first show that $v'(s)$ is monotonically increasing. In fact, for $s' \geq s$ and $0 \leq a \leq \min(s, 1-s)$, either $0 \leq a \leq \min(s', 1-s')$ or $0 \leq a + s - s' \leq \min(s', 1-s')$ must be satisfied. Let a' be a or $a + s - s'$ whoever is feasible, we have both $s' + a' \geq s + a$ and $s' - a' \geq s - a$. Specify a such that $v'(s) = (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a)$, we have

$$
v'(s') \geq (1-p)\gamma\, v(s' + a') + p\gamma\, v(s' - a') \geq v'(s).
$$

The monotonicity follows.

Similarly we let $s = b_1 b_2 \dots b_l \dots_{(2)}$. For any N, specify $n_1 \geq N$ such that $b_{n_1} = 1$ and $n_0 \geq N+2$ such that $b_{n_0} = 0$. Also let $s_0 = b_1 b_2 \dots b_{N\,(2)}$. Then for the neighbourhood set $s_0 - 2^{-(N+1)} \leq s' \leq s_0 + 2^{-(N+1)}$, $v'(s) = v(s)$ for both the ends of the interval $s_0 - 2^{-(N+1)} \in G_{N+1}$ and $s_0 + 2^{-(N+1)} \in G_{N+1}$. $|v'(s) - v'(s')|$ is then bounded by $|v(s_0 - 2^{-(N+1)}) - v(s_0 + 2^{-(N+1)})|$. As shown in Equation (8) and Equation (9), the gap between the upper and the lower bounds converges to zero when N approaches infinity. The continuity of $v'(s)$ follows. □

Lemma 10. *$v(s)$ is the unique solution of the system **(ABX)**.*

Proof. Let $v'(s) = \max_a (1-p)\gamma\, v(s+a) + p\gamma\, v(s-a)$. As per Lemma 7 we have $v(s) = v'(s)$ on the dyadic rationals $\bigcup_{\ell \geq 1} G_\ell$. Since $\bigcup_{\ell \geq 1} G_\ell$ is dense and compact on $(0, 1)$, $v(s) = v'(s)$ holds whenever both $v(s)$ and $v'(s)$ are continuous at s. Thus, for any s if there does not exist an ℓ such that $s \in G_\ell$, $v(s)$ and $v'(s)$ are continuous per Lemma 8, which then indicates $v(s) = v'(s)$. Otherwise if there exists an ℓ such that $s \in G_\ell$, as per Lemma 7 we have $v(s) = v'(s)$. Hence, the

Bellman equation (**AB**) is verified for $v(s)$. The boundary conditions (**X**) holds obviously. Finally as per Lemma 2, $v(s)$ is the unique solution to the system of Bellman equation and the boundary conditions. $\square$

Theorem 18. *Let $\gamma = 1$ and $p > 0.5$. A function $f(s)$ satisfies (**ABY**) if and only if either*

- *$f(s)$ is $v(s)$ defined in Theorem 11, or*
- *$f(0) = 0$, $f(1) = 1$, and $f(s) = C$ for all $s \in (0, 1)$, for some constant $C \geq 1$.*

Proof. It is obvious that both $f(s)$ defined above are the solutions of the system. It amounts to show that they are the only solutions. If $f(s) \leq 1$ for any s, the case has been already been discussed in the proof of Lemma 2, where $v(s)$ defined in Theorem 11 is the unique solution. For the rest of the proof, we show that $f(s) = C$ for some $C > 1$ is the unique family of solutions if there exists an s such that $f(s) > 1$.

Without the bound condition (**X**), the function $f(s)$ is not necessarily continuous on $s = 0$ and $s = 1$ and not necessarily monotonic on $s = 1$. Therefore the same arguments in the proof of Lemma 2 will not hold. However, the arguments can be extended to (**Y**) by considering the limit of $f(s)$ when s approaches 0 and 1.

By Lemma 4 the function is continuous on the open interval $(0, 1)$. Let

$$C_0 = \lim_{s \to 0^+} f(s), \;\; C_1 = \lim_{s \to 1^-} f(s).$$

Then by Lemma 3, $0 \leq C_0 < f(s) < C_1$ for $s \in (0, 1)$. Here we eliminate the case $C_0 = +\infty$ and $C_1 = +\infty$. This is because when there is a sequence of $s_t \to 0$ such that $f(s_t) > t$, then we have $f(\frac{1}{2}) \geq p\, f(s_t) + (1-p)\, f(1 - s_t) \geq (1-p)t$ for any t. Then $f(\frac{1}{2})$ does not exist. Similar arguments shows that C_1 cannot be $+\infty$.

Now specify a sequence $a_t \to \frac{1}{2}$ such that $C_0 \leq f(\frac{1}{2} - a_t) \leq C_0 + \frac{1}{t}$ and $C_1 - \frac{1}{t} \leq f(\frac{1}{2} + a_t) \leq C_1$. Then we have

$$\begin{aligned} f(\frac{1}{2}) &\geq p\, f(\frac{1}{2} - a_t) + (1-p)\, f(\frac{1}{2} + a_t) \\ &\geq p\, C_0 + (1-p)\, C_1 - \frac{1}{t}. \end{aligned}$$

As t is arbitrary we have $f(\frac{1}{2}) \geq pC_0 + (1-p)C_1$. By induction on ℓ it holds on $s \in \bigcup_{\ell \geq 1} G_\ell$ that

$$f(s) \geq C_0 + (C_1 - C_0) v(s).$$

By Lemma 4 the continuity of $f(s)$ and $v(s)$ under $\gamma = 1$, this lower bound extends beyond the dyadic rationals to the entire interval $(0, 1)$. Define $\bar{f}(s) = C_0 + (C_1 - C_0)v(s)$ for $s \in (0, 1)$, $\bar{f}(0) = C_0$, $\bar{f}(1) = C_1$. It is immediate to verify that for any $C_1 > C_0 \geq 0$, $\bar{f}(s)$ solves the system (**AY**) (without (**B**) the boundary conditions). When $C_1 - C_2 \neq 0$, by Lemma 2 Case (II) This function on $(0, 1)$ is the unique solution of the system (**AY**), given monotonicity, continuity, and the lower bound above. With the boundary conditions (**B**), we have $0 = \bar{f}(0) = C_0$ and $1 = \bar{f}(1) = C_1$, therefore $f(s) = v(s)$. This case has already been discussed as the first possible solution.

It amounts to determine $f(s)$ when $C_1 - C_0 = 0$, that is, when $f(s) = C_0$, $f(0) = 0$, $f(1)$. If $C_0 < 1$, then $f(\frac{3}{4}) < p\, f(\frac{1}{2}) + (1-p)f(1)$, which contradicts with the recursive condition (**A**). Then, $f(s) = C_0$ for some $C_0 \geq 1$ is the only set of solutions. $\square$

Lemma 20. *Let $f(s)$ satisfies (**ABZ**). If there exists $0 \leq s^- < s^+ \leq 1$ and a constant C such that $f(s^-), f(s^+) \geq C$, then $f(s) \geq C$ for all $s \in \{s^- + q(s^+ - s^-) \mid q \in \mathbb{Q}, 0 \leq q \leq 1\}$.*

Proof. The statement is immediate for $q \in \{0, 1\}$. For $0 < q < 1$ we prove the lemma by contradiction. Let $f(s^- + q(s^+ - s^-)) < C$ for some $q \in \mathbb{Q}$ while $0 < q < 1$. We define $s_0 = s^- + q(s^+ - s^-)$ and $s_{t+1} = 2s_t - s^-$ for $s_t < \frac{1}{2}(s^- + s^+)$ and $s_{t+1} = 2s_t - s^+$ for $s_t > \frac{1}{2}(s^- + s^+)$, respectively. s_{t+1} will be undefined if $s_t = \frac{1}{2}(s^- + s^+)$. Since $q \in \mathbb{Q}$, let $q = m/n$ where m and n are integers and $\gcd(m, n) = 1$. Then $(s_t - s^-)/(s^+ - s^-) = c_t/n$, where $c_t = 2^t m \bmod n$. As $\mathbb{Z}_n$ is finite, $\{s_t\}_{t \geq 0}$ can only take finite many values. Thus the sequence $\{s_t\}$ is either periodic, or terminates at some $s_t = \frac{1}{2}(s^- + s^+)$.

Then we show that $f(s_t)$ is strictly decreasing by induction. Assume that $f(s_0) > \cdots > f(s_t)$. When $s_t < \frac{1}{2}(s^- + s^+)$, $f(s_t) \geq \frac{1}{2}f(s^-) + \frac{1}{2}f(s_{t+1})$, which indicates that $f(s_{t+1}) - f(s_t) \leq f(s_t) - f(s^-) < f(s_0) - f(s^-) < 0$. When $s_t > \frac{1}{2}(s^- + s^+)$, $f(s_t) \geq \frac{1}{2}f(s_{t+1}) + \frac{1}{2}f(s^+)$, which indicates $f(s_{t+1}) - f(s_t) \leq f(s_t) - f(s^+) < 0$. The base case $f(s_1) < f(s_0)$ holds as at least one of $f(s_0) \geq \frac{1}{2}f(s^-) + \frac{1}{2}f(s_1)$ and $f(s_0) \geq \frac{1}{2}f(s_1) + \frac{1}{2}f(s^+)$ must be true. Thus we conclude that $f(s_t)$ is strictly decreasing.

If the sequence terminates at some $s_t = \frac{1}{2}(s^- + s^+)$, then $f(s_t) < f(s_1) < C$, which contradicts with $f(s_t) = f(\frac{1}{2}(s^- + s^+)) \geq \frac{1}{2}f(s^-) + \frac{1}{2}f(s^+) \geq C$. Otherwise s_t is periodic and indefinite. Denote the period as T we have $f(s_{t+T}) < f(s_t)$, which indicates $f(s_t) < f(s_t)$ as a contradiction. □

Lemma 21. *Let $f(s)$ satisfies (**ABZ**). If there exists an $\bar{s} \in \mathbb{R} \setminus \mathbb{Q}$ such that $f(\bar{s}) \geq 0$, then $f(s) \geq 0$ for all $s \in \{q\bar{s} + r \mid q, r \in \mathbb{Q}, 0 \leq q, r, \leq 1, q + r \leq 1\}$.*

Proof. Specify $s^- = \bar{s}$ and $s^+ = 1$ in Lemma 20, we have $f(\bar{s} + \frac{r}{q+r}(1 - \bar{s})) \geq 0$ whenever $0 \leq \frac{r}{q+r} \leq 1$ and $\frac{r}{q+r} \in \mathbb{Q}$. Satisfying $0 \leq q, r \leq 1$, $q + r > 0$, $q, r \in \mathbb{Q}$ will be sufficient. Specify $s^- = 0$ and $s^+ = \bar{s} + \frac{r}{q+r}(1 - \bar{s})$ in Lemma 20, we have $f(q\bar{s} + r) = f((q + r)(\bar{s} + \frac{r}{q+r}(1 - \bar{s}))) \geq 0$ whenever $q + r \leq 1$. Thus $f(q\bar{s} + r) \geq 0$ for $0 < q, r < 1$, $q, r \in \mathbb{Q}$, and $0 < q + r \leq 1$. Since the case $q = r = 0$ is immediate, the statement follows with $s \in \{q\bar{s} + r \mid q, r \in \mathbb{Q}, 0 \leq q, r, \leq 1, q + r \leq 1\}$. □