

Relatively-Secure LLM-Based Steganography via Constrained Markov Decision Processes

Yu-Shin Huang, Chao Tian, and Krishna Narayanan
 Dept. of Electrical and Computer Engineering
 Texas A&M University
 College Station, TX 77845, USA
 Email: {yushin1002, chao.tian, krn}@tamu.edu

Lizhong Zheng
 Dept. of Electrical Engineering and Computer Science
 Massachusetts Institute of Technology
 Cambridge, MA 02139, USA
 Email: lizhong@mit.edu

Abstract—THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. Linguistic steganography aims to conceal information within natural language text without being detected. An effective steganography approach should encode the secret message bits into a minimal number of language tokens while preserving the natural appearance and fluidity of the stego-texts. We present a new framework to enhance the embedding efficiency of stego-texts generated by modifying the output of a large language model (LLM). The novelty of our approach is in abstracting the sequential steganographic embedding process as a Constrained Markov Decision Process (CMDP), which takes into consideration the long-term dependencies instead of merely the immediate effects. We constrain the solution space such that the discounted accumulative total variation divergence between the selected probability distribution and the original distribution given by the LLM is below a threshold. To find the optimal policy, we first show that the functional optimization problem can be simplified to a convex problem with a finite number of variables. A closed-form solution for the optimal policy is then presented to this equivalent problem. It is remarkable that the optimal policy is deterministic and resembles water-filling in some cases. The solution suggests that usually adjusting the probability distribution for the state that has the least random transition probability should be prioritized, but the choice should be made taking into account the transition probabilities at all states instead of only the current state.

I. INTRODUCTION

In steganography, a sender Alice attempts to discreetly communicate a hidden message to Bob, the receiver, using a carrier signal, which can be in the form of text, image, audio, or video [1]–[3]. Alice transmits the encoded signal to Bob through a public channel, which is under surveillance by an eavesdropper, Eve, who seeks to detect the existence of any hidden content. A well-known example is the prisoner problem [4] where Eve is the prisoner guard in charge of deciding whether to deliver a letter to a prisoner. In this work, we consider using natural language text as the carrier signal, where the encoded message is referred to as the stego-text; see Fig. 1 for an illustration.

Traditional text steganography methods usually embed hidden messages into an existing cover text [5]–[7]. However, with the development of generative models, especially large language models (LLMs), a new approach known as coverless steganography has emerged, proving to be highly effective and secure. Generative-method-based approaches produce stego-

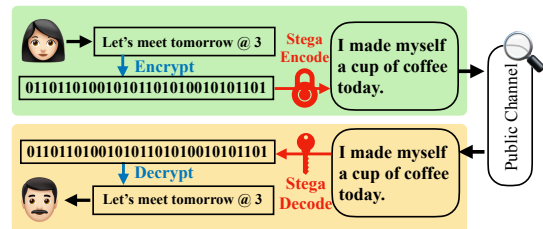


Fig. 1: Steganography system

texts that mimic natural language while also allowing for more efficient encoding of hidden information compared to traditional approaches [8]–[18]. In this approach (discussed in detail in Section II), a language model (LM) is used to produce the conditional probabilities $p(x_t|x_{t-1}, \dots, x_{t-k})$ for the token at time t and the secret message is embedded in the sampling process that samples from p_t to generate x_t . It is known that $\sum_t H(p_t)$ bits of information can be hidden in an imperceptible manner, where H is the binary entropy function.

In many scenarios, the security requirement in steganography can be relaxed, particularly when the eavesdropper Eve is computation-bounded (e.g., in a mobile device). This consideration is in fact already implicit in several previous works that adopt “near-imperceptibility” steganography [19], [20] to improve computation efficiency. Further generalizing this idea, we can view the steganography coding problem with a less-stringent security requirement as replacing the LM-based probability distribution p_t with another distribution p'_t such that $\mathbb{E}[H(p'_t)] \geq \mathbb{E}[H(p_t)]$, while maintaining an imperceptibility constraint (discussed later) and we sample the next token according to p'_t . We refer to this approach as *relatively-secure steganography*. In this work, we consider choosing such a replacement p'_t for the purpose of more efficient information embedding. This consideration is particularly important in LLM-based approaches, as several works [21]–[23] have highlighted the low embedding capacity, since p_t becomes more deterministic, leading to lower entropy and a reduced secret message embedding rate.

In a recent work [24], we considered the relatively-secure steganography coding problem, and optimized the replacement probability distribution of each token for maximum (conditional) entropy, under a constraint on the deviation from

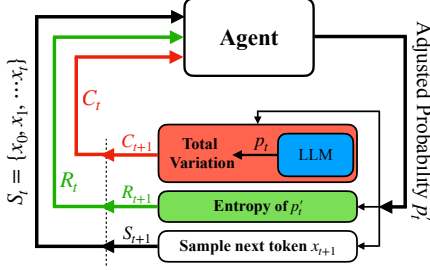


Fig. 2: Relatively-secure LM-based steganography system

the LM-based next token distribution. While this approach increases the embedding capacity, the distribution is optimized for each token position independently of others. One critical observation that motivated the current work is that the choice for the replacement probability distribution for the current token impacts not only the embedding efficiency of the immediate token, but also all future tokens.

To take into account such future impact, we propose to model the process as a decision-making agent adjusting the probability distribution for generating the next token. Selecting a particular probability distribution induces a reward R_t and a cost C_t : the reward is the entropy of the distribution (higher embedding efficiency), and the cost is the divergence from the original LM-generated distribution (unnatural text). The state of the system is the preceding tokens in the context window of the LM. As demonstrated in Fig. 2, this in fact closely resembles a Constrained Markov Decision Process (CMDP) [25]. Therefore, in this work, we formulate the relative steganography coding problem as a CMDP.

LMs are clearly too complex to model precisely. To gain insights into how p'_t should be selected, we abstract and simplify the LM into the simplest finite state model, i.e., a two-state model. Our main contribution is a closed-form solution for the optimal policy in this setting. To obtain this optimal policy, we simplify the problem from a functional optimization problem into a finite convex problem, and then establish its optimal solution in closed form. Remarkably, this implies that the optimal policy is deterministic, whereas the optimal policy for general CMDPs is usually not deterministic. Moreover, depending on the transition probability structure, the optimal policy resembles a water-filling solution in some cases. The solution suggests that usually adjusting the probability distribution for the state that has the least random transition probability should be prioritized, but the choice should be made by taking into account the transition probabilities at all states instead of only the current state.

II. PRELIMINARIES

A. LLM-based Steganography

An LLM produces the next token probability distribution $p_t = P(x_t | x_{t-1}, x_{t-2}, \dots, x_{t-k})$ conditioned on the preceding tokens in the context window of size k , which can be used to sample tokens auto-regressively to produce natural language texts [26]–[28]. We assume that the secret message bits are first encrypted using a shared random key between Alice and

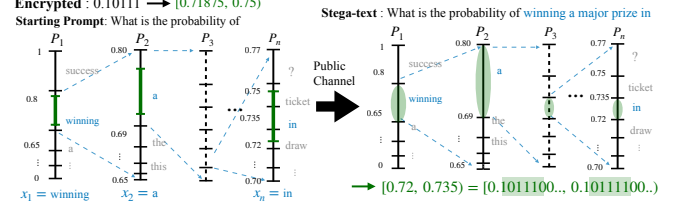


Fig. 3: Arithmetic coding based steganography

Bob, such that the encrypted message \mathbf{m} is a sequence of i.i.d. uniformly distributed bits. The key idea in embedding \mathbf{m} into a stego-text is to first use the decompressor of a compression algorithm to map the sequence of i.i.d bits \mathbf{m} into a sequence of random variables $\{X_t\}$ such that $X_t \sim p_t$ at time t . The realization of the random variable at time t , namely x_t , is the token in the stego-text at time t , and we continue this process auto-regressively.

Arithmetic coding [29] is an elegant way to generate x_t 's for *any* sequence of distributions p_t . Consider the example shown in Fig. 3 where a 5-bit encrypted message 10111 is to be embedded. The bit string is represented by the interval $\mathcal{I} = [0.101110000 \dots_2, 0.101111111 \dots_2] \simeq [0.71875, 0.75)$. We input the starting prompt “What is the probability of” to the LLM which generates p_1 at time $t = 1$, which induces a partition of the interval $[0, 1)$ and the interval \mathcal{I} corresponds to the token “winning”, implying the first token of the stego-text is $x_1 = \text{“winning”}$. Once x_1 is selected, the probability distribution p_2 is generated by adding x_1 to the prompt. The next token x_2 is chosen based on the partition of $[0, 1)$ induced by p_2 . This process is repeated to generate a stego-text of length n . During the decoding phase, Bob parses the received text into tokens and reproduces the secret bits using the exact same sequence of distributions $\{p_t\}$.

This procedure is optimal in that nearly $\sum_{t=1}^n H(p_t)$ bits from \mathbf{m} can be mapped into a stego-text of length n . It produces perfectly-secure steganography since $x_t \sim p_t$.

B. Embedding Efficiency, Deviation, and State Space Model

Low entropy distributions p_t 's imply that fewer bits of the secret message can be embedded in the stego-text. To mitigate this issue, we propose to sample from a different distribution p'_t , instead of the LLM-produced next token distribution p_t . At time- t , the problem can be naturally formulated as maximizing the entropy of p'_t , subject to the constraint that the deviation from p_t is small; this token-wise problem was studied in [24].

Mathematically, the sequential LLM-based sampling process can be viewed as a state-space model, where the state is the preceding length- k token sequence. In this work, we further simplify the state-space model to the binary state-space model where $p_s = P(S_{t+1} = 0 | S_t = s)$. We write $\mathbf{p}_s \triangleq [p_s, 1 - p_s]$ for simplicity.

C. An Introduction to CMDP

CMDP consists of several key components: the state space \mathcal{S} , action space \mathcal{A} , transition probabilities P , reward function $r : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$, a cost constraint $g : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$, a policy

$\pi : \mathcal{S} \rightarrow \Delta_{\mathcal{A}}$, the initial state distribution \mathbf{d} , and the discount factor $\gamma \in [0, 1]$, where $\Delta_{\mathcal{A}}$ represents a probability simplex with $|\mathcal{A}|$ dimensions. A policy $\pi(a|s)$ denotes the probability of selecting action a when in state s . The goal of policy optimization is to find the optimal policy $\pi(a|s)$ that guides the agent in selecting actions $a \in \mathcal{A}(s)$ for states $s \in \mathcal{S}$ in order to maximize the cumulative discounted reward over the entire timeframe, i.e.,

$$R = \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \right], \quad (1)$$

subject to the constraint

$$C = \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t g(s_t, a_t) \right] \leq b. \quad (2)$$

It is often more convenient to consider the discounted state-action pair visitation frequency $d(s, a)$ than the policy $\pi(a|s)$, also known as the occupancy measure. This quantity determines the likelihood of visiting a specific state-action pair (s, a) over time, defined as

$$d(s, a) = (1 - \gamma) \sum_{t=0}^{\infty} \gamma^t P(S_t = s, A_t = a) \quad (3)$$

The reward and cost functions can now be represented in a structured form using $d(s, a)$.

$$R = \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}(s)} r(s, a) d(s, a), \quad (4)$$

$$C = \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}(s)} g(s, a) d(s, a). \quad (5)$$

Note that $d(s, a)$ must be non-negative, and the flow conservation constraint (closely related to the Bellman's equation [30] in the dual optimization) requires that the total inflow to state s equals the total outflow:

$$\sum_{a \in \mathcal{A}(s)} d(s, a) = \mathbf{d}(s) + \gamma \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}(s')} P(s | s', a') \cdot d(s', a'). \quad (6)$$

For a stationary policy, $\pi(a|s)$ can be obtained from $d(s, a)$ as follows

$$\pi(a|s) = \frac{d(s, a)}{\sum_{a \in \mathcal{A}(s)} d(s, a)}. \quad (7)$$

We refer the readers to [31] for more details on MDPs.

III. PROBLEM FORMULATION

A. States Space, Action Space, and Policies

To formulate steganography coding as a CMDP problem, we first specify the relative components as outlined above; Fig. 4 shows the CMDP state transitions. As mentioned earlier, we abstract and simplify the state space to the simplest setting with only two states $\mathcal{S} = \{0, 1\}$. In our scenario, we define an action $a \in \mathcal{A} = [0, 1]$ as representing the likelihood of transitioning to state 0 at the subsequent time step for at a

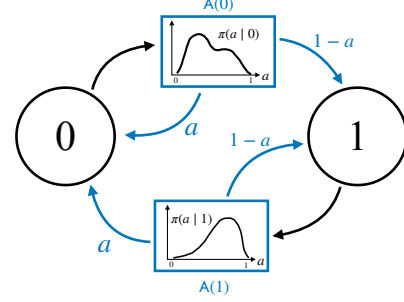


Fig. 4: Steganography with CMDP

given state $s \in \mathcal{S}$. Thus, choosing action a influences the transition probability as follows:

$$\begin{aligned} P_a(s, s') &\triangleq P(S_{t+1} = s' | S_t = s, A_t = a) \\ &= p(S_{t+1} = s' | A_t = a) \\ &= a \mathbf{1}(s' = 0) + (1 - a) \mathbf{1}(s' = 1), \end{aligned} \quad (8)$$

and the action set \mathcal{A} in fact contains uncountably many actions. For simplicity, we shall write $\mathbf{p}_a \triangleq (a, 1 - a)$.

Since the entropy of the distribution $P_a(s, s')$ directly reflects the amount of information that can be embedded in the stego-text, the reward function is simply

$$r(s, a) = H(\mathbf{p}_a), \quad s \in \{0, 1\}, a \in [0, 1] \quad (9)$$

where $H(\cdot)$ is the binary entropy function. A standard discount factor γ , $0 \leq \gamma < 1$ is assumed. The chosen probability \mathbf{p}_a at time t should not deviate too far from the original transition probability \mathbf{p}_{s_t} , therefore, the constraint cost function is

$$g(s, a) = \text{TV}(\mathbf{p}_a, \mathbf{p}_s), \quad s \in \{0, 1\}, a \in [0, 1], \quad (10)$$

where we measure the deviation by the total variation distance

$$\text{TV}(\mathbf{p}_a, \mathbf{p}_s) = |a - p_s| + |(1 - a) - (1 - p_s)| = 2|a - p_s|. \quad (11)$$

A (probabilistic) policy π is a probability distribution over the action space at state s , which can be written as

$$\pi(a|s) \triangleq P(A_t = a | S_t = s), \quad a \in \mathcal{A}, \quad (12)$$

i.e., the probability of choosing action a at state s , which is time-invariant in this setting.

B. CMDP Formulation

With the components given above, we can utilize the general CMDP framework and write out our specific problem in terms of the visitation frequency $d(s, a)$, denoted as $\mathbf{P1}$ below.

$$\mathbf{P1} : \max_{d(s, a)} : \sum_{s=0}^1 \int_0^1 H(\mathbf{p}_a) d(s, a) da \quad (13)$$

$$\text{subject to: } \int_0^1 \left(d(0, a) - \sum_{s=0}^1 \gamma a d(s, a) \right) da = \mathbf{d}_0^\gamma \quad (14)$$

$$\int_0^1 \left(d(1, a) - \sum_{s=0}^1 \gamma (1 - a) d(s, a) \right) da = \mathbf{d}_1^\gamma \quad (15)$$

$$\sum_{s=0}^1 \int_0^1 \text{TV}(\mathbf{p}_a, \mathbf{p}_s) d(s, a) da \leq b \quad (16)$$

$$d(s, a) \geq 0, \forall a \in [0, 1], s \in \{0, 1\} \quad (17)$$

where we denote the discounted initials $(1 - \gamma)\mathbf{d}(s)$ as \mathbf{d}_s^γ .

In the formulation, (14) and (15) reflect the flow conservation constraint, aligning with (6) by substituting the transition probability $P(s | s', a')$ with a and $1 - a$. It is clear that (15) can be replaced by (18)

$$\int_0^1 d(0, a) + d(1, a) da = 1, \quad (18)$$

in order to simplify the problem, since it is simply the summation of (14) and (15).

By determining the optimal $d(s, a)$ from problem **P1**, the optimal policy π can be computed via (7). For each fixed s , the visitation frequency $d(s, a)$ is a function of the state a on $[0, 1]$. Therefore, finding the optimal visitation frequency $d(s, a)$ is a functional optimization problem, even though it is a linear program in terms of $d(s, a)$.

IV. THE OPTIMAL POLICY

We establish the optimal policy π in two steps: 1) First we simplify the functional optimization to a standard convex optimization problem in \mathbb{R}^4 , and 2) We provide an explicit optimal solution to this simplified optimization problem.

A. A Simplified Optimization Problem

We first show that the optimization problem **P1** can be simplified to an optimization in \mathbb{R}^4 , given in **P3**. Mathematically, **P3** is a standard finite-dimensional convex optimization problem, which is significantly simpler than the functional optimization problem **P1**. Operationally, once we solve **P3**, we can obtain the parameters (a_0, a_1) that immediately gives us the optimal policy: at state s , choose the probability distribution $(a_s, 1 - a_s)$ to generate the next token (state). The policy itself is deterministic, and the randomness in generating the next token comes purely from the encrypted message bits. We note the generic de-randomization technique given in [25] will lead to a CMDP with two possible actions at each state, instead of only one as in our reduction.

We define an intermediate optimization problem as follows.

$$\mathbf{P2} : \max_{a_0, a_1, d_0, d_1} : d_0 H(a_0) + d_1 H(a_1) \quad (19)$$

$$\text{subject to: } (1 - \gamma a_0)d_0 - \gamma a_1 d_1 = \mathbf{d}_0^\gamma \quad (20)$$

$$d_0 + d_1 = 1 \quad (21)$$

$$d_0 \text{TV}_0(a_0) + d_1 \text{TV}_1(a_1) \leq b \quad (22)$$

$$d_0, d_1 \geq 0 \quad (23)$$

$$0 \leq a_0, a_1 \leq 1, \quad (24)$$

where we write $H(\mathbf{p}_a)$ as $H(a)$, and $\text{TV}(\mathbf{p}_a, \mathbf{p}_s)$ as $\text{TV}_s(a)$ for simplicity.

Theorem 1. $(\mathbf{P1}) = (\mathbf{P2})$, where “=” stands for equivalence between the two problems.

The proof of Theorem 1 is given in Appendix A¹ of the accompanying long version, where we utilize the convexity of the objective function and the constraint functions to show that a weighted average of the probabilistic policy is without loss of optimality. The problem **P2** is however not necessarily convex. We next show that a reparametrization of the problem leads to another equivalent problem that is indeed convex. The equivalent problem is given below.

$$\mathbf{P3} : \max_{x_0, x_1, d_0, d_1} : d_0 H\left(\frac{x_0}{d_0}\right) + d_1 H\left(\frac{x_1}{d_1}\right) \quad (25)$$

$$\text{subject to: } (d_0 - \gamma x_0) - \gamma x_1 = \mathbf{d}_0^\gamma \quad (26)$$

$$d_0 + d_1 = 1 \quad (27)$$

$$d_0 \text{TV}_0\left(\frac{x_0}{d_0}\right) + d_1 \text{TV}_1\left(\frac{x_1}{d_1}\right) \leq b \quad (28)$$

$$d_s \geq 0, s \in \{0, 1\} \quad (29)$$

$$0 \leq \frac{x_s}{d_s} \leq 1, s \in \{0, 1\}. \quad (30)$$

Clearly, we have simply replaced a_0, a_1 by $x_0 = a_0 d_0$ and $x_1 = a_1 d_1$. Since the constraints (26), (27), (29), and (30) are linear, we only need to consider the objective function (25) and the constraint (28). Recall that the perspective function $f_2(x, t)$ of a function $f_1(x)$ is defined as

$$f_2(x, t) = t f_1\left(\frac{x}{t}\right), \forall t > 0 \quad (31)$$

A well-known property of the perspective function [32] is that if $f_1(\cdot)$ is convex, $f_2(\cdot, \cdot)$ is likewise convex. Since the binary entropy function $H(\cdot)$ is concave and the total variation function is convex, (25) is indeed concave and (28) is indeed convex, implying that the problem **P3** is a convex optimization problem. When substituting a_s with $\frac{x_s}{d_s}$, it is essential to ensure d_s is non-zero, as d_s indicates the frequency of visiting state s . If $d_s = 0$, it implies state s is never visited. In such a scenario, the state can be excluded from the system. Thus, without loss of generality, we assume d_s is positive.

B. The Optimal Policy

Before presenting the solution, we first note that by symmetry, we only need to consider the case $p_0, p_1 \geq \frac{1}{2}$ and $p_1 \geq \frac{1}{2} > p_0$, since otherwise, we can simply rename the two states; more details on this relabeling can be found in Appendix C¹ in the accompanying longer version. Let us also define the following auxiliary functions:

$$\eta_0^\pm(b) = \frac{\frac{b}{2}(1 + \gamma p_1) \pm p_0(\mathbf{d}_0^\gamma + \gamma p_1)}{\pm(\gamma p_1 + \mathbf{d}_0^\gamma) + \gamma \frac{b}{2}} \quad (32)$$

$$\eta_1^\pm(b) = \frac{\frac{b}{2}(1 - \gamma p_0) \pm p_1(1 - \gamma p_0 - \mathbf{d}_0^\gamma)}{\pm(1 - \gamma p_0 - \mathbf{d}_0^\gamma) - \gamma \frac{b}{2}}. \quad (33)$$

The following two theorems give the optimal policy, the proofs of which can be found in the appendix of the accompanying long version¹.

Theorem 2. When $p_0, p_1 \geq \frac{1}{2}$, the optimal policy choices at the two states are:

¹Long version: <https://github.com/Yu-Shin-Huang/Stega-via-CMDP.git>

1) $b \in [0, b_l)$

$$\begin{cases} a_0 = p_0, a_1 = \eta_1^-(b), & \text{if } |\frac{1}{2} - p_1| \geq |\frac{1}{2} - p_0| \\ a_0 = \eta_0^-(b), a_1 = p_1, & \text{if } |\frac{1}{2} - p_1| < |\frac{1}{2} - p_0| \end{cases} \quad (34)$$

2) $b \in [b_l, b_h)$

$$a_0 = a_1 = -M(\frac{b}{2} - \mathbf{d}_0^\gamma p_0 - \mathbf{d}_1^\gamma p_1) \quad (35)$$

3) $b \in [b_h, \infty)$

$$a_0 = a_1 = \frac{1}{2} \quad (36)$$

where we defined

$$b_l = 2 \max\{(1 - \gamma p_0 - \mathbf{d}_0^\gamma)(p_1 - p_0), (\mathbf{d}_0^\gamma + \gamma p_1)(p_0 - p_1)\} \quad (37)$$

$$b_h = (2\mathbf{d}_0^\gamma + \gamma)(p_0 - p_1) + 2p_1 - 1 \quad (38)$$

$$M = \frac{1}{1 - \gamma p_0 + \gamma p_1}. \quad (39)$$

There are three regimes in the solution:

- 1) The highly-constrained regime: The optimal policy is to replace the distribution at the state that generates lower entropy tokens with a more uniform distribution, while for the other state, the distribution is maintained;
- 2) The intermediately-constrained regime: the distributions at both states should be replaced by the same (more uniform) distribution;
- 3) The unconstrained regime: the distributions at both states should be replaced by the uniform distribution.

The behavior resembles a water-filling solution in some sense; see Fig. 5a. Since p_1 is situated farther from $\frac{1}{2}$, in the highly-constrained regime, the optimal policy is to retain $a_0 = p_0$ and adjust only a_1 . Once $a_1 = p_0$, the two states become equivalent, and $a_0 = a_1$ needs to be adjusted together, until they both become $\frac{1}{2}$.

Theorem 3. When $p_1 \geq \frac{1}{2} > p_0$, the optimal policy choices at the two states are:

1) $b \in [0, b'_l)$

$$\begin{cases} a_0 = p_0, a_1 = \eta_1^-(b), & \text{if } |\frac{1}{2} - p_1| \geq |\frac{1}{2} - p_0| \\ a_0 = \eta_0^+(b), a_1 = p_1, & \text{if } |\frac{1}{2} - p_1| < |\frac{1}{2} - p_0| \end{cases} \quad (40)$$

2) $b \in [b'_l, b'_h]$

$$(a_0, a_1) \in S_b \triangleq \left\{ (a_0, a_1) \in [p_0, \frac{1}{2}] \times [\frac{1}{2}, p_1] : \Phi_-(a_1) = \Phi_+(a_0) \cap m(a_0, a_1) = \frac{b}{2} \right\} \quad (41)$$

3) $b \in (b'_h, \infty)$

$$a_0 = a_1 = \frac{1}{2} \quad (42)$$

where we define

$$b'_l = 2 \left(\frac{\mathbf{d}_0^\gamma + \gamma p_1}{1 - \gamma p_0 + \gamma p_1} \right) (\psi_0 - p_0) u(1 - p_0 - p_1)$$

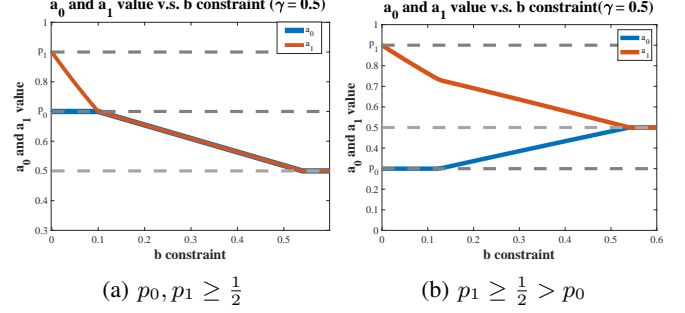


Fig. 5: a_0, a_1 value v.s. b constraint with $\mathbf{d} = [0.8, 0.2]$

$$+ 2 \left(\frac{1 - \gamma p_0 - \mathbf{d}_0^\gamma}{1 - \gamma p_0 + \gamma p_1} \right) (p_1 - \psi_1) u(-1 + p_0 + p_1) \quad (43)$$

$$b'_h = (2\mathbf{d}_0^\gamma + \gamma)(1 - p_0 - p_1) + 2p_1 - 1 \quad (44)$$

and

$$\Phi_\pm(a) = (\pm 1 + \gamma p_0 + \gamma p_1) \log \left(\frac{1 - a}{a} \right) - 2\gamma \log(1 - a) \quad (45)$$

$$\Psi_{p_0}(a_1) = \Phi_+(p_0) - \Phi_-(a_1), \quad \psi_0 = \Psi_{p_1}^{-1}(0) \quad (46)$$

$$\Psi_{p_1}(a_0) = \Phi_+(a_0) - \Phi_-(p_1), \quad \psi_1 = \Psi_{p_0}^{-1}(0) \quad (47)$$

$$m(a_0, a_1) = d_0(a_0 - p_0) + d_1(p_1 - a_1) \quad (48)$$

and $u(\cdot)$ is a unit step function in b'_l .

There are again three regimes, and the only difference from the previous case is in the second regime, the policy a_0 and a_1 are not the same; see Fig. 5b for an illustration. In the second regime, a_0 and a_1 are the solution of a pair of equations given above, which can not be solved explicitly. However, we show in the proof of this theorem that it admits a unique solution for any $b \in [b'_l, b'_h]$, which guarantees that a_0 and a_1 can be found numerically. Notice also that since ψ_0 and ψ_1 are the solutions to $\Psi_{p_0} = 0$ and $\Psi_{p_1} = 0$, respectively, they can be written as $\Phi_+(p_0) = \Phi_-(\psi_1)$ and $\Phi_+(\psi_0) = \Phi_-(p_1)$; the solution determines the boundary of the regimes.

We further note that the visitation frequency d_s can be expressed in terms of a_0 and a_1 using equation (49):

$$d_0 = \frac{\mathbf{d}_0^\gamma + \gamma a_1}{1 - \gamma a_0 + \gamma a_1}, \quad d_1 = \frac{1 - \gamma a_0 - \mathbf{d}_0^\gamma}{1 - \gamma a_0 + \gamma a_1} \quad (49)$$

V. CONCLUSION

In this study, we presented a novel formulation for relatively secure LLM-based steganography in the Constrained Markov Decision Process (CMDP) framework. This approach considers the impact on both the immediate and future token generations, and provides insights into how the optimal policy should be chosen. The solution suggests that when choosing the adjustment of the probability distribution, the choice should be made considering the LLM-generated transition probabilities at all states, and usually adjusting the distribution for the state that has the least random transition probability distribution should be prioritized.

REFERENCES

- [1] R. J. Anderson and F. A. Petitcolas, "On the limits of steganography," *IEEE Journal on selected areas in communications*, vol. 16, no. 4, pp. 474–481, 1998.
- [2] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital watermarking and steganography*. Morgan kaufmann, 2007.
- [3] N. Provos and P. Honeyman, "Hide and seek: An introduction to steganography," *IEEE security & privacy*, vol. 1, no. 3, pp. 32–44, 2003.
- [4] G. J. Simmons, "The prisoners' problem and the subliminal channel," in *Advances in Cryptology: Proceedings of Crypto 83*. Springer, 1984, pp. 51–67.
- [5] U. Topkara, M. Topkara, and M. J. Atallah, "The hiding virtues of ambiguity: quantifiably resilient watermarking of natural language text through synonym substitutions," in *Proceedings of the 8th workshop on Multimedia and security*, 2006, pp. 164–174.
- [6] C. Y. Chang and S. Clark, "Linguistic steganography using automatically generated paraphrases," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 2010, pp. 591–599.
- [7] I. Safaka, C. Fragouli, and K. Argyraki, "Matryoshka: Hiding secret communication in plain sight," in *6th USENIX Workshop on Free and Open Communications on the Internet (FOCI 16)*, 2016.
- [8] T. Fang, M. Jaggi, and K. Argyraki, "Generating steganographic text with LSTMs," *arXiv preprint arXiv:1705.10742*, 2017.
- [9] Z.-L. Yang, X.-Q. Guo, Z.-M. Chen, Y.-F. Huang, and Y.-J. Zhang, "RNN-stega: Linguistic steganography based on recurrent neural networks," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1280–1295, 2018.
- [10] Z. Ziegler, Y. Deng, and A. M. Rush, "Neural linguistic steganography," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 1210–1215.
- [11] L. Xiang, X. Wang, C. Yang, and P. Liu, "A novel linguistic steganography based on synonym run-length encoding," *IEICE transactions on Information and Systems*, vol. 100, no. 2, pp. 313–322, 2017.
- [12] Z.-L. Yang, S.-Y. Zhang, Y.-T. Hu, Z.-W. Hu, and Y.-F. Huang, "VAE-Stega: linguistic steganography based on variational auto-encoder," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 880–895, 2020.
- [13] Z. Yang, N. Wei, Q. Liu, Y. Huang, and Y. Zhang, "GAN-TStega: Text steganography based on generative adversarial networks," in *Digital Forensics and Watermarking: 18th International Workshop, IWDW 2019, Chengdu, China, November 2–4, 2019, Revised Selected Papers 18*. Springer, 2020, pp. 18–31.
- [14] C. S. de Witt, S. Sokota, J. Z. Kolter, J. Foerster, and M. Strohmeier, "Perfectly secure steganography using minimum entropy coupling," *arXiv preprint arXiv:2210.14889*, 2022.
- [15] G. Kaptchuk, T. M. Jois, M. Green, and A. D. Rubin, "Meteor: Cryptographically secure steganography for realistic distributions," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, pp. 1529–1548.
- [16] X. Zhang, K. Chen, J. Ding, Y. Yang, W. Zhang, and N. Yu, "Provably secure public-key steganography based on elliptic curve cryptography," *IEEE Transactions on Information Forensics and Security*, 2024.
- [17] S. Zhang, Z. Yang, J. Yang, and Y. Huang, "Provably secure generative linguistic steganography," *arXiv preprint arXiv:2106.02011*, 2021.
- [18] J. Ding, K. Chen, Y. Wang, N. Zhao, W. Zhang, and N. Yu, "Dis-cop: Provably secure steganography in practice based on" distribution copies", in *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2023, pp. 2238–2255.
- [19] F. Dai and Z. Cai, "Towards near-imperceptible steganographic text," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 4303–4308.
- [20] J. Shen, H. Ji, and J. Han, "Near-imperceptible neural linguistic steganography via self-adjusting arithmetic coding," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 303–313.
- [21] G. Liao, J. Yang, K. Pang, and Y. Huang, "Co-Stega: Collaborative linguistic steganography for the low capacity challenge in social media," in *Proceedings of the 2024 ACM Workshop on Information Hiding and Multimedia Security*, 2024, pp. 7–12.
- [22] M. Bai, J. Yang, K. Pang, Y. Huang, and Y. Gao, "Semantic steganography: A framework for robust and high-capacity information hiding using large language models," *arXiv preprint arXiv:2412.11043*, 2024.
- [23] K. Pang, "FreStega: A plug-and-play method for boosting imperceptibility and capacity in generative linguistic steganography for real-world scenarios," *arXiv preprint arXiv:2412.19652*, 2024.
- [24] Y.-S. Huang, P. Just, K. Narayanan, and C. Tian, "OD-Stega: LLM-based near-imperceptible steganography via optimized distributions," *arXiv preprint arXiv:2410.04328*, 2024.
- [25] E. Altman, *Constrained Markov decision processes*. Routledge, 2021.
- [26] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.
- [27] T. B. Brown, "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165*, 2020.
- [28] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar *et al.*, "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.
- [29] J. Rissanen and G. G. Langdon, "Arithmetic coding," *IBM Journal of research and development*, vol. 23, no. 2, pp. 149–162, 1979.
- [30] R. E. Bellman and S. E. Dreyfus, *Applied dynamic programming*. Princeton university press, 2015, vol. 2050.
- [31] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
- [32] S. Boyd and L. Vandenberghe, "Convex optimization," *Cambridge University Press*, 2004.

APPENDIX A
PROOF OF THEOREM 1

Assuming a general optimal solution $\tilde{d}(s, a)$ exists for problem **P1**, we claim that a new solution $d^*(s, a) = d_s^* \delta(a_s^*)$ is at least as good, where

$$\begin{cases} d_s^* = \int_0^1 \tilde{d}(s, a) da \\ a_s^* = \int_0^1 a \frac{\tilde{d}(s, a)}{d_s^*} da \end{cases} \quad (50)$$

and $\delta(\cdot)$ is a Dirac-delta function.

To prove this claim, we need to show that the function $d^*(s, a)$ meets the constraints (14)-(17), and the objective function value (13) is at least as large.

- 1) Constraints (14), (17) and (18): It is clear that d_s^* is positive for all $s \in \{0, 1\}$ and $\sum_{s=0}^1 d_s^*$ sum up to 1. Therefore $d^*(s, a)$ satisfies the constraint (17) and (18). For constraint (14),

$$\begin{aligned} & \int_0^1 (1 - \gamma a) d^*(0, a) da - \int_0^1 \gamma a d^*(1, a) da \\ &= d_0^* - \gamma a_0^* d_0^* - \gamma a_1^* d_1^* \\ &= \int_0^1 \tilde{d}(0, a) da - \gamma d_0^* \int_0^1 a \frac{\tilde{d}(0, a)}{d_0^*} da \\ & \quad - \gamma d_1^* \int_0^1 a \frac{\tilde{d}(1, a)}{d_1^*} da \\ &= \int_0^1 \tilde{d}(0, a) da - \gamma \int_0^1 a \tilde{d}(0, a) da - \gamma \int_0^1 a \tilde{d}(1, a) da \\ &= \int_0^1 (1 - \gamma a) \tilde{d}(0, a) da - \gamma \int_0^1 a \tilde{d}(1, a) da = \mathbf{d}_0^\gamma, \end{aligned} \quad (51)$$

where the equality in (51) follows from the fact that $\tilde{d}(s, a)$ satisfies the constraint (14).

- 2) Constraint (16): We write

$$\begin{aligned} b &\geq \int_0^1 \text{TV}_0(a) \tilde{d}(0, a) + \text{TV}_1(a) \tilde{d}(1, a) da \\ &= \int_0^1 d_0^* \text{TV}_0(a) \left(\frac{\tilde{d}(0, a)}{d_0^*} \right) + d_1^* \text{TV}_1(a) \left(\frac{\tilde{d}(1, a)}{d_1^*} \right) da \\ &\geq d_0^* \text{TV}_0 \left(\int_0^1 a \left(\frac{\tilde{d}(0, a)}{d_0^*} \right) da \right) \\ & \quad + d_1^* \text{TV}_1 \left(\int_0^1 a \left(\frac{\tilde{d}(1, a)}{d_1^*} \right) da \right) \\ &= d_0^* \text{TV}_0(a_0^*) + d_1^* \text{TV}_1(a_1^*) \\ &= \int_0^1 d^*(0, a) \text{TV}_0(a) + d^*(1, a) \text{TV}_1(a) da \end{aligned} \quad (52)$$

where the inequality follows the Jensen's inequality since the total variation divergence is convex.

- 3) The objective function: We wish to show that $d^*(s, a)$ achieves a value equal to or exceeding that of $\tilde{d}(s, a)$. For this purpose, we write:

$$\int_0^1 H(a) \tilde{d}(0, a) da + \int_0^1 H(a) \tilde{d}(1, a) da$$

$$\begin{aligned} &= d_0^* \int_0^1 H(a) \frac{\tilde{d}(0, a)}{d_0^*} da + d_1^* \int_0^1 H(a) \frac{\tilde{d}(1, a)}{d_1^*} da \\ &\leq d_0^* H \left(\int_0^1 a \frac{\tilde{d}(0, a)}{d_0^*} da \right) + d_1^* H \left(\int_0^1 a \frac{\tilde{d}(1, a)}{d_1^*} da \right) \\ &= d_0^* H(a_0^*) + d_1^* H(a_1^*) \\ &= \int_0^1 H(a) d^*(0, a) da + \int_0^1 H(a) d^*(1, a) da, \end{aligned} \quad (53)$$

where again the inequality is due to Jensen's inequality since the entropy function is concave.

We can now conclude that $d^*(s, a)$ is also an optimal solution to the optimization problem. The solution $d^*(s, a)$ further implies that the optimal policy is

$$\pi(a|s) = \frac{d^*(s, a)}{\int_0^1 d^*(s, a) da} = \begin{cases} 1, & a = a_s^* \\ 0, & \text{else} \end{cases} \quad (54)$$

which indicates that there is only one action a_s^* in interval $[0, 1]$ to take in each state s , i.e., a deterministic policy is optimal. Substituting the variable $d(s, a)$ in problem **P1** with $d^*(s, a)$ will yield an equivalent problem to **P2**. In other words, this optimization problem reduces the number of variables from uncountably many to merely four: the action points a_0^*, a_1^* , and the discounted state visitation distribution d_0^*, d_1^* .

APPENDIX B
PROOF OF THEOREM 2 AND THEOREM 3

To solve the convex optimization problem **P3**, our plan is to find a solution to the KKT conditions, since the problem is convex. However, the total variation constraint poses a challenge as it is not differentiable when $\frac{x_0}{d_0} = p_0$ and $\frac{x_1}{d_1} = p_1$, preventing us from performing a direct partial differentiation on this inequality constraint. Therefore, we introduce a set of auxiliary variables c_0, c_1 such that $d_0 |\frac{x_0}{d_0} - p_0| = |x_0 - d_0 p_0| \leq c_0$ and $d_1 |\frac{x_1}{d_1} - p_1| = |x_1 - d_1 p_1| \leq c_1$. The optimization problem **P3** can be reformulated as follows.

$$\max_{x_0, x_1, d_0, d_1, c_0, c_1} d_0 H \left(\frac{x_0}{d_0} \right) + d_1 H \left(\frac{x_1}{d_1} \right) \quad (55)$$

$$\text{subject to } c_0 + c_1 \leq \frac{b}{2} \quad (56)$$

$$-c_0 \leq x_0 - d_0 p_0 \leq c_0 \quad (57)$$

$$-c_1 \leq x_1 - d_1 p_1 \leq c_1 \quad (58)$$

$$d_0 - \gamma x_0 - \gamma x_1 = \mathbf{d}_0^\gamma \quad (59)$$

$$d_0 + d_1 = 1 \quad (60)$$

$$0 \leq x_0 \leq d_0, 0 \leq x_1 \leq d_1 \quad (61)$$

The Lagrangian function of this problem is:

$$\begin{aligned} \mathcal{L} &= - \left(d_0 H \left(\frac{x_0}{d_0} \right) + d_1 H \left(\frac{x_1}{d_1} \right) \right) + \lambda \left(c_0 + c_1 - \frac{b}{2} \right) \\ & \quad + \alpha_0 (-c_0 - x_0 + d_0 p_0) + \beta_0 (x_0 - d_0 p_0 - c_0) \\ & \quad + \alpha_1 (-c_1 - x_1 + d_1 p_1) + \beta_1 (x_1 - d_1 p_1 - c_1) \\ & \quad + \omega_0 (d_0 - \mathbf{d}_0^\gamma - \gamma(x_0 + x_1)) + \omega_1 (d_0 + d_1 - 1) \\ & \quad - \nu_0 x_0 - \nu_1 x_1 + \mu_0 (x_0 - d_0) + \mu_1 (x_1 - d_1) \end{aligned} \quad (62)$$

where $\lambda, \alpha_s, \beta_s, \omega_s, \nu_s, \mu_s$ are the Lagrange multipliers of constraints (56) to (61). The KKT conditions are written as follows:

1) Stationarity:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_0} &= -\log\left(\frac{d_0 - x_0}{x_0}\right) - \alpha_0 + \beta_0 \\ &\quad - \gamma\omega_0 - \nu_0 + \mu_0 = 0 \end{aligned} \quad (63)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_1} &= -\log\left(\frac{d_1 - x_1}{x_1}\right) - \alpha_1 + \beta_1 \\ &\quad - \gamma\omega_0 - \nu_1 + \mu_1 = 0 \end{aligned} \quad (64)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial d_0} &= -H\left(\frac{x_0}{d_0}\right) + \frac{x_0}{d_0} \log\left(\frac{d_0 - x_0}{x_0}\right) \\ &\quad + \alpha_0 p_0 - \beta_0 p_0 + \omega_0 + \nu_1 - \mu_0 = 0 \end{aligned} \quad (65)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial d_1} &= -H\left(\frac{x_1}{d_1}\right) + \frac{x_1}{d_1} \log\left(\frac{d_1 - x_1}{x_1}\right) \\ &\quad + \alpha_1 p_1 - \beta_1 p_1 + \omega_1 - \mu_1 = 0 \end{aligned} \quad (66)$$

$$\frac{\partial \mathcal{L}}{\partial c_0} = \lambda - \alpha_0 - \beta_0 = 0 \quad (67)$$

$$\frac{\partial \mathcal{L}}{\partial c_1} = \lambda - \alpha_1 - \beta_1 = 0 \quad (68)$$

2) Primal feasibility:

$$c_0 + c_1 \leq \frac{b}{2} \quad (69)$$

$$-c_0 - x_0 + d_0 p_0 \leq 0, \quad -c_1 - x_1 + d_1 p_1 \leq 0 \quad (70)$$

$$x_0 - d_0 p_0 - c_0 \leq 0, \quad x_1 - d_1 p_1 - c_1 \leq 0 \quad (71)$$

$$d_0 = \mathbf{d}_0^\gamma + \gamma(x_0 + x_1) \quad (72)$$

$$d_0 + d_1 = 1 \quad (73)$$

$$x_0 \geq 0, \quad x_1 \geq 0 \quad (74)$$

$$x_0 - d_0 \leq 0, \quad x_1 - d_1 \leq 0 \quad (75)$$

3) Dual feasibility:

$$\lambda \geq 0 \quad (76)$$

$$\alpha_0 \geq 0, \quad \alpha_1 \geq 0 \quad (77)$$

$$\beta_0 \geq 0, \quad \beta_1 \geq 0 \quad (78)$$

$$\nu_0 \geq 0, \quad \nu_1 \geq 0 \quad (79)$$

$$\mu_0 \geq 0, \quad \mu_1 \geq 0 \quad (80)$$

4) Complementary slackness:

$$\lambda \left(c_0 + c_1 - \frac{b}{2} \right) = 0 \quad (81)$$

$$\begin{aligned} \alpha_0(-c_0 - x_0 + d_0 p_0) \\ = \alpha_1(-c_1 - x_1 + d_1 p_1) = 0 \end{aligned} \quad (82)$$

$$\beta_0(x_0 - d_0 p_0 - c_0) = \beta_1(x_1 - d_1 p_1 - c_1) = 0 \quad (83)$$

$$\omega_0(d_0 - \mathbf{d}_0^\gamma - \gamma(x_0 + x_1)) = 0 \quad (84)$$

$$\omega_1(d_0 + d_1 - 1) = 0 \quad (85)$$

$$\nu_0 x_0 = \nu_1 x_1 = 0 \quad (86)$$

$$\mu_0(x_0 - d_0) = \mu_1(x_1 - d_1) = 0 \quad (87)$$

The solution that satisfies the KKT conditions will be presented in several cases, depending on the values of p_0 and p_1 , separated into different regimes of the cost constraint b . The expression formulas for the primal variables x_s, d_s and the dual variables ω_s, ν_s, μ_s where $s \in \{0, 1\}$ are identical across all scenarios, as presented below.

$$x_0 = a_0 d_0 \quad (88)$$

$$x_1 = a_1 d_1 \quad (89)$$

$$d_0 = \frac{\mathbf{d}_0^\gamma + \gamma a_1}{1 - \gamma a_0 + \gamma a_1} \quad (90)$$

$$d_1 = 1 - d_0 \quad (91)$$

$$\begin{aligned} \omega_0 &= \frac{-1}{\gamma} \log\left(\frac{1 - a_0}{a_0}\right) - \frac{1}{\gamma} \alpha_0 + \frac{1}{\gamma} \beta_0 \\ &= \frac{-1}{\gamma} \log\left(\frac{1 - a_1}{a_1}\right) - \frac{1}{\gamma} \alpha_1 + \frac{1}{\gamma} \beta_1 \end{aligned} \quad (92)$$

$$\omega_1 = -\log(1 - a_1) - \alpha_1 p_1 + \beta_1 p_1 \quad (93)$$

$$\nu_s = \mu_s = 0, \quad s \in \{0, 1\} \quad (94)$$

Therefore, we will not write them out particularly in each case, since they can be derived from $a_s, \alpha_s, \beta_s, s \in \{0, 1\}$.

We define the mathematical expressions below for simplifications:

$$\begin{aligned} z_0(a_0, a_1) &= (-1 - \gamma p_1) \log\left(\frac{1 - a_0}{a_0}\right) \\ &\quad + \gamma p_1 \log\left(\frac{1 - a_1}{a_1}\right) + \gamma \log\left(\frac{1 - a_0}{1 - a_1}\right) \end{aligned} \quad (95)$$

$$\begin{aligned} z_1(a_0, a_1) &= (-\gamma p_0) \log\left(\frac{1 - a_0}{a_0}\right) \\ &\quad + (-1 + \gamma p_0) \log\left(\frac{1 - a_1}{a_1}\right) + \gamma \log\left(\frac{1 - a_0}{1 - a_1}\right) \end{aligned} \quad (96)$$

A. Proof of Theorem 2

In Theorem 2, the assumption is that $p_0, p_1 \geq \frac{1}{2}$.

There are three regimes:

1) The regime $b \in [0, b_l]$: In this regime, the two cases $p_1 \geq p_0 \geq \frac{1}{2}$ and $p_0 > p_1 \geq \frac{1}{2}$ lead to different solution expressions, and let us consider the first case for now.

With the solution of a_0, a_1 given in the theorem, and in the given range of b , the range of a_0 and a_1 are

$$a_0 = p_0, \quad a_1 \in [p_0, p_1], \quad (97)$$

The variable assignments are as follows.

a) Primal variable assignments:

$$a_0 = p_0, \quad a_1 = \eta_1^-(b) \quad (98)$$

$$c_0 = 0, \quad c_1 = \frac{b}{2}, \quad (99)$$

Using (88)-(91), all the primal variables are assigned. It is important to note that x_0 and x_1 are non-zero in this setting because a_0 and a_1 are not zero, and similarly, d_0 and d_1 can be assumed to be non-zero stated in section IV-A. As a result, we are able to derive the stationary conditions (63) and (64).

b) Dual variable assignments:

$$\alpha_0 = \max\{0, Mz_0(a_0, a_1)\} \quad (100)$$

$$\beta_0 = \alpha_1 - \alpha_0 \quad (101)$$

$$\alpha_1 = Mz_1(a_0, a_1) \quad (102)$$

$$\beta_1 = 0 \quad (103)$$

$$\lambda = \alpha_1 \quad (104)$$

With (92)-(94), all the dual variables are assigned.

We wish to show that the assignments satisfy all the KKT conditions. With the primal and dual variables displayed above, one can conveniently confirm that (63)-(68), (69)-(75), (79)-(80), and (81)-(87) are satisfied by substituting these variables. It remains to demonstrate the validity of (76)-(78). It is evident that α_0 and β_1 are non-negative. Thus, we need only establish that $\alpha_1, \beta_0 \geq 0$.

For α_1 , we use the following partial derivative to show that z_1 is increasing in a_1 :

$$\frac{\partial z_1}{\partial a_1} = \frac{1 - \gamma p_0 + \gamma a_1}{(1 - a_1)a_1} \geq 0 \quad (105)$$

This immediately implies that $z_1(a_0 = p_0, a_1)$ is monotonically increasing. It suffices to show that $z_1(p_0, \frac{1}{2})$ is positive, which is the minimum value of $z_1(p_0, a_1)$ achieved in the range $a_1 \in [p_0, p_1]$. We write

$$\begin{aligned} z_1(p_0, \frac{1}{2}) &= \left[(-\gamma p_0) \log\left(\frac{1-p_0}{p_0}\right) + \gamma \log\left(\frac{1-p_0}{\frac{1}{2}}\right) \right] \\ &= \left[(1-p_0) \log(1-p_0) + p_0 \log(1-p_0) - \log \frac{1}{2} \right] \\ &= \gamma(-H(p_0) + H(\frac{1}{2})) \geq 0, \end{aligned} \quad (106)$$

from which we conclude that α_1 is non-negative since M is clearly positive. For β_0 ,

$$\beta_0 = \alpha_1 - \alpha_0 \quad (107)$$

$$= \begin{cases} -\log\left(\frac{1-a_1}{a_1}\right) + \log\left(\frac{1-a_0}{a_0}\right) & , z_0(a_0, a_1) \geq 0 \\ \alpha_1 & , z_0(a_0, a_1) < 0 \end{cases} \quad (108)$$

From equation (108), it is clear that $\beta_0 \geq 0$ when $z_0 < 0$ as $\alpha_1 \geq 0$. Similarly, for $z_0 \geq 0$, it follows that β_0 is non-negative since $a_1 \geq a_0 = p_0 \geq \frac{1}{2}$.

Now let us turn our attention to the case when $p_0 \geq p_1 \geq \frac{1}{2}$. We can determine the range of a_0 and a_1 as

$$a_0 = [p_1, p_0], \quad a_1 = p_1. \quad (109)$$

The variable assignments are as follows.

a) Primal variable assignments:

$$a_0 = \eta_0^-(b), \quad a_1 = p_1 \quad (110)$$

$$c_0 = \frac{b}{2}, \quad c_1 = 0 \quad (111)$$

Using (88)-(91), all the primal variables are assigned.

Observe that x_0 and x_1 will not be zero here as well.

b) Dual variable assignments:

$$\alpha_0 = Mz_0(a_0, a_1) \quad (112)$$

$$\beta_0 = 0 \quad (113)$$

$$\alpha_1 = Mz_1(a_0, a_1) \quad (114)$$

$$\beta_1 = \log\left(\frac{1-a_1}{a_1}\right) - \log\left(\frac{1-a_0}{a_0}\right) \quad (115)$$

$$\lambda = \alpha_0 \quad (116)$$

With (92)-(94), all the dual variables are assigned.

Same as the previous case, we wish to show that the assignments satisfy all the KKT conditions. With the primal and dual variables displayed above, one can conveniently confirm that (63)-(68), (69)-(75), (79)-(80), and (81)-(87) are satisfied by substituting these variables. It remains to demonstrate the validity of (76)-(78). It is evident that β_0 satisfies the non-negativity. Thus, we need only establish that $\alpha_0, \alpha_1, \beta_1 \geq 0$.

For α_0 , we use the following partial derivative to show that z_0 is increasing in a_0 :

$$\frac{\partial z_0}{\partial a_0} = \frac{1 + \gamma p_1 - \gamma a_0}{(1 - a_0)a_0} \geq 0 \quad (117)$$

(117) implies that $z_0(a_0, a_1 = p_1)$ is increasing in a_0 . It suffices to show that $z_0(p_1, p_1)$ is positive, which is the minimum value of $z_0(a_0, p_1)$ achieved in the range $a_0 \in [p_1, p_0]$.

$$z_0(p_1, p_1) = -\log\left(\frac{1-p_1}{p_1}\right) \geq 0 \quad (118)$$

For α_1 , it is demonstrated that $z_1(a_0, p_1)$ increases and remains non-negative at its lowest value $a_0 \in [\frac{1}{2}, p_0]$, as shown in (119) and (120).

$$\frac{\partial z_1}{\partial a_0} = \frac{\gamma p_0 - \gamma a_0}{(1 - a_0)a_0} \geq 0, \quad \forall a_0 \in [p_1, p_0] \quad (119)$$

$$\begin{aligned} z_1(\frac{1}{2}, p_1) &= (-1 + \gamma p_0) \log\left(\frac{1-p_1}{p_1}\right) \\ &\quad + \gamma \log\left(\frac{\frac{1}{2}}{1-p_1}\right) \geq 0 \end{aligned} \quad (120)$$

Furthermore, it is evident that β_1 is greater than zero since $a_0 \geq a_1 = p_1 \geq \frac{1}{2}$. This establishes that $\alpha_0, \alpha_1, \beta_1, \lambda$ are all non-negative.

2) $b \in [b_l, b_h]$: In this regime, we have the following assignments.

a) Primal variable assignments:

$$a_0 = a_1 = -M\left(\frac{b}{2} - \mathbf{d}_0^\gamma p_0 - \mathbf{d}_1^\gamma p_1\right) \quad (121)$$

$$c_0 = d_0(p_0 - a_0), \quad c_1 = d_1(p_1 - a_1). \quad (122)$$

Using (88)-(91), all the primal variables are assigned. Observe that x_0 and x_1 will not be zero here, indicating that stationary conditions can be derived.

b) Dual variable assignments:

$$\begin{cases} \alpha_0 = \alpha_1 = M \log \left(\frac{a_0}{1-a_0} \right) \\ \beta_0 = \beta_1 = 0 \\ \lambda = \alpha_1 \end{cases} \quad (123)$$

With (92)-(94), all the dual variables are assigned.

Similar to the previous discussion, we wish to show that the assignments satisfy all the KKT conditions. With the primal and dual variables displayed above, one can conveniently confirm that (63)-(68), (69)-(75), (78)-(80), and (81)-(87) are satisfied by substituting these variables. Since $a_0 \geq \frac{1}{2}$, we have that α_s, λ are non-negative, leading to (76) and (77) in this case.

3) $b \in [b_h, \infty)$: In this regime, we have

a) Primal variables:

$$a_0 = a_1 = \frac{1}{2} \quad (124)$$

$$c_0 = d_0(p_0 - \frac{1}{2}), \quad c_1 = d_1(p_1 - \frac{1}{2}). \quad (125)$$

Using (88)-(91), all the primal variables are assigned.

Observe that x_0 and x_1 will not be zero here as well.

b) Dual variables:

$$\alpha_0 = \alpha_1 = \beta_0 = \beta_1 = \rho_0 = \rho_1 = \lambda = 0. \quad (126)$$

All the dual variables are assigned using (92)-(94).

With the primal and dual variables displayed above, one can conveniently confirm that (63)-(68), (69)-(75), (76)-(80), and (81)-(87) are satisfied by substituting these variables.

B. Proof of Theorem 3

Theorem 3 holds under the assumption $p_0 < \frac{1}{2} \leq p_1$. There are again three regimes:

1) The regime $b \in [0, b'_l]$: In this regime, the two cases $|\frac{1}{2} - p_1| \geq |\frac{1}{2} - p_0|$ and $|\frac{1}{2} - p_1| < |\frac{1}{2} - p_0|$ lead to different solution expressions, and let us consider the first case for now. Observe that in the first case implies that $p_1 + p_0 - 1 \geq 0$, b'_l in this case can be written as:

$$b'_l = 2 \left(\frac{1 - \gamma p_0 - d_0^\gamma}{1 - \gamma p_0 + \gamma \psi_1} \right) (p_1 - \psi_1) \quad (127)$$

by the definition given in (43).

With the solution of a_0, a_1 given in the theorem, and in the given range of b , the range of a_0 and a_1 are

$$a_0 = p_0, \quad a_1 \in [\psi_1, p_1] \subset [\frac{1}{2}, p_1] \quad (128)$$

The variable assignments are as follows.

a) Primal variable assignments:

$$a_0 = p_0, \quad a_1 = \eta_1^-(b) \quad (129)$$

$$c_0 = 0, \quad c_1 = \frac{b}{2} \quad (130)$$

Using (88)-(91), all the primal variables are assigned. Note that in this scenario, x_0 and x_1 are non-zero,

enabling us to address the stationarity conditions in (63) and (64).

b) Dual variable assignments:

$$\alpha_0 = z_1(a_0, a_1) + z_0(a_0, a_1) = -M\Psi_{p_0}(a_1) \quad (131)$$

$$\alpha_1 = z_1(a_0, a_1) \quad (132)$$

$$\beta_0 = -z_0(a_0, a_1) \quad (133)$$

$$\beta_1 = 0 \quad (134)$$

$$\lambda = \alpha_1 \quad (135)$$

With (92)-(94), all the dual variables are assigned.

We wish to show that the assignments satisfy all the KKT conditions. With the primal and dual variables displayed above, one can conveniently confirm that (63)-(68), (69)-(75), (79)-(80), and (81)-(87) are satisfied by substituting these variables. It remains to verify (76)-(78). It is evident that β_1 is non-negative. Thus, we need only establish that $\alpha_0, \alpha_1, \beta_0 \geq 0$.

(105) shows that z_1 is monotonically increasing in a_1 , while (106) confirmed that the minimum value of z_1 is non-negative, occurring when $a_1 = \frac{1}{2}$ within the interval $a_1 \in [\psi_1, p_1] \subset [\frac{1}{2}, p_1]$. Therefore, indeed $\alpha_1 \geq 0$.

For β_0 , the following partial derivative shows that $-z_0$ is monotonically non-decreasing in a_1 :

$$\frac{\partial z_0}{\partial a_1} = \frac{-\gamma p_1 + \gamma a_1}{(1 - a_1)a_1} \leq 0, \quad \forall a_1 \in [\psi_1, p_1] \quad (136)$$

It suffices to show that $-z_0(p_0, a_1 = \frac{1}{2})$ is positive, which is the minimum value of $-z_0(p_0, a_1)$ achieved in the range $a_1 \in [\frac{1}{2}, p_1]$. We write

$$\begin{aligned} & -z_0(p_0, \frac{1}{2}) \\ &= (1 + \gamma p_1) \log \left(\frac{1 - p_0}{p_0} \right) - \gamma \log \left(\frac{1 - a_0}{\frac{1}{2}} \right) \end{aligned} \quad (137)$$

$$\begin{aligned} &= \gamma \left(\log \left(\frac{1 - p_0}{p_1} \right) - \log \left(\frac{1 - a_0}{\frac{1}{2}} \right) \right) \\ &+ (1 - \gamma + \gamma p_1) \log \left(\frac{1 - p_0}{p_0} \right) \geq 0. \end{aligned} \quad (138)$$

Therefore, $\beta_0 \geq 0$.

To demonstrate $\alpha_0 \geq 0$, it is necessary to establish that $-M\Psi_{p_0}(a_1)$ is non-negative. Employing a similar approach as earlier, we first show that $-M\Psi_{p_0}(a_1)$ increases when $a_1 \in [\psi_1, p_1]$ by way of (139). Then (140) confirms that its minimum value is zero, occurring at $a_1 = \psi_1$.

$$\begin{aligned} -\frac{\partial}{\partial a_1} M\Psi_{p_0}(a_1) &= -M\Psi'_{p_0}(a_1) = M\Phi'_-(a_1) \\ &= \frac{2\gamma a_1 + 1 - \gamma p_0 - \gamma p_1}{(1 - a_1)a_1} \geq 0 \end{aligned} \quad (139)$$

$$-M\Psi_{p_0}(\psi_1) = -M(\Phi_+(p_0) - \Phi_-(\psi_1)) = 0 \quad (140)$$

The inequality in (139) follows from the fact that $a_1 \geq \frac{1}{2}$. The above establishes that $\alpha_0, \alpha_1, \beta_0, \lambda$ are all non-negative.

Now let us turn our attention to the case when $|\frac{1}{2} - p_1| < |\frac{1}{2} - p_0|$, implying $1 - p_1 - p_0 \geq 0$. b'_l in this case can be written as:

$$b'_l = 2 \left(\frac{d_0^\gamma + \gamma p_1}{1 - \gamma \psi_0 + \gamma p_1} \right) (\psi_0 - p_0) \quad (141)$$

again by the definition given in (43).

We can determine the range of a_0 and a_1 as

$$a_0 \in [p_0, \psi_0], \quad a_1 = p_1. \quad (142)$$

The variable assignments are as follows.

a) Primal variable assignments:

$$a_0 = \eta_0^+(b), \quad a_1 = p_1 \quad (143)$$

$$c_0 = \frac{b}{2}, \quad c_1 = 0 \quad (144)$$

Using (88)-(91), all the primal variables are assigned. Note that x_0 and x_1 are also non-zero in this context, allowing us to derive stationary conditions (63) and (64).

b) Dual variable assignments:

$$\alpha_0 = 0 \quad (145)$$

$$\alpha_1 = z_1(a_0, a_1) \quad (146)$$

$$\beta_0 = -z_0(a_0, a_1) \quad (147)$$

$$\beta_1 = -z_0(a_0, a_1) - z_1(a_0, a_1) = M\Psi_{p_1}(a_0) \quad (148)$$

$$\lambda = \beta_0 \quad (149)$$

With (92)-(94), all the dual variables are assigned.

Same as in the previous case, we wish to show that the assignments satisfy all the KKT conditions. With the primal and dual variables displayed above, one can conveniently confirm that (63)-(68), (69)-(75), (79)-(80), and (81)-(87) are satisfied by substituting these variables. It remains to verify (76)-(78). It is evident that α_0 satisfies the non-negativity condition. Thus, we need only establish that $\alpha_1, \beta_0, \beta_1 \geq 0$.

For α_1 , (119) showed that z_1 is monotonically non-increasing when $a_0 \in [p_0, \psi_0]$. Moreover, (120) implies that $z_1(\frac{1}{2}, p_1)$ is positive, which is the minimum value of $z_1(a_0, p_1)$ achieved in the range $a_0 \in [p_0, \psi_0] \subset [p_0, \frac{1}{2}]$. Therefore α_1 is non-negative.

For β_0 , (117) showed that $z_0(a_0, p_1)$ is monotonically non-decreasing in a_0 , implying that β_0 is decreasing in a_0 . Moreover, we can write

$$\begin{aligned} -z_0\left(\frac{1}{2}, p_1\right) &= -\gamma p_1 \log\left(\frac{1-p_1}{p_1}\right) - \gamma \log\left(\frac{\frac{1}{2}}{1-p_1}\right) \\ &= \gamma \left(H\left(\frac{1}{2}\right) - H(p_1) \right) \geq 0, \end{aligned} \quad (150)$$

which implies that the minimum value of $-z_0(a_0, p_1)$ in the range $a_0 \in [p_0, \psi_0] \subset [p_0, \frac{1}{2}]$ is non-negative. Combining these two facts, we conclude that β_0 is also non-negative.

For β_1 , a similar procedure can be used to show its non-negativity. (151) implies that $M\Psi_{p_1}(a_0)$ is monotonically

non-increasing in $a_0 \in [p_0, \psi_0]$, given that $a_0 \leq \frac{1}{2}$. Furthermore, as shown in (152), the minimum value of $M\Psi_{p_1}(a_0)$ over the interval $a_0 \in [p_0, \psi_0]$ is zero, which is derived from the definition of ψ_0 in Theorem 3.

$$\begin{aligned} \frac{\partial}{\partial a_0} M\Psi_{p_1}(a_0) &= M\Phi'_+(a_0) \\ &= \frac{2\gamma a_0 - 1 - \gamma p_0 - \gamma p_1}{(1-a_0)a_0} \leq 0, \\ \forall a_0 \in [p_0, \psi_0] &\subset [p_0, \frac{1}{2}] \end{aligned} \quad (151)$$

$$M\Psi_{p_1}(\psi_0) = M(\Phi_+(\psi_0) - \Phi_-(a_1)) = 0 \quad (152)$$

Therefore (76)-(78) hold in this case.

2) $b \in [b'_l, b'_h]$: Let us first define the following set

$$\begin{aligned} S_b \triangleq & \left\{ (a_0, a_1) \in [p_0, \frac{1}{2}] \times [\frac{1}{2}, p_1] \mid \right. \\ & \left. \Phi_-(a_1) = \Phi_+(a_0) \text{ and } m(a_0, a_1) = \frac{b}{2} \right\} \end{aligned} \quad (153)$$

It will be shown shortly that the set S_b is a singleton set in this regime. We have the following assignments.

a) Primal variable assignments:

$$(a_0, a_1) \in S_b \quad (154)$$

$$c_0 = d_0(a_0 - p_0), \quad c_1 = d_1(p_1 - a_1) \quad (155)$$

Using (88)-(91), all the primal variables are assigned. Note that x_0 and x_1 will not be zero here as well, thus allowing the derivation of stationary conditions.

b) Dual variable assignments:

$$\alpha_0 = 0 \quad (156)$$

$$\alpha_1 = \frac{1}{2} \log\left(\frac{1-a_0}{a_0}\right) - \frac{1}{2} \log\left(\frac{1-a_1}{a_1}\right) \quad (157)$$

$$\beta_0 = \alpha_1 \quad (158)$$

$$\beta_1 = 0 \quad (159)$$

$$\lambda = \alpha_1 \quad (160)$$

With (92)-(94), all the dual variables are assigned. We can confirm that (63)-(68), (69)-(75), (79)-(80), and (81)-(87) are satisfied by substituting these variables. According to the definition of S_b , if (a_0, a_1) is an element in S_b , then it follows that $a_0 \leq \frac{1}{2} \leq a_1$. Clearly, the conditions (76)-(78) remain valid since $a_0 \leq \frac{1}{2} \leq a_1$.

3) $b \in (b_h, \infty)$: In this regime, we have

a) Primal variables:

$$a_0 = a_1 = \frac{1}{2} \quad (161)$$

$$c_0 = d_0\left(\frac{1}{2} - p_0\right), \quad c_1 = d_1\left(p_1 - \frac{1}{2}\right). \quad (162)$$

Using (88)-(91), all the primal variables are assigned. Observe that x_0 and x_1 will not be zero here as well.

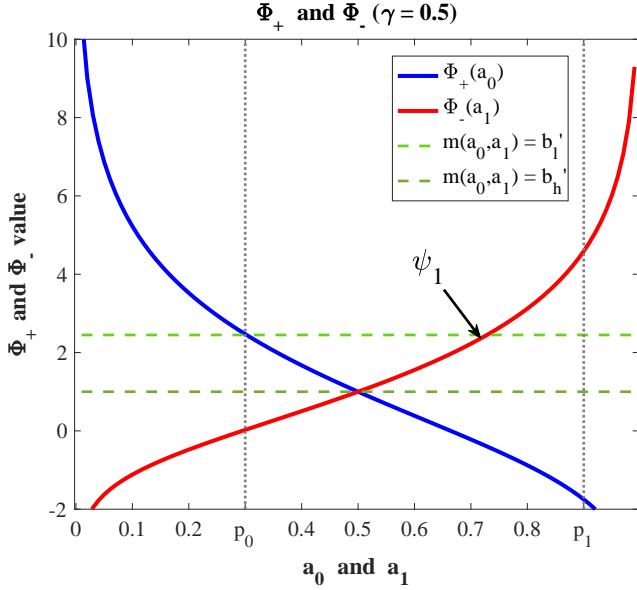


Fig. 6: The proof in appendix B-C

b) Dual variables:

$$\alpha_0 = \alpha_1 = \beta_0 = \beta_1 = \rho_0 = \rho_1 = \lambda = 0. \quad (163)$$

All the dual variables are assigned using (92)-(94).

With the primal and dual variables displayed above, we confirm that (63)-(68), (69)-(75), (76)-(80), and (81)-(87) are satisfied by substituting these variables.

C. Proof That S_b Is a Singleton Set

Let us first consider the two functions $\Phi_+(a_0)$ and $\Phi_-(a_1)$, which are shown in Fig.6. We can see that the two functions intersect at the point $a_0 = a_1 = \frac{1}{2}$. Their intersection with the darker green dash line is $(a_0, a_1) = (\frac{1}{2}, \frac{1}{2})$, which is the only element in $S_{b'_l}$; here the two functions take the same value. Similarly, their intersections with the lighter green dash line give $(a_0, a_1) = (p_0, \psi_1)$, which is the only element in $S_{b'_h}$; again the two functions take the same value. The plot illustrates that any horizontal line positioned between the two green dashed lines intersects with these two functions, and the intersection points represent the unique element in S_b for $b \in [b'_l, b'_h]$. Given this behavior, it suffices to prove the following two claims.

1) *Claim 1: Φ_+ and Φ_- are monotonic functions that increase in opposing directions in $(0, \frac{1}{2}]$ and $[\frac{1}{2}, 1)$ respectively. In other words, Φ_+ decreasing in $a \in (0, \frac{1}{2}]$ and Φ_- increasing in $a \in [\frac{1}{2}, 1)$. We prove this claim by computing the derivatives.*

$$\begin{aligned} \frac{\partial \Phi_+}{\partial a} &= \frac{-(1 - 2\gamma + \gamma p_0 + \gamma p_1)}{1 - a} - \frac{(1 + \gamma p_0 + \gamma p_1)}{a} \\ &= \frac{2\gamma a - 1 - \gamma p_0 - \gamma p_1}{(1 - a)a} \leq 0, \quad \forall a \in (0, \frac{1}{2}) \quad (164) \\ \frac{\partial \Phi_-}{\partial a} &= \frac{-(-1 - 2\gamma + \gamma p_0 + \gamma p_1)}{1 - a} \end{aligned}$$

$$= \frac{(-1 + \gamma p_0 + \gamma p_1)}{a} = \frac{2\gamma a + 1 - \gamma p_0 - \gamma p_1}{(1 - a)a} \geq 0, \quad \forall a \in [\frac{1}{2}, 1) \quad (165)$$

Moreover, as already mention, Φ_+ and Φ_- intersects at the point $\frac{1}{2}$, i.e., $\Phi_+(\frac{1}{2}) = \Phi_-(\frac{1}{2}) = -2\gamma \log \frac{1}{2}$.

2) *Claim 2: There is only one solution $(a_0, a_1) \in [p_0, \frac{1}{2}] \times [\frac{1}{2}, p_1]$ satisfying $m(a_0, a_1) = \frac{b}{2}$ for b in the range $[b'_l, b'_h]$ specified in Theorem 3. We next prove this claim. In (166) and (167) given below, we show that for different cases $|\frac{1}{2} - p_1| \geq |\frac{1}{2} - p_0|$ and $|\frac{1}{2} - p_0| > |\frac{1}{2} - p_1|$, (ψ_0, p_1) and (p_0, ψ_1) are the solutions to $m(a_0, a_1) = \frac{b}{2}$:*

$$\text{If } |\frac{1}{2} - p_1| \geq |\frac{1}{2} - p_0|,$$

$$\begin{aligned} d_0(p_0 - p_0) + d_1(p_1 - \psi_1) &= d_1(p_1 - \psi_1) \\ &= \left(\frac{1 - \gamma p_0 - \mathbf{d}_0^\gamma}{1 - \gamma p_0 + \gamma \psi_1} \right) (p_1 - \psi_1) = \frac{b'_l}{2} \quad (166) \end{aligned}$$

$$\text{If } |\frac{1}{2} - p_0| > |\frac{1}{2} - p_1|,$$

$$\begin{aligned} d_0(\psi_0 - p_0) + d_1(p_1 - p_1) &= d_0(\psi_0 - p_0) \\ &= \left(\frac{\mathbf{d}_0^\gamma + \gamma p_1}{1 - \gamma \psi_0 + \gamma p_1} \right) (\psi_0 - p_0) = \frac{b'_l}{2} \quad (167) \end{aligned}$$

Moreover, from equation (168) below, we see that $(\frac{1}{2}, \frac{1}{2})$ is the solution to $m(a_0, a_1) = \frac{b'_h}{2}$.

$$\begin{aligned} d_0(\frac{1}{2} - p_0) + d_1(p_1 - \frac{1}{2}) &= \frac{1}{2}(d_0 - d_1) - d_0 p_0 + d_1 p_1 \\ &= \frac{1}{2} \left(\frac{\mathbf{d}_0^\gamma + \frac{1}{2}\gamma - 1 + \frac{1}{2}\gamma + \mathbf{d}_0^\gamma}{1 - \frac{1}{2}\gamma + \frac{1}{2}\gamma} \right) \\ &\quad - p_0 \left(\frac{\mathbf{d}_0^\gamma + \frac{1}{2}\gamma}{1 - \frac{1}{2}\gamma + \frac{1}{2}\gamma} \right) + p_1 \left(\frac{1 - \frac{1}{2}\gamma - \mathbf{d}_0^\gamma}{1 - \frac{1}{2}\gamma + \frac{1}{2}\gamma} \right) \\ &= \left(\mathbf{d}_0^\gamma + \frac{\gamma}{2} \right) (1 - p_0 - p_1) + p_1 - \frac{1}{2} = \frac{b'_h}{2} \quad (168) \end{aligned}$$

As a consequence, when $b = b'_l$ and $b = b'_h$, the solutions to the equation are situated within the interval $[p_0, \frac{1}{2}] \times [\frac{1}{2}, p_1]$. Since $\Phi_+(a_0)$ and $\Phi_-(a_1)$ are monotonic, and the mapping $m(a_0, a_1)$ is continuous, we conclude that there is a unique solution (a_0, a_1) to the equations $\Phi_+(a_0) = \Phi_-(a_1)$ and $m(a_0, a_1) = \frac{b}{2}$ for each $b \in [b'_l, b'_h]$.

APPENDIX C PROOF OF REDUCTION OF CASES

In scenarios where $p_0, p_1 \leq \frac{1}{2}$ and $p_0 \geq \frac{1}{2} \geq p_1$, we interchange state 0 to 1' and state 1 to 0' as the order of states is inconsequential. We first note that in our setting, $p_0 = p_{0 \rightarrow 0}$ and $p_1 = p_{1 \rightarrow 0}$, meaning that p_0 is the probability from state 0 to 0 and p_1 is the probability from state 1 to 0. After renaming the state from 0, 1 to 1', 0', the probability p_0, p_1 now becomes

$$p_0 = p_{0 \rightarrow 0} \stackrel{\text{rename}}{=} p_{1' \rightarrow 1'} \triangleq 1 - p'_1 \quad (169)$$

$$p_1 = p_{1 \rightarrow 0} \stackrel{\text{rename}}{=} p_{0' \rightarrow 1'} \triangleq 1 - p'_0 \quad (170)$$

Therefore, for case $p_0, p_1 \leq \frac{1}{2}$, the renamed optimization problem will have $p'_0, p'_1 \geq \frac{1}{2}$. Use the results in Theorem 2 to solve for optimal a'_0, a'_1 , then convert back to a_0, a_1 .

$$a'_0 = a_{0' \rightarrow 0'} \stackrel{\text{recover}}{=} a_{1 \rightarrow 1} = 1 - a_1 \Rightarrow a_1 = 1 - a'_0 \quad (171)$$

$$a'_1 = a_{1' \rightarrow 0'} \stackrel{\text{recover}}{=} a_{0 \rightarrow 1} = 1 - a_0 \Rightarrow a_0 = 1 - a'_1 \quad (172)$$

Similar steps for case $p_0 \geq \frac{1}{2} \geq p_1$.