Function Optimization-based Schemes for Designing Continuous Action Learning Automata

by

Haoye Lu

Thesis submitted in partial
fulfillment of the requirements for the
Master of Computer Science degree

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

The field of Learning Automata (LA) has been studied and analyzed extensively for more than four decades; however, almost all the papers have concentrated on the LA working in Environments that have a finite number of actions. This is a well-established model of computation, and expedient, ϵ-optimal and absolutely expedient machines have been designed for stationary and non-stationary Environments. There are only a few papers which deal with Environments possessing an infinite number of actions. These papers assume a well-defined and rather simple uni-modal functional form, like the Gaussian function, for the Environment's infinite reward probabilities.

This thesis pioneers the concept and presents a series of continuous action LA (CALA) algorithms that do not require the function of the Environment's infinite reward probabilities to obey a well-established uni-modal functional form. Instead, this function can be, but not limited to, a multi-modal function as long as it satisfies some weak constraints. Moreover, as our discussion evolves, the constraints are further relaxed. In all these cases, we demonstrate that the underlying machines converge in an ϵ-optimal manner to the optimal action of an infinite action set. Based on the CALA algorithms proposed, we report a global maximum search algorithm, which can find the maximum points of a real-valued function by sampling the function's values that could be contaminated by noise.

This thesis also investigates the performance limit of the action-taking scheme, sampling actions based on probability density functions, which is used by all currently available CALA algorithms. In more details, given a reward function, we define an index of the function which is the least upper bound of the performance that a CALA algorithm can possibly achieve. Besides, we also report a CALA algorithm that meets this upper bound in an ϵ-optimal manner.

By investigating the problem from a different perspective, we argue that the algorithms proposed are closely related to the family of "Stochastic Point Location" problems involving either discretized steps or d-ary parallel machines. The thesis includes the detailed proofs of the assertions and highlights the niche contributions within the broader theory of learning. To the best of our knowledge, there are no comparable results reported in the literature.

Acknowledgements

My most sincere gratitude to Prof. Amiya Nayak for providing me guidance, support, advice and ideas throughout my thesis work. I wholeheartedly thank him for his precious supervisions.

My deepest appreciations to Prof. John Oommen for verifying my work and providing significant suggestions.

Besides, I would also like to thank my family and friends for keeping my motivation high and supporting me throughout.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Finding close relations among the results that were previously considered unrelated, would be the most fascinating fruits of research. Such analogs have appeared in the fundamental sciences for centuries, for example, in Physics, when it concerns wave theory and light. Without belaboring this point from a philosophic perspective, we submit a similar thesis within the field of computerized Learning Theory.

Researchers have studied, analyzed and worked within the field of Learning Automata (LA) for many years. Apart from schemes that have a continuous value space of action-taking probability, significant research has also been done on those with a discretized space. Further, while most of the results within the topic of LA have concentrated on Environments possessing a finite number of actions, a few pioneering papers [18, 19, 46, 48, 55] boast contributions when the number of actions is infinite. All of them, categorically assume a well-defined and rather simple uni-modal functional form, like the Gaussian function, for the Environment's infinite reward probabilities. From a third, but completely different perspective, there has been a lot of work done within the family of so-called "Stochastic Point Location" (SPL) problems [20, 36, 41, 42, 52, 61, 62]. The proposed solutions to the SPL are either discretized or work with d-ary parallel ma-

chines. This papers attempts to tie all the consequent loose ends, and proposes a single theoretical foundation that knits them together.

Firstly, our concern is the field of LA in which the Environment has infinite actions [18, 19, 46, 48, 55]. However, secondly, within this domain, unlike the prior art, we pioneer the concept in which the uni-modal functional form for the Environment's infinite reward probabilities, does not obey a well-established form. Thirdly, as our theory does not have direct constraints on the shape of this function, the potential functional forms include, but are not limited to, multi-modal functions. Fourthly, all of our solutions further operate by moving in the search domain in a step-oriented manner, thus also embracing the field of Discretized LA [24, 25, 35, 37, 38, 64]. In a nutshell, by putting them all together, we propose solutions to LA problems that deal with an infinite number of actions, where the functional form for the infinite reward probabilities are arbitrarily general and the proposed solution moves within the solution space in a discretized manner, as the one does in the SPL. Thus, apart from the results that we present being novel and pioneering, they are also extremely captivating and fascinating.

This introductory section *briefly* presents the existing literature about the above-mentioned distinct areas, and then addresses the problem at hand, and our contributions.

1.1 Learning Automata

We first briefly survey the well-trodden field of LA. For more than fifty years[1], this field, initiated by Tsetlin [56], has been studied as a typical model for learning in random Environments. This domain has actually served as the precursor for the area of Reinforcement Learning. Unlike other fields of Artificial Intelligence (AI), like game playing,

[1] In the interest of brevity, we assume that the reader is fairly well-versed in the fundamental concepts of LA and their convergence properties. The review here is thus, though still comprehensive, necessarily brief. That being said, excellent surveys of the field can be found in [22, 30, 31, 33, 45, 55].

problem solving, natural language processing etc., an LA, by definition, operates in a *random* medium. Consequently, the "Teacher" (formally refereed to as the "Environment") can respond differently and stochastically, for the same query, at different time instances.

The basic tenets of the field of LA are the following:

- The LA interacts with a random Environment that offers it a set of actions;
- The LA adaptively attempts to learn the optimal action offered by the Environment;
- To render the problem non-trivial, the Environment is stochastic;
- At each iteration (or time step), the LA selects one action and communicates it to the Environment;
- This, in turn, *stochastically* triggers either a reward or a penalty from the Environment;
- Based on the "history" and this response, the LA adjusts its action selection strategy;
- The goal is that the LA adapts itself, or "converges", to the optimal action.

More specifically, the LA uses the knowledge acquired in the past iterations, and changes its state either deterministically or stochastically in order to make a "wiser" decision in the next iteration. In this manner, the LA, though lacking a complete knowledge about the Environment, is able to learn its optimal action by repeatedly interacting with it.

FSSA and VSSA: Initial LA were designed to be Fixed Structure Stochastic Automata (FSSA). Their state update and decision functions are time invariant [31, 56], as

in the case of the Tsetlin and Krinsky machine. Variable Structure Stochastic Automata (VSSA) were initially characterized by *continuous functions* that update the probability of selecting the various actions. These permit the action probabilities to take any value in $[0, 1]$, and have been extensively designed and analyzed. Representatives of VSSA include the Linear Reward-Penalty (L_{R-P}) scheme, the Linear Reward-Inaction (L_{R-I}) scheme, the Linear Inaction-Penalty (L_{I-P}) scheme and the Linear Reward-ϵPenalty ($L_{R-\epsilon P}$) scheme [22, 31], all of which use linear updating functions. As a general principle, the L_{R-I} and $L_{R-\epsilon P}$ schemes assign more importance to reward responses than to penalties; they are also ϵ-optimal in all stationary environments. This is also the case with FSSA, where, for example, the Krinsky automaton, which treats rewards significantly "more seriously" than penalties, is ϵ-optimal in all stationary environments, while the Tsetlin automaton, which treats rewards and penalties equally, is only ϵ-optimal when the largest reward probability is greater than 0.5 [31]. VSSA schemes which invoke nonlinear functions have also been designed and analyzed [22, 23, 31].

Applications of LA: With regard to applications, the entire field of LA and stochastic learning has had a myriad of applications listed in [22, 30, 31, 45, 55]. Besides, the applications also include solutions for problems in network and communications [29, 34, 44], network call admission, traffic control, quality of service routing [3, 4, 59], distributed scheduling [49], training hidden Markov models [21], neural network adaptation [28], intelligent vehicle control [57, 58], service selection [60] and even fairly theoretical problems such as graph partitioning [39] and string taxonomy [40]. Besides these fairly generic applications, with a little insight, LA can be used to assist in solving (by, indeed, learning the associated parameters) the stochastic resonance problem [11], the stochastic sampling problem in computer graphics [12], the problem of determining roads in aerial images by using geometric-stochastic models [6], and various location problems [7]. Similar

learning solutions can also be used to analyze the stochastic properties of the random waypoint mobility model in wireless communication networks [9], to achieve spatial point pattern analysis codes for GISs [47], to digitally simulate wind field velocities [43], to interrogate the experimental measurements of global dynamics in magneto-mechanical oscillators [13], and to analyze spatial point patterns [5].

1.2 Contributions

In the context of what we have explained above, the main contributions of this thesis are as follows:

1. We have devised a series of LA-based schemes in which the action space is continuous and thus infinite. All the Continuous Action Learning Automaton (CALA) schemes we have proposed adopt a linear search-based strategy for searching optimal actions and thus are named as the Linear Search-Based CALA (LSCALA) algorithms.

2. Unlike the state-of-the-art, we have shown that, for all the schemes proposed, the functional form for the Environment's infinite reward probabilities, if uni-modal, does not need to obey a well-established functional form.

3. Unlike the state-of-the-art, we have shown that the functional form for the Environment's infinite reward probabilities can be, but not limited to, multimodal.

4. Based on the LSCALA scheme, we have also devised an algorithm that can find the global maximum of a real-valued function by sampling its noisy function values.

5. For a reward function ϕ, we have defined an associated index, denoted by $\tilde{\phi}$, referred to as the Optimal Reward Probability (ORP) of ϕ. We have shown that $\tilde{\phi}$ is the

least upper bound of reward probability that a CALA, taking actions sampled from the actions' probability distribution, can possibly achieve. Besides, we have reported an LSCALA algorithm whose long-term reward probability converges to it in an ϵ-optimal manner.

6. We have proven that all the LSCALA algorithms reported are ϵ-optimal under some weak assumptions on the reward functions. Besides, as the discussion evolves, we have shown that those assumptions can be further relaxed. Compared to the assumptions required by the currently-available algorithms, our assumptions are much more general and can, in practice, be easily satisfied. To the best of our knowledge, these algorithms are the first a few CALA-based algorithms that have been proven to be ϵ-optimal in Environments of the type listed above.

7. To support all these claims, we have provided simulation results to corroborate our theoretical work.

8. Finally, we have provided a single framework which merges three distinct phenomena, namely that of working with infinite actions, working within a "discretized" paradigm, and relating these to the SPL problem.

To the best of our knowledge, all of these results are novel and pioneering.

1.3 Outline of the Thesis

The rest of the thesis is organized as follows:

- In Chapter 2, we briefly survey the respective fields of LA with a large number or even an infinite number of actions, "Discretization", and the SPL.

- Chapter 3 introduces the original LSCALA algorithm. We discuss the idea of "Discretization" behind the algorithm design as well as the relation between the algorithm and the SPL problem. We also prove that the algorithm is ϵ-optimal if the reward function is strictly positive and Lipschitz continuous. Besides, we provide some experimental results to corroborate our analysis.

- In Chapter 4, we show that the Lipschitz continuity of the reward function is unnecessary for the ϵ-optimality of the LSCALA algorithm introduced in Chapter 3. Instead, we prove that the continuity of the function near its global maximum points is sufficient. Besides, by slightly modifying the original algorithm, we show that we can get rid of the condition that the reward function is strictly positive. Some experimental results are provided at the end of the chapter to support our analysis results.

- Chapter 5 reports a global maximum search method based on the original LSCALA algorithm. The method can find the global maximum points of a real valued function according to the sampled function values that may be contaminated by noise.

- Chapter 6 discusses the performance limit of the LA algorithms that take actions according some continuous probability density functions. This strategy have been used by all currently-available CALA algorithms. Given a reward function ϕ, an index $\tilde{\phi}$, named the Optimal Reward Probability (ORP), is defined. We have proved that $\tilde{\phi}$ is the least upper bound of the long term reward probability that any CALA algorithms can possibly achieve. We have also reported two LSCALA algorithms whose long-term reward probabilities meet $\tilde{\phi}$ in an ϵ-optimal manner. Similar to the previous chapters, we also have provided necessary experimental

results to demonstrate the validity of our claims.

- Chapter 7 concludes the thesis and includes some avenues for future work.

Chapter 2

State-of-the-Art in the Relevant Fields

As we have previously stated, our goal is to present a new LA scheme for a continuous action space. Our solution is based on the idea of discretization, which is also related to the SPL problem. To place these in perspective, it is prudent for us to briefly catalog the state-of-the-art in these three areas.

2.1 Designing LA when R is Large or Infinite

The first paradigm, which is central to this paper is that of designing LA when R is large or infinite. The literature here deals with two distinct cases, i.e., when the number of actions is large or even countably infinite, and then when the action space is continuous. We consider each of them separately.

LA for a Large Number of Actions: Designing LA when the number of actions involved, R, is large is complex, for the following reasons:

1. In the case of FSSA, one requires N-states for each of the R actions. As the

Environment responds, the LA, for the most part (i.e., except at the so-called boundary states) moves within the states of a single action, and it can take a large number of iterations (for example, in the Krinsky LA) for the machine to even *enter* the boundary state of another action. Before all the actions are even visited, in certain environments, it could take tens of thousands of iterations for all the actions to be visited.

2. In the case of FSSA, since the machines are almost always ergodic, the Markov chain lingers in its transient behavior for a long time before convergence. Thus, when the number of actions is large, one deals with an $R \cdot N \times R \cdot N$-sized Markov chain, and this adds to the sluggishness of the machine.

3. In the case of VSSA, the above two concerns are mitigated by the use of the action probability vector. This has noticeable advantages, when R is relatively small, for example, of the order of 10. In this case, the action probability vector has the dimension R, and all the actions have a reasonable probability of being chosen. This permits the LA to discriminate between the various actions, and to converge to the superior one. However, when R is large, many of the action probabilities can have very small values and may not even be chosen, thus rendering the principle motivating VSSA to be void.

4. In the context of VSSA, for example, in linear schemes, the probabilities which are decreased are multiplied by a constant. Thus for any R, typically $R - 1$ of these probabilities may have to be decreased. Notice that when R is large, the decrement of these $R - 1$ probabilities can make a "non-small-step" change in the probability that is being increased. This will, consequently, significantly hinder the convergence of the machine, inasmuch as all the convergence proofs depend on

the theory of "small-step" random processes. The same assertion is valid for the discretized families of LA.

5. The families of continuous and discrete pursuit algorithms, described above, are universally accepted to be the fastest reported LA because they augment the action probability vector with a vector of the estimates of the reward probabilities. When R is large, this poses a problem of a disproportionate magnitude, because all the R actions have to be sampled a reasonably large number of times so that the inferior actions can be filtered out. Thus, pursuit LA are also quite sluggish when R is large.

LA for a Continuous Action Set: Within this family, two Continuous Action Learning Automata (CALA) algorithms have been reported, both of which were devised in the 1990's.

1. The first algorithm was due to Santharam *et al.*, and its convergence was proven under some strict assumptions [48]. In particular, if the reward function ϕ is continuously differentiable and its derivative is Lipschitz, the final result converges to a point arbitrarily close to a local maximum of ϕ by forcing some parameters to the limit (zero or infinity) [55].

2. An alternative implementation of CALA is the Continuous Action Reinforcement Learning Automaton (CARLA), introduced by Howell *et al.* [18, 19]. Rodrguez *et al.* reported an improvement of the CARLA method in terms of the computation effort required and the local convergence properties [46]. However, they confessed that their analysis was made under really strict assumptions, which confines the application of the algorithm considerably to a narrow scope.

The solution that we propose in this paper, resolves all of these issues.

2.2 Concept of Discretization

The second paradigm, whose properties we intend to capture, deals with the phenomenon of discretization. In practice, the relatively slow rate of convergence of continuous LA constituted a limiting factor in their applicability. In order to increase their speed of convergence, the concept of discretizing the probability space was introduced in [35, 53]. This concept is implemented by restricting the probability of choosing an action to be one of a finite number of values in $[0, 1]$. If the values allowed are equally spaced in this interval, the discretization is said to be *linear*, otherwise, it is called *non-linear*. Following the discretization concept, many of the continuous VSSA have been discretized. As a matter of fact, discretized versions of almost all the continuous LA have been reported [24, 25, 35, 37, 38, 64]. As alluded to earlier, in this research, we shall incorporate the concept of moving in a discretized and stepwise manner.

2.3 Stochastic Point Location Problem

We now formulate the Stochastic Point Location (SPL) problem, which is the third paradigm that we consider. To do this, we assume that there is a Learning Mechanism (LM) whose task is to determine the optimal value of some variable (or parameter), μ. We assume that there is an optimal choice for μ - an unknown value[1], say $\mu^* \in [0, 1]$. The SPL attempts to learn μ^*. Although μ^* is unknown to the LM, we assume that it has responses from an intelligent "Environment" E (*via* a Teacher/Oracle) which is capable of informing it whether the current value of μ is too small or too big. To render the problem both meaningful and distinct from its deterministic version, we emphasize that the responses from E are assumed "faulty". Thus, E may tell the LM to increase μ

[1]In the SPL, we assume that μ is any number in the interval $[0, 1]$. The question of generalizing this is rather straightforward.

when it should be decreased, and *vice versa*. However, to render the problem tangible, we initially assume that the probability of receiving an intelligent response is $p > 0.5$.

The quantity "p" reflects on the "effectiveness" of the Environment, E. Thus, whenever the current $\mu < \mu^*$, E correctly suggests that the LM increases μ with probability p. It simultaneously could have incorrectly recommended that it decreases μ with probability $(1-p)$. Similarly, whenever $\mu > \mu^*$, E advises the LM to decrease μ with probability p, and to increase it with probability $(1-p)$.

The crucial issue that we have to address is how to update our guess $\mu(n)$, at time n, of μ^*. Each updating mechanism yields a different potential solution to the SPL. The goal, of course, is for the proposed solution to converge to a value $\mu(\infty)$, whose expected value is arbitrarily close to μ^*. In line with the accepted terminology in the field of LA, we characterize such a scheme as being ϵ-optimal.

The reported solutions to the SPL are listed below:

1. The first paper proposed the problem and pioneered a solution operating in a discretized space [36];

2. The Continuous Point Location with Adaptive Tertiary Search (CPL-ATS) solution was one in which *three* LA worked in parallel to resolve it [41];

3. The extension of the latter, namely the Continuous Point Location with Adaptive d-ARY Search (CPL-AdS), used d LA in parallel [42], and it could operate in truth-telling and deceptive Environments;

4. The General CPL-AdS methodology extended the CPL-AdS to possess all the properties of the latter, but it could also operate in non-stationary Environments [20];

5. In the Hierarchical Stochastic Search on the Line (HSSL), the authors proposed that the LM moved to distant points in the interval (modelled hierarchically), and

specified by a tree [61];

6. The Symmetrical Hierarchical Stochastic Search on the Line (SHSSL) symmetrically enhanced the HSSL to work in deceptive Environments [62];

7. The Adaptive Step Search (ASS), used historical information within the last three steps to determine the current step size [52].

This concludes our brief survey of the SPL and permits us to proceed with our specific contributions.

Chapter 3

Linear Search-based CALA (LSCALA)

In this chapter, we propose the original Linear Search-Based CALA (LSCALA) and show that it is ϵ-optimal. The implementation of the algorithm is introduced in Section 3.1. Also, we discuss how the algorithm is related to the SPL problem, as well as, the "discretization" idea behind the design. To show the ϵ-optimality of the algorithm, we first analyze the convergence properties of LSCALA in Section 3.2. In particular, we show that its behaviour can be characterized by a Markov Chain (MC) model that asymptotically converges to its unique limiting distribution. Based on this, we discuss its long-term reward probability followed by proving its ϵ-optimality in Section 3.3[1]. In Section 3.4, we demonstrate some representative experimental results to corroborate our analysis claims. Finally, we conclude this chapter in Section 3.5.

[1]The analysis is made under the assumption that the reward function is strictly positive and Lipschitz continuous. We show that these assumptions can be considerably relaxed in Chapter 4.

3.1 Approach

To design the LSCALA, we assume, without loss of generality, that $[0, 1]$ is the action space of the Environment whose reward function is $\phi : [0, 1] \to [0, 1]$. Let $\mathbb{Z}_k$ denote integer set $\{0, 1, \cdots k-1\}$. Then for $n \in \mathbb{N}^*$, define a list of points

$$\mathbb{P}^n = [\theta_i^n := i/n : i \in \mathbb{Z}_{n+1}]$$

that are evenly scattered on the action space. Since, later, LSCALA searches the maximum of ϕ according to $\phi(\theta_i^n)$. We name θ_i^n probing points (PP), and n the PP Density (PPD). We abbreviate $\phi(\theta_i^n)$ to ϕ_i whenever there is no ambiguity. Let $\mathcal{P}$ be a pointer of $\mathbb{P}^n$ such that $\mathcal{P} = i$ means it points to θ_i^n. Pick some $m \in \mathbb{N}^*$. Define a variable $\mathcal{O}_m$ (named the order of $\mathcal{P}$) with two operations, **Inc**$\mathcal{O}$ and **Dec**$\mathcal{O}$. They set

$$\mathcal{O}_m = \mathcal{O}_m + 1 \mod m \qquad \& \qquad \mathcal{O}_m = \mathcal{O}_m - 1 \mod m,$$

respectively. Besides, if $\mathcal{O}_m = 0$ after performing **Inc**$\mathcal{O}$, we say $\mathcal{O}_m$ overflows. Similarly, if $\mathcal{O}_m = m - 1$ after implementing **Dec**$\mathcal{O}$, we say $\mathcal{O}_m$ underflows.

Then we implement LSCALA as follows. Initialize $\mathcal{P}$ and $\mathcal{O}_m$ with random initial values from Z_{n+1} and Z_m, respectively. The LA sets variable $\mathbb{I} = \mathcal{P} - 1$ or $\mathcal{P} + 1$ with the same probability 0.5. If $\mathbb{I} \notin \mathbb{Z}_{n+1}$[2], the LA gets a penalty without consulting the Environment. Else, it takes action $\theta_{\mathbb{I}}^n$ and has probability $\phi_{\mathbb{I}}$ to get a reward and $(1-\phi_{\mathbb{I}})$ a penalty. If the LA gets a penalty, it leaves $\mathcal{P}$ and $\mathcal{O}_m$ unchanged. Otherwise, it implements **Inc**$\mathcal{O}$ (if $\mathbb{I} = \mathcal{P}+1$) or **Dec**$\mathcal{O}$ (if $\mathbb{I} = \mathcal{P}-1$) followed by updating $\mathcal{P}$ according to Eq. (3.1). Finally, the LA repeats the procedure from setting $\mathbb{I}$. The pseudo-code of

[2]This is possible when $\mathcal{P} = 0$ or n.

the algorithm is given in Algorithm 1.

$$\mathcal{P} = \begin{cases} \mathcal{P} - 1 & \text{if } \mathcal{O}_m \text{ underflows} \\ \mathcal{P} + 1 & \text{if } \mathcal{O}_m \text{ overflows,} \\ \mathcal{P} & \text{else.} \end{cases} \tag{3.1}$$

The design of $\mathcal{O}_m$ slows down the transition of $\mathcal{P}$. With bigger m, the pointer $\mathcal{P}$ requires a longer time in average to move to another PP. Thence, we name m the Transition Resistance of $\mathcal{P}$ (TR). We later show that after a sufficient number of iterations, larger m makes it more likely for $\mathcal{P}$ to point to a θ_i^n that has the largest ϕ_i.

Algorithm 1: Linear Search-Based CALA (LSCALA)

Input: reward function ϕ, PPD n, TR m.

Method:

Pick $\mathcal{P} \in Z_{n+1}$ and $\mathcal{O} \in Z_m$, randomly.;
while ***True*** **do**
  $\mathbb{I} \leftarrow$ choose $\mathcal{P}+1$ or $\mathcal{P}-1$ randomly;
  if $\mathbb{I} \in Z_{n+1}$ **then**
    resp $\leftarrow$ the environment's response given input $\theta_{\mathbb{I}}^n$;
    if *resp is reward* **then**
      if $\mathbb{I} = \mathcal{P}+1$ **then**
        Inc$\mathcal{O}$;
      else
        Dec$\mathcal{O}$;
      end if
      if $\mathcal{O}_m$ *underflows* **then**
        $\mathcal{P} \leftarrow \mathcal{P}-1$
      else if $\mathcal{O}_m$ *overflows* **then**
        $\mathcal{P} \leftarrow \mathcal{P}+1$
    end if
  end if
end while

The following remark guarantees that the updating is always within the pertinent

bounds:

Remark 1. *If $\mathcal{P} \in \mathbb{Z}_{n+1}$ initially, it stays in $\mathbb{Z}_{n+1}$ forever. Assume $\mathcal{P} = 0$ in an iteration. Then the LA cannot get a reward by choosing $\mathbb{I} = \mathcal{P} - 1 = -1$ as it is not in $\mathbb{Z}_{n+1}$. So the LA has no chance to implement* **Dec**$\mathcal{O}$*, and $\mathcal{O}_m$ cannot underflow. Namely, $\mathcal{P}$ cannot decrease to -1. Similarly, $\mathcal{P}$ cannot be updated to $n+1$ when $\mathcal{P} = n$. Therefore, $\mathcal{P}$ always stays in $\mathbb{Z}_{n+1}$.*

Based on the above formulation and algorithm, we can make the following pertinent assertions:

1. The updating is purely of a discretized flavor, thus satisfying one of the primary goals of the paper;

2. It is analogous to the scheme for the SPL [36], thus satisfying another primary goal of the paper;

3. The quantity μ of Section 2.3 is precisely in line and analogous to the quantity λ utilized here.

3.2 The Markovian Analysis on LSCALA

In this section, we shall formally analyze the convergence properties of Algorithm LSCALA. The main goal is to derive a closed-form expression for the limiting distribution of $\mathcal{P}$.

In Section 3.2.1, we show that the transition of $\mathcal{P}$ and $\mathcal{O}_m$ is a MC which has a unique stationary distribution. As a result, the limiting distribution of $\mathcal{P}$ exists. We give a closed-form expression of the limiting distribution and show its validity. In Section 3.2.2, we provide a more constructive way to find $\mathcal{P}$'s limiting distribution. Compared to the method we used in Section 3.2.1, this method is more intuitive and heuristic.

Our analysis is performed under the following assumption.

Assumption 1. *The reward function* $\phi(\lambda) > 0$ *for all* $\lambda \in [0, 1]$.

3.2.1 The Limiting Distribution of $\mathcal{P}$

The main goal of this section is to derive a closed-form expression for the limiting distribution of $\mathcal{P}$ (Theorem 4). For the sake of conciseness, we write $\mathcal{P}$ and $\mathcal{O}_m$ together as a tuple $(\mathcal{P}, \mathcal{O}_m)$. We first show that the transitions of $(\mathcal{P}, \mathcal{O}_m)$ can be characterized by a MC. Then we prove that, regardless of the initial distribution, the distribution of $(\mathcal{P}, \mathcal{O}_m)$ converges to a unique stationary distribution (denoted by π^*) after a sufficient number of iterations (Theorem 3). A closed-form expression of π^* is also provided. For a fixed $\mathcal{P}$, we accumulate its stationary probabilities in π^* over $\mathcal{O}_m$. Then a closed-form expression of $\mathcal{P}$'s limiting distribution follows (Theorem 4).

Figure 3.1 lists all the possible values of $(\mathcal{P}, \mathcal{O}_m)$ in a matrix-like presentation. Each dot represents a possible value. The dots in a column have the same $\mathcal{P}$, and those in a row have the same $\mathcal{O}_m$. The arrowed lines connecting the dots give the potential transition path of $(\mathcal{P}, \mathcal{O}_m)$. When **Dec**$\mathcal{O}$ happens, if $(\mathcal{P}, \mathcal{O}_m)$ is not at the first row (that is, $\mathcal{O}_m > 0$), $\mathcal{O}_m$ decreases by one without triggering the underflow. So $(\mathcal{P}, \mathcal{O}_m)$ moves to the dot above in the same column. Otherwise, $\mathcal{O}_m = 0$, and $\mathcal{O}_m$ underflows. So $\mathcal{P}$ decreases by one and $\mathcal{O}_m$ is set to $m - 1$. Namely, $(\mathcal{P}, \mathcal{O}_m)$ moves to the last dot of the previous column. Regardless of the case, when the LA implements **Dec**$\mathcal{O}$, let **Prev**$[(\mathcal{P}, \mathcal{O}_m)]$ be the dot to which $(\mathcal{P}, \mathcal{O}_m)$ moves. A similar transition pattern can be observed when the LA implements **Inc**$\mathcal{O}$. Specifically, if $\mathcal{O}_m \neq m - 1$, $(\mathcal{P}, \mathcal{O}_m)$ moves to the next dot below without changing the column. Else, it moves from the last dot of current column to the first dot of the right one. Whichever the case, let **Next** $[(\mathcal{P}, \mathcal{O}_m)]$ denote the dot to which $(\mathcal{P}, \mathcal{O}_m)$ moves. If neither **Dec**$\mathcal{O}$ nor **Inc**$\mathcal{O}$ is performed, $(\mathcal{P}, \mathcal{O}_m)$

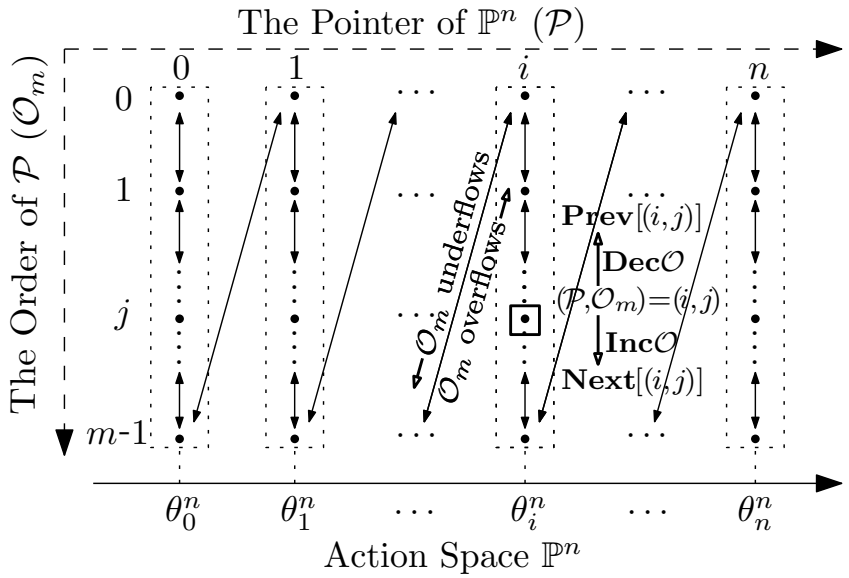


Figure 3.1: The transition graph of $(\mathcal{P}, \mathcal{O}_m)$ with the corresponding action $\theta_i^n \in \mathbb{P}^n$. Each dot represents a possible value of $(\mathcal{P}, \mathcal{O}_m)$. If two dots connected by an arrowed line, then $(\mathcal{P}, \mathcal{O}_m)$ could move from one dot to the other within one iteration. Suppose that $(\mathcal{P}, \mathcal{O}_m) = (i, j)$. If the LA implements **Dec**$\mathcal{O}$, then $(\mathcal{P}, \mathcal{O}_m)$ moves to **Prev**$[(i, j)]$, which equals $(i, j-1)$ if $j > 0$ and $(i-1, m-1)$ if $j = 0$ (that is, $\mathcal{O}_m$ underflows). If **Inc**$\mathcal{O}$ is performed, $(\mathcal{P}, \mathcal{O}_m)$ is updated to **Next** $[(i, j)]$, which is $(i, j+1)$ if $j < m-1$ and $(i+1, 0)$ if $j = m-1$ (namely, $\mathcal{O}_m$ overflows).

stays at the current dot.

Let $(\mathcal{P}, \mathcal{O}_m)_t$ denote the $(\mathcal{P}, \mathcal{O}_m)$ at the t-th iteration before the update. Suppose that $(\mathcal{P}, \mathcal{O}_m)_t = (i, j)$. The LA has probability 0.5 to take action θ_{i-1}^n and then has probability ϕ_{i-1} to get a reward. So it has probability $0.5\phi_{i-1}$ to implement **Dec**$\mathcal{O}$ and thus $(\mathcal{P}, \mathcal{O}_m)_{t+1} = \mathbf{Prev}[(i, j)]$. Namely,

$$Pr\left[(\mathcal{P}, \mathcal{O}_m)_{t+1} = \mathbf{Prev}[(i, j)] \,|(\mathcal{P}, \mathcal{O}_m)_t = (i, j)\right] = 0.5\phi_{i-1}. \tag{3.2}$$

Similarly, for $\mathbf{Inc}\mathcal{O}$, we have

$$Pr\left[(\mathcal{P},\mathcal{O}_m)_{t+1} = \mathbf{Next}\left[(i,j)\right] \middle| (\mathcal{P},\mathcal{O}_m)_t = (i,j)\right] = 0.5\phi_{i+1}. \tag{3.3}$$

If neither $\mathbf{Dec}\mathcal{O}$ nor $\mathbf{Inc}\mathcal{O}$ is performed, $(\mathcal{P},\mathcal{O}_m)$ stays at the current dot. That is,

$$Pr[(\mathcal{P},\mathcal{O}_m)_{t+1} = (i,j) | (\mathcal{P},\mathcal{O}_m)_t = (i,j)] = 1 - 0.5(\phi_{i-1} + \phi_{i+1}). \tag{3.4}$$

Remark 2. *Eqs (3.2)-(3.4) are not applicable when $\mathcal{P}$ equals zero or n since ϕ_{-1} and ϕ_{n+1} are not defined. To fix the problem, we set $\phi_{-1} = \phi_{n+1} = 0$ considering that the LA cannot implement* $\mathbf{Dec}\mathcal{O}$ *when $\mathcal{P} = 0$ and cannot perform* $\mathbf{Inc}\mathcal{O}$ *when $\mathcal{P} = n$ (Remark 1).*

To sum up, by Eqs (3.2)-(3.4), $(\mathcal{P},\mathcal{O}_m)_{t+1}$ follows a distribution fully determined by $(\mathcal{P},\mathcal{O}_m)_t$. Therefore, the transition of $(\mathcal{P},\mathcal{O}_m)$ is a MC, which is plotted in Figure 3.2. The figure, derived from Figure 3.1 (the dots in Figure 3.1 corresponds to the states in Figure 3.2), presents the transition probabilities of $(\mathcal{P},\mathcal{O}_m)$ given its current value. The transition probability is determined by $\mathcal{P}$, which can be seen from Eqs (3.2)-(3.4). In more details, assume $(\mathcal{P},\mathcal{O}_m) = (i,j)$. It moves to $\mathbf{Prev}[(i,j)]$ and $\mathbf{Next}\left[(i,j)\right]$ with probability $u_i = 0.5\phi_{i-1}$ and $d_i = 0.5\phi_{i+1}$, respectively. Otherwise, it stays at (i,j) with probability $s_i = 1 - 0.5(\phi_{i-1} + \phi_{i+1})$. According to Remark 2, $\phi_{-1} = \phi_{n+1} = 0$, which implies $u_0 = d_n = 0$. So, although $(\mathcal{P},\mathcal{O}_m)$ starts from some states outside the red cycle, it finally leaves the state and never comes back. In other words, these states are transient. By Assumption 1, $\phi(\lambda) > 0$ for all $\lambda \in [0,1]$. Then $\phi_i > 0$ for $i \in Z_{n+1}$, and thus all u_i and d_i except u_0 and d_n are greater than zero. So, any two states in the red cycle can communicate with each other. Combining with the fact that once $(\mathcal{P},\mathcal{O}_m)$ enters the red cycle, it cannot leave (since $u_0 = d_n = 0$), the states in the red cycle form

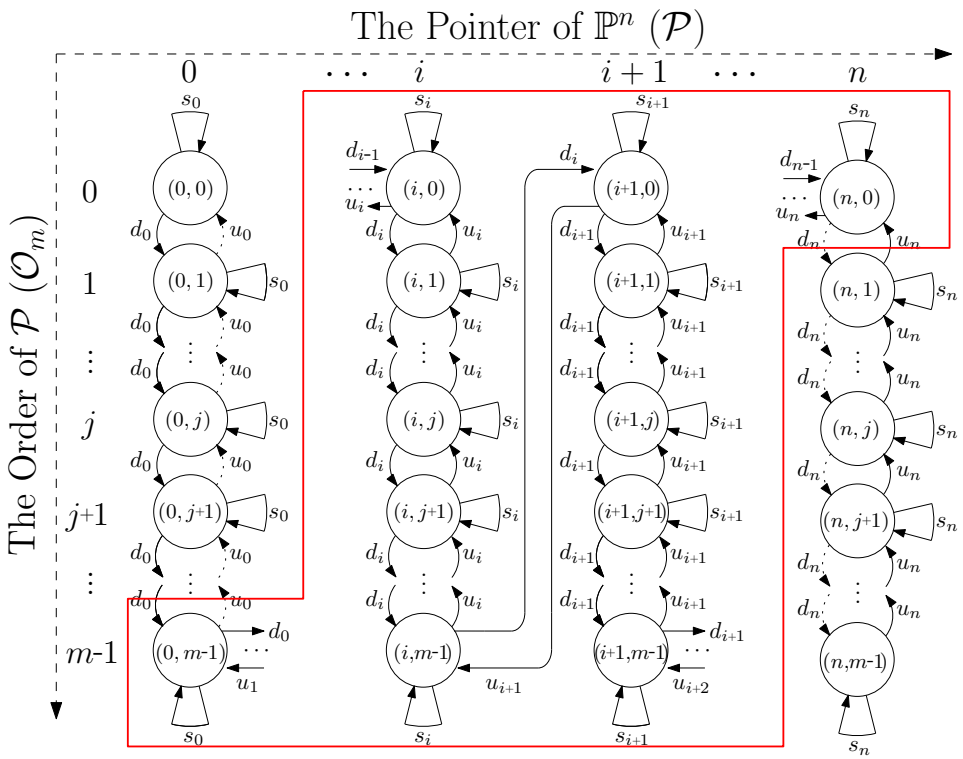


Figure 3.2: The MC of $(\mathcal{P}, \mathcal{O}_m)$ with the transition probabilities, where $u_i = 0.5\phi_{i-1}$, $d_i = 0.5\phi_{i+1}$ and $s_i = 1 - 0.5(\phi_{i-1} + \phi_{i+1})$. Note that the dotted transition edges in the first and the last columns do not exist as the transition probabilities $u_0 = d_n = 0$. This makes transient the states outside the red cycle. The states within the cycle make up the only CCC of the MC.

the only CCC[3] of the MC. Thus, we have the following theorem,

Theorem 1. *As $t \to \infty$, the MC in Figure 3.2 converges to a MC restricted to the states of the CCC (the red cycle).*

In the sequel, we call the MC of $(\mathcal{P}, \mathcal{O}_m)$ the Original MC (OMC) and the one, to which it converges, the Restricted MC (RMC).

The uniqueness of the CCC also implies

Theorem 2. *The OMC and RMC have a unique stationary distribution.*

[3]That is, starting from any state in the class, we have a non-zero probability of reaching any states in the class and zero probability of escaping from the class [32, pp. 10-11].

Proof. Since the OMC only has one CCC, the class is recurrent [32, Theorem 1.5.6]. Therefore, the OMC has a unique stationary distribution π^* [27, Proposition 1.28]. The OMC itself is a CCC. So the same conclusion follows. □

Let $\pi(t)$ denote the distribution of $(\mathcal{P}, \mathcal{O}_m)$ at the t-th iteration and $\pi^*_{i,j}$ the probability of $(\mathcal{P}, \mathcal{O}_m) = (i, j)$ in the stationary distribution π^*. Then we have,

Theorem 3. $\lim_{m\to\infty} \pi(t) = \pi^*$ *regardless of the initial distribution* $\pi(1)$. *Besides,* $\pi^*_{i,j}$ *equals*

$$C \cdot \phi_i^m \phi_{i-1}^{m-1-j} \phi_{i+1}^j, \tag{3.5}$$

for $(i, j) \in \mathbb{Z}_{n+1} \times \mathbb{Z}_m$, *where* C *is some constant that makes the sum of all* $\pi^*_{i,j}$ *equal to one.*

Proof. Suppose that γ^* is the stationary distribution of the RMC, $\gamma(t)$ its distribution at the t-th iteration, and $\gamma^*_{i,j}$ the probability that $(\mathcal{P}, \mathcal{O}_m) = (i, j)$ in γ^*. Then, by Theorem 1, Theorem 3 follows if we show

1. $\lim_{t\to\infty} \gamma(t) = \gamma^*$ regardless of $\gamma(1)$,
2. $\gamma^*_{i,j} = C\phi_i^m \phi_{i-1}^{m-1-j} \phi_{i+1}^j$ for all states of the RMC, and
3. Eq (3.5) is zero for the states exclusively owned by OMC.

For the first two points, we show

Lemma 1. $\lim_{t\to\infty} \gamma(t) = \gamma^*$ *regardless of* $\gamma(1)$. *And*

$$\gamma^*_{i,j} = C \cdot \phi_i^m \phi_{i-1}^{m-1-j} \phi_{i+1}^j. \tag{3.6}$$

Proof. For state $(0, m-1)$, it has a loop with transition probability $s_0 = 1 - 0.5(\phi_{-1} + \phi_1) = 1 - 0.5\phi_1 > 0$. So the state is aperiodic, which implies the RMC is also aperiodic

[32, Lemma 1.8.2]. Since the RMC is a communicating class, it is irreducible. So, regardless of $\gamma(1)$, $\lim_{t\to\infty}\gamma(t) = \gamma^*$ [32, Theorem 1.8.3].

Then we show Eq (3.6) holds by showing

$$\gamma^*_{u,v} \cdot p_{(u,v)\to(r,s)} = \gamma^*_{r,s} \cdot p_{(r,s)\to(u,v)}$$

for any two states (u, v) and (r, s) in the RMC, where $p_{(u,v)\to(r,s)}$ and $p_{(r,s)\to(u,v)}$ are the transition probabilities from (u, v) to (r, s) and (r, s) to $(u.v)$ respectively [8, Cor 6.1]. In Figure 3.2, the states of the RMC are connected by transition edges in cascades. So it is sufficient to verify the equation when $(u, v) = (i, j)$ and $(r, s) = (i, j+1)$ for $0 < i < n$ and $0 \le j < m-1$, and $(u, v) = (i, m-1)$ and $(r, s) = (i+1, 0)$ for $0 \le i < n$. For the first case,

$$\gamma^*_{i,j} \cdot d_i = C \cdot \phi_i^m \phi_{i-1}^{m-1-j} \phi_{i+1}^j (0.5\phi_{i+1}) = C \cdot \phi_i^m \phi_{i-1}^{m-1-(j+1)} \phi_{i+1}^{j+1} (0.5\phi_{i-1}) = \gamma^*_{i,j+1} \cdot u_i.$$

For the second,

$$\gamma^*_{i,m-1} \cdot d_i = C \cdot \phi_i^m \phi_{i+1}^{m-1} (0.5\phi_{i+1}) = C \cdot \phi_{i+1}^m \phi_i^{m-1} (0.5\phi_i) = \gamma^*_{i+1,0} \cdot u_{i+1}.$$

So Eq (3.6) holds. □

For the third point, when $i = 0$ and $j = 0, 1, \cdots m-2$, since $\phi_{-1} = 0$, Eq (3.5) becomes

$$C \cdot \phi_0^m \phi_{-1}^{m-1-j} \phi_1^j = 0.$$

Similarly, for $i = n$ and $j = 1, \cdots m-1$, as $\phi_{n+1} = 0$, Eq (3.5) can be written as

$$C \cdot \phi_n^m \phi_{n-1}^{m-1-j} \phi_{n+1}^j = 0.$$

Therefore, Theorem 3 follows. □

Let $\widehat{\pi}_i^*$ denote the $Pr[\mathcal{P} = i]$ in the limiting distribution of $\mathcal{P}$. In other words, $\widehat{\pi}_i^* = \sum_{j=0}^{m-1} \pi_{i,j}^*$. Then

Theorem 4. *For* $0 \leq i \leq n$,

$$\widehat{\pi}_i^* = \frac{\mathbb{G}_i(m)}{\sum_{i=0}^{n} \mathbb{G}_i(m)}, \tag{3.7}$$

where

$$\mathbb{G}_i(m) = \phi_i^m \left(\sum_{j=0}^{m-1} \phi_{i-1}^{(m-1)-j} \phi_{i+1}^j \right). \tag{3.8}$$

Proof. By Theorem 3,

$$\widehat{\pi}_i^* = \sum_{j=0}^{m-1} \pi_{i,j}^* = C \cdot \phi_i^m \sum_{j=0}^{m-1} \phi_{i-1}^{m-1-j} \phi_{i+1}^j = C \cdot \mathbb{G}_i(m). \tag{3.9}$$

Since

$$1 = \sum_{i=0}^{n} \sum_{j=0}^{m-1} \pi_{i,j} = \sum_{i=0}^{n} \widehat{\pi}_i^* = C \cdot \sum_{i=0}^{n} \mathbb{G}_i(m),$$

then

$$C = \frac{1}{\sum_{i=0}^{n} \mathbb{G}_i(m)}.$$

Inserting it to Eq (3.9), we get Eq (3.7). □

3.2.2 A Constructive Method to Derive the Closed Form Expression of the Limiting Distribution of $\mathcal{P}$

In Section 3.2.1, we directly give the closed-form expression of the limiting distribution of $\mathcal{P}$ followed by showing its validity. In this section, we provide a more constructive method for finding the expression. So the main goal is to reprove Theorem 4. Our discussion assumes that we already know the existence of the stationary distribution π^*

of the MC demonstrated in Figure 3.2.

Theorem 5 discusses the expression of π^* regarding the states in the first and last columns of the chain. These states cannot be covered by the general formulas, Eqs. (3.10) and (3.11), introduced in Theorem 6. The general relationship among $\pi^*_{i,k}$ is discussed in Theorem 6. In particular, it constructs a relation between two consecutive states in the MC. Eq (3.10) shows the relation between the last state of any specific column and the first state of the next column while Eq (3.11) shows the one between two states within the same column. Corollary 1 gives a closed-form expression of $\pi^*_{i,k}$ in term of $\pi^*_{0,k}$. Lemma 3 derives an algebraic result for calculating the sum of a geometric sequence. The result allows us to get a closed-form expression of $\widehat{\pi}^*_i$ by taking the sum of all $\pi^*_{i,k}$ within one column. In more details, combining the result with Corollary 1, we can get a closed-form expression of $\widehat{\pi}^*_i$ in terms of $\pi^*_{i,0}$. Theorem 8 derives a closed-form expression for $\pi_{i,0}$ in terms of $\widehat{\pi}^*_0$. Together with Theorem 7, an expression for $\pi_{i,0}$ in terms of $\widehat{\pi}^*_0$ is derived in Lemma 4. Since $\sum_{i=0}^{n} \widehat{\pi}^*_i = 1$, we normalize the expression and derive the closed-form expression of $\widehat{\pi}^*_i$.

We need the following lemma to continue the discussion.

Lemma 2 (Prop 1.28 of [27])**.** *Suppose a MC of finite states has a stationary distribution π^*. Let π^*_i denotes the probability of staying in state i when π^* is reached. Then for all transient states*[4] *i, $\pi^*_i = 0$.*

Recall Figure 3.2. In the MC of $(\mathcal{P}, \mathcal{O}_m)$, the states outside the red cycle are transient. Hence, by Lemma 2, its stationary distribution π^* must satisfy

Theorem 5. *For $0 \leq k < m-1$, $\pi^*_{0,k} = 0$. Also, for $0 < k \leq m-1$, $\pi^*_{n,k} = 0$.*

Besides, for two consecutive states of the MC,

[4]The original proposition in [27] uses "inessential states" instead of "transient states". According to Remark 1.24 of [27], when the state space is finite, "essential state" is equivalent to "recurrent state". Then a state that is not recurrent is transient, which is also called inessential.

Theorem 6. *$\pi^*_{i,j}$ satisfies the following equations:*

1. *For $0 \leq i \leq n-1$,*

$$\pi^*_{i+1,0} = \frac{\phi_{i+1}}{\phi_i} \cdot \pi^*_{i,m-1}. \tag{3.10}$$

2. *For $1 \leq i \leq n$, and $0 \leq j \leq m-1$,*

$$\pi^*_{i,j+1} = \frac{\phi_{i+1}}{\phi_{i-1}} \cdot \pi^*_{i,j}. \tag{3.11}$$

Proof. Let s and s' be two consecutive states in Figure 3.2. In particular, $s' = \mathbf{Next}[s]$. Assume that π^* is reached. Let π^*_s and $\pi^*_{s'}$ denote the probabilities that $(\mathcal{P}, \mathcal{O}_m) = s$ and s'. Besides, let $p_{ss'}$ be the transition probability from s to s' and $p_{s's}$ the one from s to s'. In Figure 3.2, the states are connected by transition edges in cascades. So by putting a divider between s and s', we partition the states into two groups (see Figure 3.3). In particular, let G_s be the set of states on path $(0,0) \leftrightarrow (0,1) \leftrightarrow \cdots \leftrightarrow s$ and G'_s the one on $s' \leftrightarrow \cdots \leftrightarrow (n, m-2) \leftrightarrow (n, m-1)$.

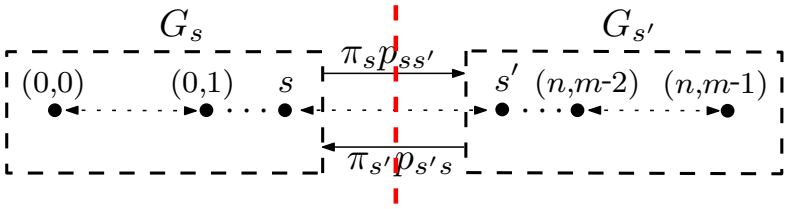


Figure 3.3: The divider between s and s' partitions the states of the MC into two groups, G_s and G'_s. When π^* is reached, the rate that $(\mathcal{P}, \mathcal{O}_m)$ moves from s to s' equals the one in the reversed direction. In other words, $\pi^*_s \cdot p_{ss'} = \pi^*_{s'} \cdot p_{s's}$.

Consider the scenario when the stationary distribution π^* has been attained. The rate at which $(\mathcal{P}, \mathcal{O}_m)$ moves from G_s to G'_s equals the one from G'_s to G_s. If this were not the case, $Pr[(\mathcal{P}, \mathcal{O}_m) \in G_s]$ and $Pr[(\mathcal{P}, \mathcal{O}_m) \in G_{s'}]$ change, which conflicts with the

assumption that the chain has attained its stationary distribution. Thence,

$$\pi^*_s \cdot p_{ss'} = \pi^*_{s'} \cdot p_{s's}. \tag{3.12}$$

If $s = (i, m-1)$, then $s' = (i+1, 0)$ (see Figure 3.2). Eq (3.12) becomes

$$\pi^*_{i,m-1} \cdot d_i = \pi^*_{i+1,0} \cdot u_{i+1} \iff \pi^*_{i+1,0} = \frac{d_i}{u_{i+1}} \cdot \pi^*_{i,m-1} = \frac{\phi_{i+1}}{\phi_i} \cdot \pi^*_{i,m-1}.$$

If $s = (i, j)$, for $0 \leq j < m-1$, $s' = (i, j+1)$. Eq (3.12) becomes

$$\pi^*_{i,j} \cdot d_i = \pi^*_{i,j+1} \cdot u_i \iff \pi^*_{i,j+1} = \frac{d_i}{u_i} \cdot \pi^*_{i,j} = \frac{\phi_{i+1}}{\phi_{i-1}} \cdot \pi^*_{i,j}.$$

□

The corollary below demonstrates the relation between a state and the first state of the column containing it.

Corollary 1. *For* $1 \leq i \leq n$,

$$\pi^*_{i,j} = \left(\frac{\phi_{i+1}}{\phi_{i-1}}\right)^j \cdot \pi^*_{i,0}. \tag{3.13}$$

Proof. The result is proven by induction. For the base case, consider the setting when $j = 0$, Eq (3.13) holds trivially. Assume Eq (3.13) holds for $j = k$. Then for $k+1$, by Eq (3.11),

$$\pi^*_{i,k+1} = \frac{\phi_{i+1}}{\phi_{i-1}} \cdot \pi^*_{i,k} = \frac{\phi_{i+1}}{\phi_{i-1}} \left(\frac{\phi_{i+1}}{\phi_{i-1}}\right)^k \pi^*_{i,0} = \left(\frac{\phi_{i+1}}{\phi_{i-1}}\right)^{k+1} \pi^*_{i,0},$$

which is also true. □

Based on Theorem 6, we have the following expression for the limiting distribution

of $\mathcal{P}$. Recall that

$$\widehat{\pi}_i^* = \sum_{j=0}^{m-1} \pi_{i,j}^*. \tag{3.14}$$

Theorem 7. *If* $i = 0$,

$$\widehat{\pi}_0^* = \pi_{0,m-1}^*. \tag{3.15}$$

Otherwise, for $1 \leq i \leq n$,

$$\widehat{\pi}_i^* = \pi_{i,0}^* \cdot \frac{\sum_{j=0}^{m-1} \phi_{i-1}^{(m-1)-j} \phi_{i+1}^{j}}{\phi_{i-1}^{m-1}}. \tag{3.16}$$

To prove Theorem 7, we need the following algebraic result, which deals with the sums of geometric sequences.

Lemma 3. *For* $p, q \in \mathbb{R}^+$,

$$\sum_{i=0}^{m-1} \left(\frac{p}{q}\right)^k = \frac{\sum_{k=0}^{m-1} q^{(m-1)-k} p^k}{q^{m-1}}. \tag{3.17}$$

Proof. If $p \neq q$, then Eq (3.17) is a geometric series with non-one constant ratio. So,

$$\begin{aligned}\sum_{i=0}^{m-1} \left(\frac{p}{q}\right)^k &= \frac{1-(p/q)^m}{1-p/q} = \frac{q^m - p^m}{q-p} \cdot \frac{1}{q^{m-1}} \\ &= \frac{\sum_{k=0}^{m-1} q^{(m-1)-k} p^k}{q^{m-1}}.\end{aligned}$$

Otherwise, $p = q$. Then,

$$\begin{aligned}\sum_{i=0}^{m-1} \left(\frac{p}{q}\right)^k &= \sum_{i=0}^{m-1} 1 = m = \frac{\sum_{k=0}^{m-1} q^{(m-1)-k} q^k}{q^{m-1}} \\ &= \frac{\sum_{k=0}^{m-1} q^{(m-1)-k} p^k}{q^{m-1}}.\end{aligned}$$

□

Proof of Theorem 7. By Theorem 5, $\pi^*_{0,k} = 0$ if $0 \le k < m-1$. So

$$\widehat{\pi}^*_0 = \sum_{j=0}^{m-1} \pi^*_{0,j} = \pi^*_{0,m-1}.$$

For $1 \le i \le n$, we have

$$\begin{aligned}
\widehat{\pi}^*_i &= \sum_{j=0}^{m-1} \pi^*_{i,j} = \pi^*_{i,0} \cdot \sum_{j=0}^{m-1} \left(\frac{\phi_{i+1}}{\phi_{i-1}}\right)^j && \text{(by Eq (3.13))} \\
&= \pi^*_{i,0} \cdot \frac{\sum_{j=0}^{m-1} \phi_{i-1}^{(m-1)-j} \phi_{i+1}^j}{\phi_{i-1}^{m-1}}. && \text{(by Lemma 3)}
\end{aligned}$$

□

For the states of the first row in Figure 3.2, we have

Theorem 8. *For* $1 \le i \le n$,

$$\pi^*_{i,0} = \frac{\phi_{i-1}^{m-1}\phi_i^m}{\phi_0^m \phi_1^{m-1}} \cdot \widehat{\pi}^*_0. \tag{3.18}$$

Proof. We prove the theorem by induction. For $i = 0$, Eq (3.18) holds by Eqs (3.10) and (3.15). Assume Eq (3.18) holds for $i = k$, then for $i = k+1$, we have

$$\begin{aligned}
\pi^*_{k+1,0} &= \frac{\phi_{k+1}}{\phi_k} \pi^*_{k,m-1} && \text{(by Eq (3.10))} \\
&= \frac{\phi_{k+1}}{\phi_k} \cdot \left(\frac{\phi_{k+1}}{\phi_{k-1}}\right)^{m-1} \cdot \pi^*_{k,0} && \text{(by Corollary 1)} \\
&= \frac{\phi_{k+1}}{\phi_k} \cdot \left(\frac{\phi_{k+1}}{\phi_{k-1}}\right)^{m-1} \cdot \frac{\phi_{k-1}^{m-1}\phi_k^m}{\phi_0^m \phi_1^{m-1}} \widehat{\pi}^*_0 \\
&= \frac{\phi_k^{m-1}\phi_{k+1}^m}{\phi_0^m \phi_1^{m-1}} \cdot \widehat{\pi}^*_0.
\end{aligned}$$

Thence, Eq (3.18) holds for $1 \le i \le n$. □

Then we can derive the expression for the limiting distribution of $\mathcal{P}$.

Lemma 4. *For* $0 \leq i \leq n$,

$$\widehat{\pi}_i^* = K \cdot \mathbb{G}_i(m), \tag{3.19}$$

where $K = \frac{\widehat{\pi}_0^*}{\phi_0^m \phi_1^{m-1}}$.

Proof. We use proof by induction. For $i = 0$, as $\phi_{-1} = 0$,

$$\mathbb{G}_0(m) = \phi_0^m \left(\sum_{k=0}^{m-1} \phi_{-1}^{(m-1)-k} \phi_1^k \right) = \phi_0^m \phi_1^{m-1}.$$

Then we have,

$$K \cdot \mathbb{G}_i(m) = \frac{\widehat{\pi}_0^* \cdot \mathbb{G}_0(m)}{\phi_0^m \phi_1^{m-1}} = \widehat{\pi}_0^* \cdot \frac{\phi_0^m \phi_1^{m-1}}{\phi_0^m \phi_1^{m-1}} = \widehat{\pi}_0^*.$$

Assume Eq (3.19) holds for $i = k$. Then for $i = k+1$,

$$\begin{aligned}
\widehat{\pi}_{k+1}^* &= \pi_{k,0}^* \cdot \frac{\sum_{j=0}^{m-1} \phi_{i-1}^{(m-1)-j} \phi_{i+1}^k}{\phi_{i-1}^{m-1}} && \text{(by Theorem 7)} \\
&= \left[\frac{\phi_{i-1}^{m-1} \phi_i^m}{\phi_0^m \phi_1^{m-1}} \widehat{\pi}_0^* \right] \cdot \frac{\sum_{j=0}^{m-1} \phi_{i-1}^{(m-1)-j} \phi_{i+1}^j}{\phi_{i-1}^{m-1}} && \text{(by Theorem 8)} \\
&= \frac{\widehat{\pi}_0^*}{\phi_0^m \phi_1^{m-1}} \left[\phi_i^m \sum_{j=0}^{m-1} \phi_{i-1}^{(m-1)-j} \phi_{i+1}^j \right] = K \cdot \mathbb{G}_i(m),
\end{aligned}$$

which completes the proof. □

Second Proof of Theorem 4. As $\sum_{i=0}^{n+1} \widehat{\pi}_i^* = 1$, by Lemma 4,

$$\sum_{i=0}^{n+1} \widehat{\pi}_i^* = K \sum_{i=0}^{n+1} \mathbb{G}_i(m) \implies K = \frac{1}{\sum_{i=0}^{n+1} \mathbb{G}_i(m)}.$$

Combined with Eq (3.19), Eq (3.7) follows. □

3.3 The Reward Probability and ϵ-optimality

The main goal of this section is to show that Algorithm LSCALA is ϵ-optimal. More precisely, suppose that the reward function ϕ has a global maximum ϕ_α at λ_α (that is, for all $\lambda \in [0,1]$, $\phi(\lambda) \leq \phi(\lambda_\alpha) = \phi_\alpha$). We shall formally prove that when the PPD n and TR m approach to infinity, the LA's reward probability approaches ϕ_α after a long-term execution. Simultaneously, the analysis also indicates that $\mathcal{P}$ points to a θ_i^n almost surely that ϕ_i is arbitrarily close to ϕ_α. Apart from Assumption 1, the rest of the analysis in this section also needs

Assumption 2. *The reward function ϕ is Lipschitz continuous and has no global maximum points at the boundaries of the action space (that is, not at zero or one).*[5]

Note that a Lipschitz continuous function is defined as follows.

Definition 1 (Lipschitz continuity)**.** *For function $f : X \subseteq \mathbb{R} \to Y \subseteq \mathbb{R}$, if there is a $K > 0$ such that for all $x_1, x_2 \in X$, $|f(x_1) - f(x_2)| \leq K|x_1 - x_2|$, then f is Lipschitz continuous with the Lipschitz constant K.*

Let $Rw_n(m)$ denote the reward probability of the LA after a long-term execution, which is also the reward probability after π^* is reached. Suppose the reward function ϕ has a global maximum ϕ_α The main goal of this section is to show LSCALA is ϵ-optimal (Theorem 11). That is,

$$\lim_{n\to\infty} \lim_{m\to\infty} Rw_n(m) = \phi_\alpha. \tag{3.20}$$

Also, we show that as $n, m \to \infty$, $\mathcal{P}$ equals some i almost surely that ϕ_i is arbitrarily close to ϕ_α (Corollary 2).

[5]This assumption is considerably relaxed in Chapter 4. In particular, the continuity near the global maximum points is sufficient to imply the ϵ-optimality.

Let $\mathcal{R}_i$ denote the reward probability given $\mathcal{P} = i$. When $\mathcal{P} = i$, the LA randomly takes either θ_{i-1}^n (with reward probability ϕ_{i-1}) or θ_{i+1}^n (with reward probability ϕ_{i+1}). So we have

$$\mathcal{R}_i = 0.5\phi_{i-1} + 0.5\phi_{i+1} = 0.5(\phi_{i-1} + \phi_{i+1}). \tag{3.21}$$

Then $Rw_n(m)$ can be written as

$$Rw_n(m) = \sum_{i=0}^{n} Pr[\mathcal{P} = i] \, Pr[\text{get a reward} | \mathcal{P} = i] = \sum_{i=0}^{n} \widehat{\pi}_i^* \mathcal{R}_i. \tag{3.22}$$

Let $\hat{\phi}_\alpha(n) = \max\{\phi_i : i \in Z_{n+1}\}$ and correspondingly, $\tilde{I}(n) = \{i \in Z_{n+1} : \phi_i = \hat{\phi}_\alpha(n)\}$, the set of indices of ϕ_i that gives the maximum. The next theorem shows $\hat{\phi}_\alpha(n)$ is a good approximation of ϕ_α when n is sufficiently large, and the θ_i^n with the largest ϕ_i is not at the boundaries.

Theorem 9. $\lim_{n\to\infty} \hat{\phi}_\alpha(n) = \phi_\alpha$. *Besides, there exists* $N \in \mathbb{N}$ *such that for all* $n > N$, $0, n \notin \tilde{I}(n)$.

Proof. Assume that ϕ has a global maximum point at λ_α. Since the distance between two consecutive θ_i^n is $\frac{1}{n}$, the distance between λ_α and the closest θ_i^n (denoted by θ_α^n) is less than $\frac{1}{n}$. Note that $\phi(\theta_\alpha^n) \leq \hat{\phi}_\alpha(n) \leq \phi_\alpha$. As ϕ is Lipschitz continuous (Assumption 2), let $K > 0$ be the Lipschitz constant. Then

$$0 \leq \phi_\alpha - \hat{\phi}_\alpha(n) \leq |\phi_\alpha - \phi(\theta_\alpha^n)| \leq K|\lambda_\alpha - \theta_\alpha^n| < \frac{K}{n}. \tag{3.23}$$

Therefore, $\lim_{n\to\infty} \hat{\phi}_\alpha(n) = \phi_\alpha$. As ϕ does not have a global maximum point at the boundaries (Assumption 2), $\lambda_\alpha \neq 0$ or 1. Pick $N \in \mathbb{N}$ such that $\frac{K}{N} < \min(\phi_\alpha - \phi_0, \phi_\alpha - \phi_n)$.

For all $n > N$, by Eq (3.23),

$$\phi_\alpha - \hat{\phi}_\alpha(n) < \frac{K}{n} < \frac{K}{N} < \min(\phi_\alpha - \phi_0, \phi_\alpha - \phi_n).$$

Then $\phi_\alpha - \hat{\phi}_\alpha(n) < \min(\phi_\alpha - \phi_0, \phi_\alpha - \phi_n) \leq \phi_\alpha - \phi_0$ implies $\hat{\phi}_\alpha(n) > \phi_0$. Similarly, $\hat{\phi}_\alpha(n) > \phi_n$. Hence, $0, n \notin \tilde{I}(n)$ by definition. □

Theorem 10 below reveals that for large n , given $\mathcal{P} \in \tilde{I}(n)$, the reward probability of the LA is roughly $\hat{\phi}_\alpha(n)$. Moreover, as $m \to \infty$, in limiting distribution of $\mathcal{P}$, $Pr\left[\mathcal{P} \in \tilde{I}(n)\right]$ converges to one.

Theorem 10. *When n is sufficiently large, $\mathcal{R}_i \approx \hat{\phi}_\alpha(n)$ for $i \in \tilde{I}(n)$. Besides,*

$$\lim_{m\to\infty} \sum_{i\in\tilde{I}(n)} \widehat{\pi}_i^* = 1. \tag{3.24}$$

Moreover, for $i \in \mathbb{Z}_{n+1} \setminus \tilde{I}(n)$, $\lim_{m\to\infty} \widehat{\pi}_i^* = 0$.

Proof. Let $K > 0$ be the Lipschitz constant of ϕ (Assumption 2). Then $|\phi_k - \phi_{k+1}| \leq K|\theta_k^n - \theta_{k+1}^n| = \frac{K}{n}$ where $k = 0, 1, \cdots, n-1$. So when n is sufficiently large, we have $\phi_k \approx \phi_{k+1}$. Theorem 9 shows $0, n \notin \tilde{I}(n)$ when n is big enough. Then for $i \in \tilde{I}(n)$, ϕ_{i-1} and ϕ_{i+1} are not ϕ_{-1} or ϕ_{n+1}, and thus, $\phi_{i-1} \approx \phi_i \approx \phi_{i+1}$. So by Eq (3.21), $\mathcal{R}_i = 0.5(\phi_{i-1} + \phi_{i+1}) \approx 0.5(\phi_i + \phi_i) = 0.5(\hat{\phi}_\alpha(n) + \hat{\phi}_\alpha(n)) = \hat{\phi}_\alpha(n)$. Recall Theorem 4. As $\phi_k \approx \phi_{k+1}$, for $1 \leq i \leq n-1$,

$$\mathbb{G}_i(m) = \phi_i^m \left(\sum_{k=0}^{m-1} \phi_{i-1}^{(m-1)-k} \phi_{i+1}^k \right) \approx m\phi_i^{2m-1}. \tag{3.25}$$

For $i = 0$, as $\phi_{-1} = 0$,

$$\mathbb{G}_0(m) = \phi_0^m \left(\sum_{k=0}^{m-1} 0^{(m-1)-k} \cdot \phi_1^k \right) \approx \phi_0^{2m-1}. \tag{3.26}$$

Similarly, for $i = n$, as $\phi_{n+1} = 0$,

$$\mathbb{G}_n(m) = \phi_n^m \left(\sum_{k=0}^{m-1} \phi_{n-1}^{(m-1)-k} \cdot 0^k \right) \approx \phi_n^{2m-1}. \tag{3.27}$$

Since $0, n \notin \tilde{I}(n)$ when n is big enough, then by Theorem 4,

$$\begin{aligned}
\frac{1}{\sum_{i\in\tilde{I}(n)} \widehat{\pi}_i^*} =& \frac{\sum_{i=0}^{n} \mathbb{G}_n(m)}{\sum_{i\in\tilde{I}(n)} \mathbb{G}_n(m)} = \frac{\sum_{i\in\tilde{I}(n)} \widehat{\pi}_i^*}{\sum_{i\in\tilde{I}(n)} \widehat{\pi}_i^*} + \frac{\sum_{i\in\{0,n\}} \widehat{\pi}_i^* + \sum_{i\in\mathbb{Z}_{n+1}\setminus(\tilde{I}(n)\cup\{0,n\})} \widehat{\pi}_i^*}{\sum_{i\in\tilde{I}(n)} \widehat{\pi}_i^*} \\
=& 1 + \frac{\phi_0^{2m-1} + \phi_n^{2m-1} + \sum_{i\in\mathbb{Z}_{n+1}\setminus(\tilde{I}(n)\cup\{0,n\})} m\phi_i^{2m-1}}{\sum_{i\in\tilde{I}(n)} m \left[\hat{\phi}_\alpha(n)\right]^{2m-1}} \\
=& 1 + \frac{o(1) + \sum_{i\in\mathbb{Z}_{n+1}\setminus(\tilde{I}(n)\cup\{0,n\})} o(1)}{\sum_{i\in\tilde{I}(n)} 1} \\
=& 1 + o(1).
\end{aligned}$$

Therefore,

$$\lim_{m\to\infty} \sum_{i\in\tilde{I}(n)} \widehat{\pi}_i^* = \lim_{m\to\infty} 1/(1 + o(1)) = 1,$$

and thus $\lim_{m\to\infty} \sum_{i\in\mathbb{Z}_{n+1}\setminus\tilde{I}(n)} \widehat{\pi}_i^* = 0$ as $\sum_{i\in\mathbb{Z}_{n+1}} \widehat{\pi}_i^* = 1$. So for $i \in \mathbb{Z}_{n+1} \setminus \tilde{I}(n)$, $\lim_{m\to\infty} \widehat{\pi}_i^* = 0$. □

Corollary 2. *As $n, m \to \infty$, in the limiting distribution of $\mathcal{P}$, the probability converges to one that $\mathcal{P}$ points a θ_i^n with ϕ_i arbitrarily close to ϕ_α.*

Proof. The result is immediate from Theorems 9 and 10. □

Theorem 11 (ϵ-optimal)**.**

$$\lim_{n\to\infty} \lim_{m\to\infty} Rw_n(m) = \phi_\alpha. \tag{3.28}$$

Proof. Recall Eq (3.22). According to Theorem 10, by choosing big enough n, we have

$$\begin{aligned}\lim_{m\to\infty} Rw_n(m) &= \lim_{m\to\infty} \sum_{i\in\tilde{I}(n)} \widehat{\pi}_i^* \mathcal{R}_i + \lim_{m\to\infty} \sum_{i\in\mathbb{Z}_{n+1}\setminus\tilde{I}(n)} \widehat{\pi}_i^* \mathcal{R}_i \\ &= \lim_{m\to\infty} \hat{\phi}_\alpha(n) \sum_{i\in\tilde{I}(n)} \widehat{\pi}_i^* + \sum_{i\in\mathbb{Z}_{n+1}\setminus\tilde{I}(n)} 0\cdot\mathcal{R}_i \\ &= \lim_{m\to\infty} \hat{\phi}_\alpha(n)\cdot 1 = \hat{\phi}_\alpha(n).\end{aligned}$$

Then by Theorem 9,

$$\lim_{n\to\infty}\lim_{m\to\infty} Rw_n(m) = \lim_{n\to\infty} \hat{\phi}_\alpha(n) = \phi_\alpha,$$

which completes the proof. □

3.4 Experimental Results

The LSCALA algorithm has been rigorously tested in a number of environments, and the results we have obtained are both interesting and conclusive. Although a considerable number of simulations have been done for various combinations of ϕ, m and n, in the interest of space and brevity we merely cite a few demonstrative examples. We emphasize, however, that the results submitted are typical and representative of the results obtained for the other settings. The experiments were conducted by choosing relatively small values of m and n. This was done so that the readers can verify our results within a reasonable amount of time. Of course, more accurate results can be obtained by increasing the values of m and n at the cost of slowing the speed with which the distribution of $\mathcal{P}$ converges to its asymptotic distributional form.

Algorithm LSCALA is tested with the Environment that provides responses according

to the reward function

$$\phi(x) = \min(0.1\sin(5x) + 0.05\sin(16x + 4) + 0.87, 1), \tag{3.29}$$

and the one that provides responses based on

$$\phi(x) = \begin{cases} -4x^2 + 1.12x + 0.7716, & \text{if } x \leq 0.3, \\ \min\left(1.1\exp\left(-9x^2 + 9x - 2.25\right), 1\right) & \text{if } 0.3 < x \leq 0.8, \\ 4x^2 + 3.69 - 7.2x & \text{elsewhere.} \end{cases} \tag{3.30}$$

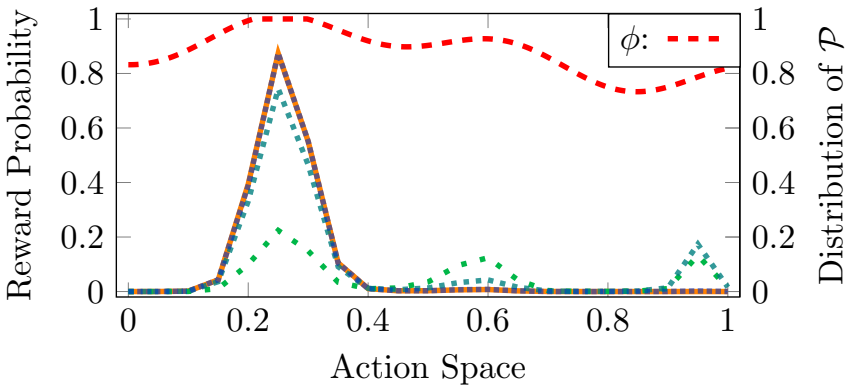
For both cases, $\phi_\alpha = 1$.

As one can see, the above two reward functions, plotted in Figures 3.4 and 3.5 respectively by the red dashed line, are greater than zero on its domain and do not have a global maximum point at zero or one. Their Lipschitz continuities can be seen by its bounded first derivative [51, Cor 6.4.20][6]. So ϕ satisfies Assumptions 1 and 2, and thus Algorithm LSCALA is applicable.

Figure 3.4 plots the results of the simulation where the reward function ϕ is defined by Eq (3.29), $n = 20$ and $m = 30$. We ran a group of $100,000$ LA implementing the LSCALA algorithm and plotted the estimated distribution curve of $\mathcal{P}$ and the reward rate as a function of the number of iterations. The results of the theoretical "infinite" number of iterations have also been calculated as per Theorem 4 and Eq (3.22). Additional theoretical results for other combinations of m and n have been provided in Table 3.1. From the table, we can clearly see that $Rw_n(m)$ approaches to ϕ_α as m and n increase.

Figure 3.5 plots the results of the simulation where the reward function ϕ is defined

[6]For both reward functions $\phi(x)$ plotted in Figures 3.4 and 3.5, there are finite number of points at which the function is not differentiable. However, the function is still continuous at these points, which so does not affect the Lipschitz continuity.



Distr. of **id**				
# iterations:	1×10^4	1×10^5	1×10^6	∞
Reward Rate:	92.33%	95.50%	98.21%	98.25%

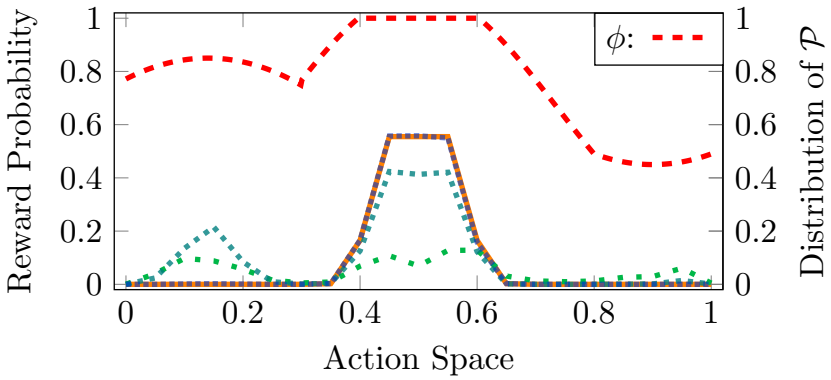
Figure 3.4: The estimated distribution curve of $\mathcal{P}$ as a function of the number of iterations. The estimation is made by running an ensemble of $100,000$ LA implementing Algorithm LSCALA with $n = 20$, $m = 30$. The environment gives responses according to the reward function ϕ defined by Eq (3.29).

Table 3.1: The $Rw_n(m)$ of the LA implementing Algorithm LSCALA. The reward function ϕ is defined by Eq (3.29).

n \ m	10	30	50	100
10	93.50%	95.00%	95.10%	95.11%
20	95.48%	98.25%	98.70%	99.05%
50	96.26%	99.22%	99.58%	99.77%
100	96.40%	99.39%	99.73%	99.89%

by Eq (3.30), $n = 20$ and $m = 32$. We ran an ensemble of $15,000$ LA implementing the LSCALA algorithm and plotted the estimated density curve of $\mathcal{P}$ and the reward rate as a function of the number of iterations. The results of the theoretical "infinite" number of iterations have also been calculated as per Theorem 4 and Eq (3.22). In the interest of completeness, additional theoretical results for other combinations of m and n have been provided in Table 3.2.

From both experiments, we can observe that the density curve of $\mathcal{P}$ and the reward rate converge to their theoretical counterparts, which effectively corroborates our analysis



Distr. of $\mathcal{P}$				
# iterations:	1×10^3	1×10^5	2×10^6	∞
Reward Rate:	85.52%	94.98%	99.05%	99.13%

Figure 3.5: The estimated density curve of $\mathcal{P}$ and the reward rate as a function of the number of iterations. The estimation is made by running an ensemble of 15,000 LA implementing the original LSCALA algorithm with $n = 20$, $m = 32$ and the reward function ϕ defined by Eq (3.30).

Table 3.2: The $Rw_n(m)$ of the LA implementing the LSCALA algorithm. The reward function ϕ is defined by Eq (3.30).

n \ m	8	16	32	64
10	93.17%	95.88%	97.54%	98.62%
20	96.44%	98.41%	99.13%	99.53%
30	97.18%	98.97%	99.47%	99.72%
40	97.47%	99.2%	99.62%	99.80%

about the distribution of $\mathcal{P}$ and $Rw_n(m)$ for a fixed m and n. Tables 3.1 and 3.2 clearly show that $Rw_n(m)$ approaches to ϕ_α when m and n increase. This trend also experimentally confirms our assertion that the LSCALA algorithm is ϵ-optimal.

3.5 Summary

In this chapter, we have reported the original LSCALA algorithm. We have discussed the "discretization" idea behind the algorithm design as well as the its relation with the

SPL problem. We have shown that the behaviour of the LSCALA can be characterized by a MC, which converges to its unique stationary distribution asymptotically. The LSCALA's long-term behaviour is then discussed, and we have proved that the algorithm is ϵ-optimal. Finally, we have provided a few experimental results, which can strongly corroborate our theoretical analysis.

Chapter 4

Proof of ϵ-optimality of LSCALA with Relaxed Preconditions and Enhanced LSCALA Algorithm

In this chapter, we shall show that the conditions that imply the ϵ-optimality of the original LSCALA algorithm in Chapter 3 can be considerably relaxed.

The analysis of this chapter consists of two parts. In Section 4.1, we show that the Lipschitz continuity of the reward function is unnecessary. Instead, we prove that the continuity near the global maximum points is enough to imply the algorithm's ϵ-optimality. Based on this, in Section 4.2, we further show that we do not need the condition that the reward function is strictly positive (Assumption 1). This is done by introducing a new CALA, named Enhanced LSCALA (E-LSCALA), by slightly modifying the original LSCALA algorithm. Then in Section 4.3, we demonstrate some simulation results to support our claims. Finally, we summarize our work in Section 4.4.

4.1 Proof of ϵ-optimality of LSCALA with Relaxed Preconditions

This section gives the proof of the ϵ-optimality of Algorithm LSCALA with a considerably relaxed version of Assumption 2. In particular, we show that Algorithm LSCALA is ϵ-optimal under Assumption 1 and

Assumption 3. *The set of local maxima can be written as a finite union of intervals, which could possibly be degenerate. Besides, for each interval containing the global maxima, it is nested by some open interval $I \subseteq [0,1]$ in which ϕ is continuous.*

Instead of assuming the continuity of the reward function at each point of the action space in Assumption 2, Assumption 3 only imposes constraints upon the intervals near the global maximum points. Our analysis in this section shows that the continuity near the reward function's global maximum points is sufficient to imply the ϵ-optimality of the LSCALA algorithm, which thus significantly broadens the algorithm's applicable scope and flexibility.

4.1.1 Rationale for ϵ-optimality Proofs

To provide the rationale for proof of the ϵ-optimality, we recall Theorem 4 which shows that:

$$\widehat{\pi}_i^* = \frac{\mathbb{G}_i(m)}{\sum_{j=0}^{n} \mathbb{G}_j(m)}, \tag{4.1}$$

where $\mathbb{G}_i(m) = \phi_i^m \left(\sum_{k=0}^{m-1} \phi_{i-1}^{(m-1)-k} \phi_{i+1}^k \right)$. We first demonstrate that for some i, $\widehat{\pi}_i^*$ approaches to a constant and for the rest of the indices, $\widehat{\pi}_i^*$ vanishes as $m \to \infty$. However, for all $\mathbb{G}_i(m)$, $\lim_{m\to\infty} \mathbb{G}_i(m) = 0$, which makes the numerator and denominator of Eq. (4.1) approach to zero together. So it is hard to analyze Eq. (4.1) directly. Instead,

we need to find some $C > 0$, such that for some i, $\frac{\mathbb{G}_i(m)}{C^m} \to \infty$ or a constant (in which case the corresponding $\widehat{\pi}_i^*$ goes to a constant), and for the rest of the indices, $\frac{\mathbb{G}_i(m)}{C^m} \to 0$ (in which case, the corresponding $\widehat{\pi}_i^*$ vanishes)[1]. In order to choose such C, we provide a sufficient condition when $\frac{\mathbb{G}_i(m)}{C^m}$ does and does not converge to zero in Lemmas 5 and 6, respectively. Interestingly, the C in Lemmas 5 and 6 involves two consecutive ϕ_i. So it is natural to choose $M = \max\{\phi_k\phi_{k+1} : 0 \leq k < n\}$ as the proper C to analyze $\widehat{\pi}_i^*$ when $m \to \infty$.

Having demonstrated this, we further let $\mathbb{K}$ denote the set of k that gives the maximum. By Lemma 5, we can find a set of i such that $\frac{\mathbb{G}_i(m)}{C^m} \to 0$. Let $\mathbb{L}$ be the set that contains the rest of i. Then, by Lemma 6, there exists such i that $\frac{\mathbb{G}_i(m)}{C^m} \to \infty$ or constant. So we can conclude that for $i \notin \mathbb{L}$, $\widehat{\pi}_i^*$ vanishes, and this is shown in Theorem 12. Let $\tau = \min\{\phi_{l-1}, \phi_{l+1} : l \in \mathbb{L}\}$ be a lower bound of the reward probability given $\mathcal{P} \in \mathbb{L}$. Since when $m \to \infty$, $\mathcal{P} = i$ for some $i \in \mathbb{L}$ almost surely, we show, in Lemma 7, that the long-term gross reward probability is lower bounded by τ. Lemmas 8 and 9 together show that $\lim_{n\to\infty} \tau = \phi_\alpha$, the global maximum of ϕ, since the variable x_τ that $\phi(x_\tau) = \tau$, approaches to its global maximum point as $n \to \infty$. Then, the ϵ-optimality follows (Theorem 13).

4.1.2 The Formal Optimality-related Analysis

We shall now prove the ϵ-optimality of the scheme under Assumption 3. We use the same notations introduced in Section 3.3. Recall that $Rw_n(m)$ denote the reward probability of the LA after a long-term execution, which is also called the Gross Reward Probability

[1]It is important to make sure both kinds of i exist. If all $\mathbb{G}_i(m)$ approach to infinity or zero, no conclusion can be drawn.

in the sequel. Recall Theorem 4,

$$\mathbb{G}_i\left(m\right) = \phi_i^m \left(\sum_{j=0}^{m-1} \phi_{i-1}^{(m-1)-j} \phi_{i+1}^j \right).$$

Lemma 5. *For* $0 \leq i \leq n$, *and* $C > max(\phi_{i-1}\phi_i, \phi_i\phi_{i+1})$, $\mathbb{G}_i\left(m\right) = o\left(C^m\right)$.

Proof. We prove this by considering two mutually exclusive and exhaustive cases, namely when $\phi_{i-1} = \phi_{i+1}$, and then when $\phi_{i-1} \neq \phi_{i+1}$.

Firstly, if $\phi_{i-1} = \phi_{i+1}$, then $C > \phi_{i-1}\phi_i = \phi_i\phi_{i+1}$. Further,

$$\mathbb{G}_i\left(m\right) = \phi_i^m \left(\sum_{k=0}^{m-1} \phi_{i-1}^{(m-1)-k} \phi_{i+1}^k \right) = m\phi_i^m \phi_{i+1}^{m-1}.$$

Thus,

$$\lim_{m\to\infty} \frac{\mathbb{G}_i\left(m\right)}{C^m} = \lim_{m\to\infty} \frac{m\phi_i^m \phi_{i+1}^{m-1}}{C^m} = \frac{1}{\phi_{i+1}} \lim_{m\to\infty} \frac{m\phi_i^m \phi_{i+1}^m}{C^m} = \frac{1}{\phi_{i+1}} \lim_{m\to\infty} mq^m, \tag{4.2}$$

where $q = \frac{\phi_i\phi_{i+1}}{L} < 1$. By invoking L'Hospital's Rule,

$$\lim_{m\to\infty} mq^m = \lim_{m\to\infty} \frac{m}{q^{-m}} = \lim_{m\to\infty} \frac{1}{-q^{-m}\ln q} = \lim_{m\to\infty} \frac{1}{-\ln q} q^m = 0.$$

Consequently,

$$\lim_{m\to\infty} \frac{\mathbb{G}_i\left(m\right)}{C^m} = \frac{1}{\phi_{i-1}} \lim_{m\to\infty} mq^m = \frac{1}{\phi_{i-1}} \cdot 0 = 0.$$

Secondly, if $\phi_{i-1} \neq \phi_{i+1}$, without loss of generality, assume $\phi_{i-1} > \phi_{i+1}$ (so $\phi_{i-1} > 0$). Then

$$\mathbb{G}_i\left(m\right) = \phi_i^m \left(\sum_{k=0}^{m-1} \phi_{i-1}^{(m-1)-k} \phi_{i+1}^k \right) = \phi_{i-1}^{-1}(\phi_i^m \phi_{i-1}^m) \sum_{k=0}^{m-1} \left(\frac{\phi_{i+1}}{\phi_{i-1}} \right)^k.$$

So,

$$\lim_{m\to\infty}\frac{\mathbb{G}_i\left(m\right)}{C^m}=\lim_{m\to\infty}\phi_{i-1}^{-1}\left(\frac{\phi_i\phi_{i-1}}{C}\right)^m\cdot\sum_{k=0}^{m-1}\left(\frac{\phi_{i+1}}{\phi_{i-1}}\right)^k.$$

Since $\phi_{i-1}>\phi_{i+1}$, $C>\max\{\phi_{i-1}\phi_i,\phi_i\phi_{i+1}\}=\phi_{i-1}\phi_i$, which implies $\frac{\phi_i\phi_{i-1}}{C}<1$. Besides, the geometric series $\lim_{m\to\infty}\sum_{k=0}^{m-1}\left(\frac{\phi_{i+1}}{\phi_{i-1}}\right)^k$ converges to some constant d as $\frac{\phi_{i+1}}{\phi_{i-1}}<1$. Thence,

$$\lim_{m\to\infty}\frac{\mathbb{G}_i\left(m\right)}{C^m}=\phi_{i-1}^{-1}\cdot d\lim_{m\to\infty}\left(\frac{\phi_i\phi_{i-1}}{C}\right)^m=0. \tag{4.3}$$

Hence, the lemma follows. □

Lemma 6. *For $0\leq i\leq n$, set $C=max(\phi_{i-1}\phi_i,\phi_i\phi_{i+1})$, then $\lim_{m\to\infty}\frac{\mathbb{G}_i(m)}{C^m}$ either converges to a constant or goes to infinity.*

Proof. We again prove this by considering two mutually exclusive and exhaustive cases, namely when $\phi_{i-1}=\phi_{i+1}$, and then when $\phi_{i-1}\neq\phi_{i+1}$.

In the first case, when $\phi_{i-1}=\phi_{i+1}$, then $C=\phi_{i-1}\phi_i=\phi_i\phi_{i+1}$. Recall Eq (4.2),

$$\lim_{m\to\infty}\frac{\mathbb{G}_i\left(m\right)}{C^m}=\frac{1}{\phi_{i+1}}\lim_{m\to\infty}\frac{m\phi_i^m\phi_{i+1}^m}{C^m}=\frac{1}{\phi_{i+1}}\lim_{m\to\infty}m=\infty. \tag{4.4}$$

For the second case, when $\phi_{i-1}\neq\phi_{i+1}$, without loss of generality, assume $\phi_{i-1}>\phi_{i+1}$ (so $\phi_{i-1}>0$). Then $\phi_{i-1}\phi_i>\phi_i\phi_{i+1}$. So $C=\phi_{i-1}\phi_i$. Now let $d=\lim_{m\to\infty}\sum_{k=0}^{m-1}\left(\frac{\phi_{i+1}}{\phi_{i-1}}\right)^k$. Then

$$\begin{aligned}\lim_{m\to\infty}\frac{\mathbb{G}_i\left(m\right)}{C^m}&=\lim_{m\to\infty}\phi_{i-1}^{-1}\left(\frac{\phi_i\phi_{i-1}}{C}\right)^m\sum_{k=0}^{m-1}\left(\frac{\phi_{i+1}}{\phi_{i-1}}\right)^k\\&=\phi_{i-1}^{-1}d\lim_{m\to\infty}\left(\frac{\phi_i\phi_{i-1}}{C}\right)^m=d\cdot\phi_{i-1}^{-1}1=d\cdot\phi_{i-1}^{-1},\end{aligned}$$

which is a constant. The result is thus proven. □

Let us now consider the scenario when n is a fixed value. Let

$$M = \max\{\phi_k \phi_{k+1} : 0 \leq k < n\}, \tag{4.5}$$

and $\mathbb{K}$ be the set of indices $\{k\}$ that yields the maximum. Furthermore, let $\mathbb{L}$ denote the set of $\mathcal{P}$ whose expressions of $\mathbb{G}_l(m)$ (see Eq (3.8)) involve ϕ_k and ϕ_{k+1} for $k \in \mathbb{K}$.

Remark 3. *For $l \in \mathbb{L}$, l is either equal to or consecutive to some $k \in \mathbb{K}$. Thus, if we pick $l \in \mathbb{L}$, the maximum M is attained at either $\phi_{l-1}\phi_l$ or $\phi_l\phi_{l+1}$, implying that the values of k are either $k = l-1$ or $k = l$.*

Theorem 12 shows if $l \notin \mathbb{L}$, $\widehat{\pi}_l^*$ vanishes as $m \to \infty$, which implies that the probability mass function is concentrated around some points contained in l.

Theorem 12. *For $l \notin \mathbb{L}$,*

$$\lim_{m\to\infty} \widehat{\pi}_l^* = 0. \tag{4.6}$$

Proof. Since $l \notin \mathbb{L}$, $M > \max(\phi_{l-1}\phi_l, \phi_l\phi_{l+1})$. By Lemma 5, $\mathbb{G}_l(m) = o(M^m)$. So by Theorem 4, we have,

$$\begin{aligned}
\lim_{m\to\infty} \widehat{\pi}_l^* &= \lim_{m\to\infty} \frac{\mathbb{G}_l(m)}{\sum_{j=0}^n \mathbb{G}_j(m)} = \lim_{m\to\infty} \frac{\mathbb{G}_l(m)/M^m}{\sum_{j=0}^n \mathbb{G}_j(m)/M^m} \\
&= \lim_{m\to\infty} \frac{o(1)}{\sum_{j\notin\mathbb{L}} o(1) + \sum_{j\in\mathbb{L}} \mathbb{G}_j(m)/M^m} \\
&= \lim_{m\to\infty} \frac{o(1)}{\sum_{j\in\mathbb{L}} \mathbb{G}_j(m)/M^m}.
\end{aligned}$$

According to Lemma 6, $\lim_{m\to\infty} \sum_{j\in\mathbb{L}} \mathbb{G}_j(m)/M^m$ either goes to infinity or approaches to a constant. Thus, Eq (4.6) follows. □

Lemma 7. *As m goes to infinity, the Gross Reward Probability, $Rw_n(m)$, satisfies*

$$\lim_{m\to\infty} Rw_n(m) \geq \tau, \tag{4.7}$$

where $\tau = \min\{\phi_{l-1}, \phi_{l+1} : l \in \mathbb{L}\}$.

Proof. As per Theorem 12,

$$\lim_{m\to\infty} \sum_{l\in\mathbb{L}} \widehat{\pi}_l^* = 0 + \lim_{m\to\infty} \sum_{l\in\mathbb{L}} \widehat{\pi}_l^* = \lim_{m\to\infty} \sum_{i\notin\mathbb{L}} \widehat{\pi}_l^* + \lim_{m\to\infty} \sum_{l\in\mathbb{L}} \widehat{\pi}_l^* = \lim_{m\to\infty} \sum_{l=0}^{n} \widehat{\pi}_l^* = 1. \tag{4.8}$$

Using the definition of the Gross Reward Probability, $Rw_n(m)$, from Eq (3.22),

$$\begin{aligned}
\lim_{m\to\infty} Rw_n(m) &= \lim_{m\to\infty} \sum_{l=0}^{n} \widehat{\pi}_l^* \cdot \mathcal{R}_l = \lim_{m\to\infty} \sum_{l\in\mathbb{L}} \widehat{\pi}_l^* \cdot \mathcal{R}_l && \text{(by Theorem 12)} \\
&= \lim_{m\to\infty} \sum_{l\in\mathbb{L}} \widehat{\pi}_l^* \cdot 0.5(\phi_{l-1} + \phi_{l+1}) && \text{(by Eq (3.21))} \\
&\geq \lim_{m\to\infty} \sum_{l\in\mathbb{L}} \widehat{\pi}_l^* \cdot \tau = 1 \cdot \tau = \tau, && \text{(by Eq (4.8))}
\end{aligned}$$

proving the result. □

Before we prove that the scheme converges to the global maximum, we need the following result.

Lemma 8. *Using the above value of τ, let $x_\tau = \frac{s}{n} \in [0,1]$ such that $\phi(x_\tau) = \phi_s = \tau$. Then $\lim_{n\to\infty} |x_\tau - y| = 0$, where $y = \frac{k}{n}$ for some $k \in \mathbb{K}$.*

Proof. Recall $\tau = \min\{\phi_{l-1}, \phi_{l+1} : l \in \mathbb{L}\}$. So $|s-l| = 1$ for some $l \in \mathbb{L}$. By Remark 3, there is some $k \in \mathbb{K}$ such that $|l-k| \leq 1$. This implies $|s-k| \leq |s-l| + |l-k| \leq 1+1 = 2$. Equivalently, $|x_\tau - y| = |\frac{s}{n} - \frac{k}{n}| \leq \frac{2}{n}$. Therefore, $0 \leq \lim_{n\to\infty} |x_\tau - y| \leq \lim_{n\to\infty} \frac{2}{n} = 0$. So $\lim_{n\to\infty} |x_\tau - y| = 0$. □

Let ϕ_α denote the global maximum of the reward function ϕ. Then the next lemma shows τ approaches to ϕ_α when n goes to infinity.

Lemma 9. *The limit of τ obeys:*

$$\lim_{n\to\infty} \tau = \phi_\alpha. \tag{4.9}$$

Proof. Let α be the point in $[0,1]$ such that $\phi(\alpha) = \phi_\alpha$. Note that the reward function ϕ satisfies Assumptions 3. So α is contained by some open interval I on which ϕ is continuous. When n increases, $y = \frac{k}{n}$ $(k \in \mathbb{K})$ approaches to α and is finally contained by I. Due to the continuity, we have $\lim_{n\to\infty} \phi(y) = \phi_\alpha$. By Lemma 8, $\lim_{n\to\infty} |x_\tau - y| = 0$. So $\lim_{n\to\infty} \phi(x_\tau) = \lim_{n\to\infty} \phi(y) = \phi_\alpha$. In other words, $\lim_{n\to\infty} \tau = \phi_\alpha$. □

The final theorem proving the ϵ-optimality follows.

Theorem 13. *The LSCALA scheme is ϵ-optimal because*

$$\lim_{n\to\infty} \lim_{m\to\infty} Rw_n(m) = \phi_\alpha. \tag{4.10}$$

Proof. To initiate the proof we recall Eq (3.21),

$$\mathcal{R}_j = 0.5(\phi_{j-1} + \phi_{j+1}) \leq 0.5 \cdot (\phi_\alpha + \phi_\alpha) = \phi_\alpha. \tag{4.11}$$

We now invoke Eq (3.22) to yield,

$$Rw_n(m) = \sum_{i=0}^{n} \widehat{\pi}_i^* \cdot \mathcal{R}_i \leq \sum_{i=0}^{n} \widehat{\pi}_i^* \cdot \phi_\alpha = 1 \cdot \phi_\alpha = \phi_\alpha. \tag{4.12}$$

Combining this with the results of Lemmas 7 and 9, we have

$$\phi_\alpha = \lim_{n\to\infty} \tau \leq \lim_{n\to\infty} \lim_{m\to\infty} Rw_n(m) \leq \lim_{n\to\infty} \phi_\alpha = \phi_\alpha, \tag{4.13}$$

which completes the proof. □

4.2 Enhanced LSCALA Algorithm

In this section, we shall show that Assumption 1 is not necessary. This will be achieved by marginally modifying the LSCALA algorithm. By only assuming that the reward function satisfies Assumption 3, we prove that the new scheme is also ϵ-optimal.

The modification is achieved by considering an enhanced reward function, which enhances ϕ so as to satisfy Assumption 1. This is done by adding a "subsistent" reward probability $c \in (0, 1)$ to ϕ. To be more specific, whenever the LA obtains a penalty for some $\lambda \in [0, 1]$, the new function still considers it to be a reward with probability c. In other words, the reward probability associated with the value x becomes:

$$\psi(x) = \phi(x) + c \cdot (1 - \phi(x)) = (1 - c) \cdot \phi(x) + c, \tag{4.14}$$

which is lower bounded by $c > 0$. The reader can easily verify that if ϕ satisfies Assumption 3, then the modified function, ψ, does too. We refer to the modified version of LSCALA which incorporates this modification as the Enhanced LSCALA, or E-LSCALA.

In the sequel, whenever an LA executes the modified algorithm, we shall refer to the probability of a reward given by the Environment, specified by ϕ, as the "*Objective* Reward Probability". On the other hand, the one is called "*Subjective* Reward Probability" (specified by ψ) that includes the subsistent reward probability c and determines the

updates of $(\mathcal{P}, \mathcal{O}_m)$. Let $\tilde{\psi}$ denote the global maximum of ψ. After the stationary distribution is reached, the overall objective reward probability is called the Gross Objective Reward Probability and is denoted by $Rw_n^o(m)$. Besides, its subjective counterpart is named the Gross Subjective Reward Probability and is denoted by $Rw_n^s(m)$. The main goal of the rest of this section is to show that

$$\lim_{n\to\infty}\lim_{m\to\infty} Rw_n^o(m) = \phi_\alpha.$$

Remark 4. *The Objective Reward Probability is the one we care about and thus, the one we try to maximize. In contrast, the Subjective Reward Probability is merely an auxiliary concept that allows us to use the results that have been proven in Section 4.1.*

Recall that we abbreviate $\phi(\theta_i) = \phi(i/n)$ to ϕ_i for $0 \le i \le n$ in Section 3.1, and set $\phi_{-1} = \phi_{n+1} = 0$ in Remark 2. Here we use the same convention to define ψ_i. In particular, let $\psi_i = \psi^*(i/n)$ for $0 \le i \le n$. Also, set $\psi_{-1} = \psi_{n+1} = 0$.

Note that for $0 \le i \le n$,

$$\psi_i = (1-c)\cdot\phi_i + c. \tag{4.15}$$

The transformation defined by Eq (4.14) only involves compression (multiplying by $(1-c)$) and shifting (adding the quantity c), which does not change the *relative order* of the function values. More precisely, $\phi(x) \le \phi(y)$ if and only if $\psi(x) \le \psi(y)$. Consequently, if $\alpha \in [0,1]$ satisfies $\phi(\alpha) = \phi_\alpha$, then we also have $\psi(\alpha) = \psi_\alpha$. According to Eq (4.14),

$$\psi_\alpha = (1-c)\cdot\phi_\alpha + c. \tag{4.16}$$

We now define $\mathbb{K}_\psi$ and $\mathbb{L}_\psi$ in the same way as we earlier defined $\mathbb{K}$ and $\mathbb{L}$, except that we replace ϕ by ψ. Given an LA that implements the E-LSCALA, the distribution of its $\mathcal{P}$ is the same as the one that implements the original LSCALA with the reward function

ψ. To simplify the explanation, we use the same notation $\widehat{\pi}_i^*$ to denote the probability that $\mathcal{P} = i$ after the stationary distribution is reached. We now state and prove some results that are true as n is sufficiently large.

Lemma 10. *The boundary points, $0, n \notin \mathbb{L}_\psi$ when n is sufficiently large.*

Proof. To prove this, we recall Assumption 3, from which we know that every globally maximum point (or interval) is nested by an open interval $I \subseteq [0, 1]$. This implies that the boundaries 0 (corresponding to group 0) and 1 (corresponding to group n) cannot be the global maximum points. Thence, by choosing a large enough value for n, for $k \in \mathbb{K}_\psi$, both $|0 - k|$ and $|n - k|$ are greater than unity and thus, $0, n \notin \mathbb{L}_\psi$ by virtue of Remark 3. □

To proceed with the analysis, to simplify the readability, we use notations similar to what was defined in Eq (3.21). Let $\mathcal{R}_j^o$ and $\mathcal{R}_j^s$ be the Objective and Subjective Reward Probabilities when $\mathcal{P} = j$. In particular,

$$\mathcal{R}_j^o = 0.5(\phi_{j-1} + \phi_{j+1}) \qquad \text{and} \qquad \mathcal{R}_j^s = 0.5(\psi_{j-1} + \psi_{j+1}). \tag{4.17}$$

Then $Rw_n^o(m)$ and $Rw_n^s(m)$ can be written as

$$Rw_n^o(m) = \sum_{i=0}^{n} \widehat{\pi}_i^* \cdot \mathcal{R}_i^o \qquad \text{and} \qquad Rw_n^s(m) = \sum_{i=0}^{n} \widehat{\pi}_i^* \cdot \mathcal{R}_i^s. \tag{4.18}$$

For $0 < i < n$,

$$\begin{aligned} \mathcal{R}_j^s &= 0.5(\psi_{j-1} + \psi_{j+1}) = 0.5[(1-c)\phi_{j-1} + c + (1-c)\phi_{j+1} + c] \\ &= 0.5(1-c)(\phi_{j-1} + \phi_{j+1}) + c = (1-c)\mathcal{R}_j^o + c. \end{aligned} \tag{4.19}$$

Remarkably, the relation does not hold for $i = 0$ or n because $\phi_{-1} = \psi_{-1} = 0$ and

$\phi_{n+1} = \psi_{n+1} = 0$. However, this does not affect the proof of the ϵ-optimality as $\widehat{\pi}_0^*$ and $\widehat{\pi}_n^*$ equal zero when $n \to \infty$ as per Lemma 10 and Theorem 12. This leads us to the main result of this setting.

Lemma 11. *When n is sufficiently large,*

$$\lim_{m\to\infty} Rw_n^s(m) = (1-c) \lim_{m\to\infty} Rw_n^o(m) + c. \tag{4.20}$$

Proof. According to Lemma 10, when n is sufficiently large, $0, n \notin \mathbb{L}$. Then,

$$\begin{aligned}
\lim_{m\to\infty} Rw_n^s(m) &= \lim_{m\to\infty} \sum_{i=0}^{n} \widehat{\pi}_i^* \cdot \mathcal{R}_i^s && \text{(by Eq (4.18))} \\
&= \lim_{m\to\infty} \sum_{i\in\mathbb{L}} \widehat{\pi}_i^* \cdot \mathcal{R}_i^s && \text{(by Theorem 12)} \\
&= \lim_{m\to\infty} \sum_{i\in\mathbb{L}} \widehat{\pi}_i^* \cdot [(1-c)\mathcal{R}_i^o + c] && \text{(by Lemma 10 and Eq (4.19))} \\
&= (1-c) \lim_{m\to\infty} \left[\sum_{i\in\mathbb{L}} \widehat{\pi}_i^* \cdot \mathcal{R}_i^o\right] + c \\
&= (1-c) \lim_{m\to\infty} Rw_n^o(m) + c. && \text{(by Eq (4.18) and Theorem 12)}
\end{aligned}$$

Whence, the theorem is proven. □

Theorem 14. *The algorithm E-LSCALA is ϵ-optimal. In other words,*

$$\lim_{n\to\infty} \lim_{m\to\infty} Rw_n^o(m) = \phi_\alpha. \tag{4.21}$$

Proof. As per Theorem 13, $Rw_n^s(m)$ of the LA implementing E-LSCALA with reward function ϕ satisfies,

$$\lim_{n\to\infty} \lim_{m\to\infty} Rw_n^s(m) = \psi_\alpha. \tag{4.22}$$

Combining the above with Lemma 11, we have

$$\psi_\alpha = \lim_{n\to}\lim_{m\to\infty} Rw_n^s(m) = (1-c)\lim_{n\to}\lim_{m\to\infty} Rw_n^o(m) + c. \tag{4.23}$$

Equivalently,

$$\lim_{n\to}\lim_{m\to\infty} Rw_n^o(m) = \frac{\psi_\alpha - c}{1-c}.$$

Rearranging Eq (4.16), we have $\phi_\alpha = \frac{\psi_\alpha - c}{1-c}$; thus, Eq (4.21) follows. □

4.3 Experimental Results

In this section, we demonstrate the strength of LSCALA and E-LSCALA algorithms by experimental results. Regarding Algorithm LSCALA, we focus on the reward functions that fail Assumption 2 but still satisfy Assumption 3. Although a significant number of tests have been implemented on both algorithms by choosing different combinations of ϕ, m, n and c, we only present a few here for the sake of conciseness. Note that the results we provide are typical and representative of the results obtained for the other configurations. In order to let the readers verify our results within a reasonable amount of time, the values of m and n we pick here are relatively small. Certainly, more accurate results can be obtained by increasing the values of m and n at the cost of slowing the convergence speed of the distribution of $\mathcal{P}$ which converges to its limiting distribution asymptotically.

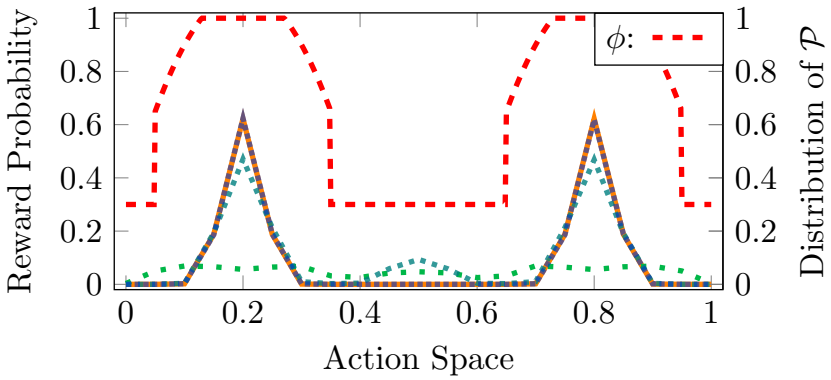
The first experiment was implemented with the reward function

$$\phi(x) = \begin{cases} \min(-20 \cdot (x - 0.2)^2 + 1.1, 1), & \text{if } 0.05 < x < 0.35, \\ \min(-20 \cdot (x - 0.8)^2 + 1.1, 1), & \text{if } 0.65 < x < 0.95, \\ 0.3, & \text{elsewhere.} \end{cases} \tag{4.24}$$

The function is plotted in Figure 4.1, which has the global maximum $\phi_\alpha = 1$. Obviously, the function is lower bounded by 0.3 and thus satisfies Assumption 1. Notice that the function is discontinuous at $x = 0.05$, 0.35, 0.65 and 0.95. So Assumption 2 does not hold, and then the analysis presented in Section 3.3 cannot be applied. In other words, based on Section 3.3, we cannot conclude that the LA scheme implementing Algorithm LSCALA is ϵ-optimal when interacting with the environment that gives rewards according to the reward function defined in Eq (4.24). However, the function satisfies Assumption 3 because for each interval consisting of global maximum points, it is nested by another interval on which the function is continuous. As a result, the analysis performed in Section 4.1 is applicable, and thus, the ϵ-optimality indeed holds.

Figure 4.1 plots[2] the results of the simulation where $n = 20$ and $m = 32$. We ran an ensemble of $20,000$ LA implementing the original LSCALA algorithm and plotted the estimated density curve of $\mathcal{P}$ and the reward rate as a function of the number of iterations. The results of the theoretical "infinite" number of iterations have also been calculated as per Theorem 4, Eq (3.21) and Eq (3.22). In the interest of completeness, additional theoretical results for other combinations of m and n have been provided in Table 4.1.

[2]In Figure 4.1, the density curve of $\mathcal{P}$ are scaled for a better representation. So readers may not expect that the area below the density curve is one. We apply the same treatment in Figure 4.2.



Distr. of $\mathcal{P}$				
# iterations:	1×10^2	1×10^3	4×10^5	∞
Reward Rate:	74.86%	90.03%	98.04%	98.06%

Figure 4.1: The estimated density curve of $\mathcal{P}$ and the reward rate as a function of the number of iterations. The estimation is made by running an ensemble of 20,000 LA implementing the LSCALA algorithm with $n = 20$, $m = 32$ and the reward function ϕ defined by Eq (4.24).

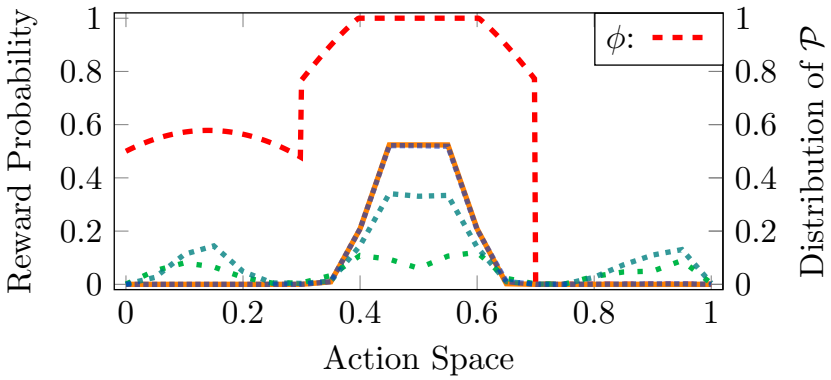
Table 4.1: The $Rw_n(m)$ of the LA implementing the LSCALA algorithm. The reward function ϕ is defined by Eq (4.24).

n \ m	10	16	32	64
10	84.89%	86.54%	88.14%	89.03%
20	95.92%	96.98%	98.06%	98.81%
50	98.61%	99.11%	99.52%	99.74%
80	98.98%	99.41%	99.72%	99.85%

The second experiment was performed with a reward function defined by

$$\phi(x) = \begin{cases} -4x^2 + 1.12x + 0.5, & \text{if } x < 0.3, \\ \min(1.1 \cdot \exp(-9x^2 + 9x - 2.25), 1) & \text{if } 0.3 \leq x < 0.7, \\ 0 & \text{elsewhere.} \end{cases} \tag{4.25}$$

The minimum of the function is zero and so Assumption 1 is not satisfied. Consequently, only the E-LSCALA algorithm can handle it. By running an ensemble of 20,000



Distr. of $\mathcal{P}$				
# iterations:	1×10^3	1×10^4	4×10^5	∞
Reward Rate:	65.92%	74.50%	98.51%	98.83%

Figure 4.2: The estimated density curve of $\mathcal{P}$ and the reward rate as a function of the number of iterations. The estimation is made by running an ensemble of 20,000 LA implementing the E-LSCALA algorithm with $n = 20$, $m = 32$, $c = 0.3$ and the reward function ϕ defined by Eq (4.25).

LA implementing the enhanced algorithm with $n = 20$, $m = 32$ and $c = 0.3$, we obtain the simulation results presented in Figure 4.2. The power of the scheme is obvious.

We have also plotted the estimated density curves of $\mathcal{P}$ and the corresponding reward rates for various numbers of iterations. The values for infinite number of iterations are calculated by applying the closed-form expression from Theorem 4 and the formula of $Rw_n^o(m)$ in Eq (4.18). Here, one should note that the computation of the stationary distribution of $\mathcal{P}$ is based on the *"Subjective"* reward function $\psi = 0.7 \cdot \phi + 0.3$ (see Eq (4.15)). Again, additional theoretical results for more choices of m and n are listed in Table 4.2, from which the reader should be able to appreciate the strength of the algorithms and the theoretical results.

From both experiments, we can observe that the density curve of $\mathcal{P}$ and the reward rate converge to their theoretical counterparts, which effectively corroborates our analysis about the distribution of $\mathcal{P}$ and $Rw_n(m)$ for a fixed m and n. Tables 4.1 and 4.2

Table 4.2: The $Rw_n^o(m)$ of the LA implementing the E-LSCALA algorithm with $c = 0.3$. The reward function ϕ is defined by Eq (4.25).

n \ m	10	16	32	64
10	91.59%	93.89%	96.39%	98.00%
20	96.93%	97.94%	98.83%	99.35%
50	98.38%	99.06%	99.57%	99.78%
80	98.54%	99.21%	99.69%	99.86%

clearly show that $Rw_n(m)$ approaches to ϕ_α when m and n increase. This trend also experimentally confirms our assertions that both LSCALA and E-LSCALA algorithms are ϵ-optimal.

4.4 Summary

In this chapter, we have shown that the assumptions we made in Chapter 3 can be considerably relaxed. In particular, our analysis shows that to ensure the ϵ-optimality of the algorithm, it is sufficient that the reward function is continuous near its global maximum points. Based on the original LSCALA algorithm, an enhanced version has been proposed. By adding a "subsistent" reward probability, the enhanced algorithm does not require the reward function to be strictly positive. The experimental results have been provided at the end of this chapter, which coincide with our theoretical analysis results.

Chapter 5

CALA-based Global Maximum Search

Finding the extreme values of an objective function plays a dominant role in almost all optimization problems. In some problems, local convergence is sufficient [14, 16]. But, in most of the cases, an algorithm of global convergence is preferable [17, 50].

In this chapter, based on Algorithm LSCALA, we devise a new algorithm, named CALA-GMS, for finding the global maximum of an objective function (denoted by f) by sampling its values which may be contaminated by noise.

More specifically, without loss of generality, we assume that f has domain $[0, 1]$. Let W be the Random Variable (r.v.) that generates noise and $w : \mathbb{R} \rightarrow \mathbb{R}$ its probability density function (pdf). By picking $\lambda \in [0, 1]$, a noisy function value $\tilde{f}(\lambda) = f(\lambda) + \epsilon$ is sampled, where $\epsilon \sim W$. We will show that by sampling the noisy function values, an LA scheme implementing Algorithm CALA-GMS can finally learn the global maximum points of f[1].

[1]Our analysis assumes that f is Lipschitz continuous

5.1 Approach

Algorithm CALA-GMS is implemented by adding an "interpreter" to the original LSCALA algorithm introduced in Chapter 3. Regarding the original algorithm, we have proved that it is ϵ-optimal, which implies that when m and n are sufficiently large, the LA learns the maximum of the reward function after a sufficient number of iterations. So if we use an interpreter to map the objection function f to a function of range $[0, 1]$, we can use the LSCALA algorithm to find its global maximum points and in turn to find the ones of f.

We implement the interpreter as follows. Choose a strictly increasing function $S : \mathbb{R} \to (0, 1)$ that is Lipschitz continuous (a potential choice is the sigmoid function). Then for all $\lambda \in [0, 1]$, $(\tilde{f}(\lambda))$ is in $(0, 1)$, and we use it as the reward probability of taking action λ to train the LA. In particular, we implement Algorithm CALA-GMS as follows.

In the original LSCALA, when the LA takes action λ, the Environment samples function value $f(\lambda)$ and gets $\tilde{f}(\lambda)$. Then the Environment has probability $S(\tilde{f}(\lambda))$ to give a reward and $1 - S(\tilde{f}(\lambda))$ a penalty.

In Section 5.2, we show that Algorithm CALA-GMS has the globally convergency. In particular, we prove that, an LA scheme implementing the algorithm can find the global maximum of f almost surely when m and n approach to infinity.

5.2 Analysis

Let f_α denote the global maximum of the objective function f. In this section, our main goal is to prove

Theorem 15. *If m and n approach to infinity, the probability converges to one that $\mathcal{P}$ points a θ_i^n with $f(\theta_i)$ arbitrarily close to f_α after a sufficient number of iterations.*

We complete the analysis under the following assumption.

Assumption 4. *The objective function $f : [0,1] \to \mathbb{R}$ is Lipschitz continuous, and its global maximum points are not at zero or one.*

Let $\phi(\lambda)$ be the reward probability given that the LA takes action λ in Algorithm CALA-GMS. Then we have

$$\phi(\lambda) = E[S(\tilde{f}(\lambda))]. \tag{5.1}$$

By Corollary 2, if $n, m \to \infty$, the probability converges to one that $\mathcal{P}$ points a θ_i^n with $\phi(\theta_i)$ arbitrarily close to the global maximum of ϕ. So, to prove Theorem 15, we only need to show

1. $\phi(\lambda)$ and $f(\lambda)$ have global maximum points at the same λ, and
2. ϕ satisfies Assumptions 1 and 2.

For the first point, it is sufficient to prove

Theorem 16. *For $x_1, x_2 \in [0,1]$, $f(x_1) > f(x_2)$ if and only if $\phi(x_1) > \phi(x_2)$.*

Proof. ($\Rightarrow$) Assume $f(x_1) > f(x_2)$. Then for $t \in \mathbb{R}$, $S(f(x_1)+t) > S(f(x_2)+t)$ as S is strictly increasing. Recall that w is the pdf of the noise. Then,

$$\begin{aligned}\phi(x_1) = E[S(\tilde{f}(x_1))] &= \int S(f(x_1)+t)w(t)dt \\ &> \int S(f(x_2)+t)w(t)dt = E[S(\tilde{f}(x_2))] = \phi(x_2).\end{aligned}$$

($\Leftarrow$) Assume that $\phi(x_1) > \phi(x_2)$. If $f(x_1) \leq f(x_2)$, then

$$\begin{aligned}\phi(x_1) = E[S(\tilde{f}(x_1))] &= \int S(f(x_1)+t)w(t)dt \\ &\leq \int S(f(x_2)+t)w(t)dt = E[S(\tilde{f}(x_2))] = \phi(x_2),\end{aligned}$$

which is a contradiction. Thence, $f(x_1) > f(x_2)$. □

By Assumption 4, since f does not have a global maximum point at zero or one, Theorem 16 implies that neither does ϕ. As the range of S is $(0, 1)$, for all $\lambda \in [0, 1]$, $\phi(\lambda) = E[S(\tilde{f}(\lambda))] > E[0] = 0$. Therefore, to complete the proof of Theorem 15, we only need to show

Theorem 17. *ϕ is Lipschitz continuous.*

Lemma 12. *For a r.v. X, $|E(X)| \leq E(|X|)$.*

Proof. Since $-E(|X|) = E(-|X|) \leq E(X) \leq E(|X|)$, $|E(X)| \leq |E(|X|)| = E(|X|)$. □

Proof of Theorem 17. Let M and N be the Lipschitz constants of S and f. Pick $x_1, x_2 \in [0, 1]$. We have

$$\begin{aligned}
|\phi(x_1) - \phi(x_2)| &= \left|E[S(\tilde{f}(x_1))] - E[S(\tilde{f}(x_2))]\right| \\
&= \left|E[S(\tilde{f}(x_1)) - S(\tilde{f}(x_2))]\right| \\
&\leq E\left[|S(\tilde{f}(x_1)) - S(\tilde{f}(x_2))|\right] && \text{(by Lemma 12)} \\
&= \int |S(f(x_1) + t) - S(f(x_2) + t)|\, w(t)dt \\
&\leq \int M\, |(f(x_1) + t) - (f(x_2) + t)|\, w(t)dt \\
&\leq \int M\, |(f(x_1) - f(x_2))|\, w(t)dt \\
&= MN|x_1 - x_2| \int w(t)dt = MN|x_1 - x_2|.
\end{aligned}$$

Therefore, ϕ is Lipschitz continuous by definition. □

5.3 Experimental Results

We demonstrate how Algorithm CALA-GMS works by testing it with the objective function

$$f(x) = -(0.5 - 2.5x)\sin(9x + 1). \tag{5.2}$$

The noise is simulated by $W \sim N(0, 0.1)$. Besides, we use

$$S(x) = 5 \cdot (e^{-x} + 5)^{-1} \tag{5.3}$$

for mapping the noisy function values into $(0, 1)$. Note that S is strictly increasing and Lipschitz continuous as its first derivative is positive and bounded. The function f is plotted in Figure 5.1 by the red dotted line, and it does not have a global maximum point at zero or one. Its Lipschitz continuity can also be seen by its bounded first derivative. So Assumption 4 is satisfied, and Algorithm CALA-GMS is applicable. In Figure 5.1, we

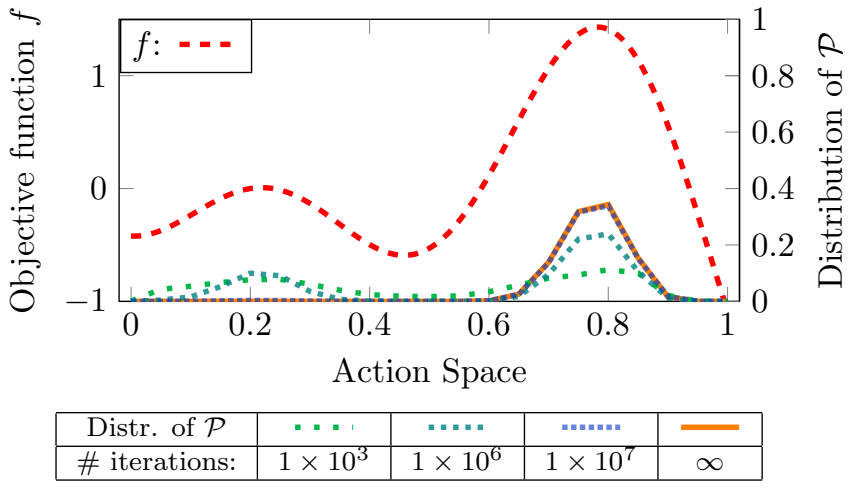


Figure 5.1: The estimated distribution curve of $\mathcal{P}$ as a function of the number of iterations. The estimation is made by running an ensemble of 100,000 LA performing Algorithm CALA-GMS with $n = 20$ and $m = 25$. The objective function is defined by Eq (5.2), and the function for mapping its function values into range $(0, 1)$ is defined by Eq (5.3).

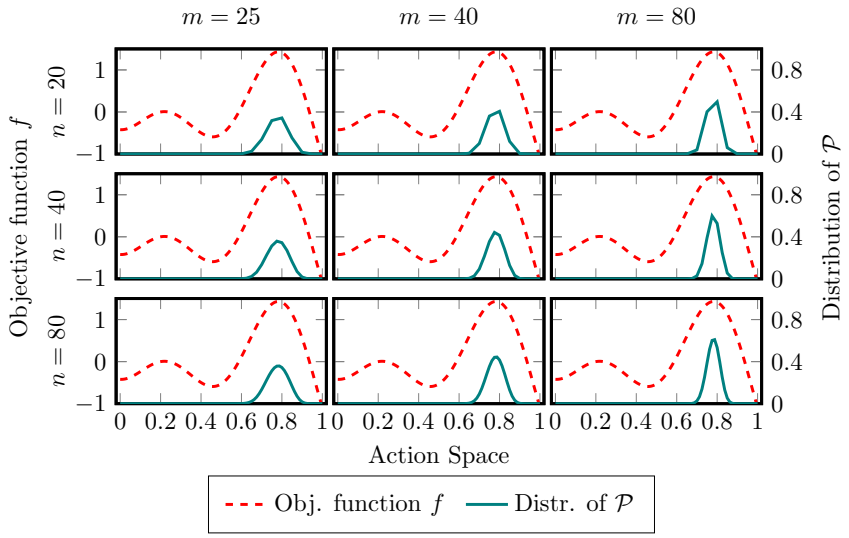


Figure 5.2: The theoretical distribution of $\mathcal{P}$ with various m and n. The objective function is defined by Eq (5.2).

plot the estimated distribution of $\mathcal{P}$ by running an ensemble of $100,000$ LA performing Algorithm CALA-GMS with $n = 20$ and $m = 25$. And the theoretical curve after infinite number of iterations is calculated through Theorem 4 and Eq (3.22) with the reward function $E[S(\tilde{f}(x))]$. Additional theoretical results are provided in Figure 5.2. It is clear that, when n and m increase, the limiting distribution of $\mathcal{P}$ concentrates to the point at which f reaches its global maximum.

From the experiment, we can observe that the distribution curve of $\mathcal{P}$ converges to its theoretical counterpart, which effectively corroborates our analysis about the limiting distribution of $\mathcal{P}$ for a fixed m and n. When m and n increase, Figure 5.2 clearly shows that the limiting distribution of $\mathcal{P}$ shrinks to the global maximum point of f. The trends experimentally confirm our assertions that CALA-GMS algorithm is globally convergent.

5.4 Summary

In this chapter, we have proposed a global maximum searching algorithm named CALA-GMS. The algorithm can find the global maximum points of a real-valued function by sampling its noisy function values. The experimental results coincide with our theoretical analysis, which thus strongly support our claims.

Chapter 6

The Performance Limit of Action Probability Distribution-based CALA Algorithms

All currently available LA algorithms take actions according to some action probability distributions. By interacting with the Environment, the LA modifies the distribution for the intention that the probability of taking the optimal action is maximized. All reported CALA algorithms, including those we have introduced in Chapters 3, 4 and 5, inherit such scheme.

In this chapter, we discuss the limit of the theoretical performance of the CALA implementing such scheme. In particular, for a reward function ϕ, we define an associated index $\tilde{\phi}$, referred to as the Optimal Reward Probability (ORP). We prove that it is the least upper bound of the long-term reward probability that any action probability distribution-based CALA algorithms can possibly achieve.

6.1 Preface

Our proof is constructive. In particular, in Section 6.2, we first define the ORP and show that it is an upper bound of the long-term reward probability for any CALA algorithms. Then we devise a new CALA algorithm, named G-LSCALA, that can handle more general reward functions than those we have discussed in the previous chapters in Section 6.3. We analyze the algorithm and show that its long-term reward probability reaches the ORP in an ϵ-optimal manner in Section 6.4, which implies that the ORP is in fact the least upper bound.

Apart from the proof, in Section 6.5, we provide an enhanced version of the G-LSCALA algorithm by applying a similar idea used in Section 4.2. We also show that the ϵ-optimality holds in the enhanced algorithm. After this, we give some comments and examples to show the strength of the new algorithms in Section 6.6. In this section, we also discuss the relation between the ORP and the reward function's global maximum points, which have been heavily used in the ϵ-optimality proofs of all classical LA problems. In Section 6.7, we use some experimental results to corroborate the claims implied by our theoretical analysis. Finally, in Section 6.8, we make a summary to conclude this chapter.

For the sake of conciseness and without loss of generality, our discussion assumes that the domain of the reward function ϕ is $[0, 1]$. To avoid tedious discussions upon the boundary cases, we further define an auxiliary function ϕ^* that extends the domain of ϕ to the real line.

$$\phi^*(x) = \begin{cases} \phi(x) & \text{if } x \in [0,1], \\ 0 & \text{elsewhere.} \end{cases} \tag{6.1}$$

6.2 Optimal Reward Probability (ORP)

In this section, we define the ORP of ϕ and discuss some typical properties of it. More discussions are given in Section 6.6. In Theorem 19, we prove that the eORP is an upper bound of the reward probability that a CALA can theoretically achieve. We defer the proof that it is also the least upper bound to the end of Section 6.4.2.

All the discussions in this section are made under the following assumption,

Assumption 5. *The reward function ϕ is integrable.*

Remark 5. *If a function is integrable on intervals I_1 and I_2, then it is also integrable on $I_1 \cup I_2$. Therefore, if ϕ is integrable, then so is the auxiliary function ϕ^*.*

Let $\mathcal{I}(x,l) = \int_x^{x+l} \phi^*$ and $\mathcal{I}(I) = \int_I \phi^*$. Given that $I = (a,b)$, $\mathcal{I}(I) = \int_I \phi^* = \int_a^b \phi^* = \mathcal{I}(a, b-a)$. If $l = \frac{1}{n}$ for some $n \in \mathbb{N}^*$, we write $\mathcal{I}\left(x, \frac{1}{n}\right) = \mathcal{I}^n(x)$. The average value of function f over an interval I is defined as $\frac{\int_I f}{|I|}$, where $|I|$ denotes the length of I. Specifically, for ϕ^*,

$$\mathcal{A}(I) = \frac{\mathcal{I}(I)}{|I|} \qquad \text{and} \qquad \mathcal{A}(x,l) = \frac{\mathcal{I}(x,l)}{l}.$$

The definition can be trivially generalized to handle mutually exclusive intervals $\{I_i\}_{i=0}^n$. In particular,

$$\mathcal{A}(\cup_{i=0}^n I_i) = \frac{\mathcal{I}(\cup_{i=0}^n I_i)}{\sum_{i=0}^n |I_i|}.$$

Similarly, if $|I| = \frac{1}{n}$ for some $n \in \mathbb{N}^*$,

$$\mathcal{A}^n(x) = \frac{1}{1/n} \cdot \mathcal{I}^n(x) = n\mathcal{I}^n(x).$$

Moreover, set

$$\tilde{\mathcal{A}}^n = \sup_{x\in[0,1]} \mathcal{A}^n(x)\,.$$

As ϕ^* is upper bounded by one, for all $n \in \mathbb{N}^*$ and $x \in [0,1]$,

$$\mathcal{A}^n(x) = n\mathcal{I}^n(x) \leq n\left(\frac{1}{n}\cdot 1\right) = 1.$$

Thence, $\sup_{n\in\mathbb{N}^*}\sup_{x\in[0,1]}\mathcal{A}^n(x)$ exists, and we name it the ORP of ϕ.

Definition 2 (Optimal Reward Probability)**.** *Given a reward function $\phi : [0,1] \to [0,1]$, its Optimal Reward Probability (ORP) is defined as,*

$$\tilde{\phi} = \sup_{n\in\mathbb{N}^*}\sup_{x\in[0,1]}\mathcal{A}^n(x) = \sup_{n\in\mathbb{N}^*}\tilde{\mathcal{A}}^n.$$

Theorem 18 shows that $\tilde{\mathcal{A}}^n$ attains its supremum $\tilde{\phi}$ when n approaches to infinity.

Theorem 18. $\lim_{n\to\infty}\tilde{\mathcal{A}}^n = \tilde{\phi}$.

In order to make the proof well-organized, we first prove some lemmas.

Lemma 13. *Suppose that $\{I_i\}_{i=1}^N$, a set of mutually exclusive bounded intervals, satisfies $\mathcal{A}\big(\cup_{i=1}^N I_i\big) > c$. Then the set contains at least one I_i such that $\mathcal{A}(I_i) > c$.*

Proof. We prove the lemma by contradiction. If the opposite is true, then $\mathcal{A}(I_i) \leq c$ for all i. Since

$$\mathcal{A}\big(\cup_{i=1}^N I_i\big) = \frac{\mathcal{I}\left(\cup_{i=0}^n I_i\right)}{\sum_{i=0}^n |I_i|} > c,$$

we have

$$c\cdot\sum_{i=1}^N |I_i| < \mathcal{I}\left(\cup_{i=1}^N I_i\right).$$

On the other side, as $\mathcal{A}(I_i) \leq c$ for all i,

$$\mathcal{I}\left(\cup_{i=1}^{N} I_i\right) = \sum_{i=1}^{N} \mathcal{I}\left(I_i\right) = \sum_{i=1}^{N} |I_i| \, \mathcal{A}(I_i) \leq \sum_{i=1}^{N} |I_i| \cdot c = c \cdot \sum_{i=1}^{N} |I_i| < \mathcal{I}\left(\cup_{i=1}^{N} I_i\right),$$

which is a contradiction. □

Corollary 3. *If $\tilde{\mathcal{A}}^n > c$, then for all $k \in \mathbb{N}^*$, $\tilde{\mathcal{A}}^{kn} > c$.*

Proof. Pick $k \in \mathbb{N}^*$. Since $\tilde{\mathcal{A}}^n > c$, there exists $x_0 \in [0,1]$ such that $\mathcal{A}^n(x_0) > c$. Partition interval $\left(x_0, x_0 + \frac{1}{n}\right)$ into k subintervals of length $\frac{1}{kn}$. Then, Lemma 13 shows that at least one subinterval I satisfies $\mathcal{A}(I) > c$. Hence, $\tilde{\mathcal{A}}^{kn} \geq \mathcal{A}(I) > c$. □

Corollary 4. *For all $p, q \in \mathbb{N}^*$ and $x \in [0,1]$, $\mathcal{A}\left(x, \frac{p}{q}\right) \leq \tilde{\phi}$.*

Proof. Assume the opposite is true. Then there exist some $p, q \in \mathbb{N}^*$ and $x_0 \in [0,1]$ such that $\mathcal{A}\left(x_0, \frac{p}{q}\right) > \tilde{\phi}$. Evenly partition interval $\left(x_0, x_0 + \frac{p}{q}\right)$ into p subintervals of length $\frac{1}{q}$. By Lemma 13, there exists at least one interval (denoted by I) that $\mathcal{A}(I) > \tilde{\phi}$. Hence, $\tilde{\phi} \geq \tilde{\mathcal{A}}^q \geq \mathcal{A}(I) > \tilde{\phi}$, a contradiction. □

Lemma 14. *For any functions g_1 and g_2 defined on some set X, we have*

$$\sup_{x \in X}(g_1 - g_2) \geq \sup_{x \in X} g_1 - \sup_{x \in X} g_2. \tag{6.2}$$

Proof. Notice that for any functions f_1 and f_2 defined on X, we have

$$\sup_{x \in X}(f_1 + f_2) \leq \sup_{x \in X} f_1 + \sup_{x \in X} f_2.$$

So we have

$$\sup_{x \in X} g_1 = \sup_{x \in X}[(g_1 - g_2) + g_2] \leq \sup_{x \in X}(g_1 - g_2) + \sup_{x \in X} g_2.$$

Hence, Eq (6.2) follows. □

Lemma 15. *If* $[0, 1 - \frac{1}{n}] \subset I$, $\sup_{x\in I} \mathcal{A}^n(x) = \sup_{x\in[0,1]} \mathcal{A}^n(x) = \tilde{\mathcal{A}}^n$.

Proof. Let $s = \sup_{x\in[0,1-\frac{1}{n}]} \mathcal{A}^n(x)$. We first show $s = \sup_{x\in I} \mathcal{A}^n(x)$ if $[0, 1 - \frac{1}{n}] \subset I$. For any interval I containing $[0, 1 - \frac{1}{n}]$, we have

$$\sup_{x\in\mathbb{R}} \mathcal{A}^n(x) \geq \sup_{x\in I} \mathcal{A}^n(x) \geq \sup_{x\in[0,1-\frac{1}{n}]} \mathcal{A}^n(x) = s.$$

So it is sufficient to show $\sup_{x\in\mathbb{R}} \mathcal{A}^n(x) = s$. We show it by proving that for $x \notin [0, 1-\frac{1}{n}]$, there exists $y \in [0, 1 - \frac{1}{n}]$ such that

$$\mathcal{A}^n(x) \leq \mathcal{A}^n(y)\,. \tag{6.3}$$

Without loss of generality, assume that $x > 1 - \frac{1}{n}$. If $x > 1$, $\mathcal{A}^n(x) = n\int_x^{x+\frac{1}{n}} \phi^* = 0$. Then for any $y \in [0, 1 - \frac{1}{n}]$, Eq (6.3) holds. Otherwise, for $x \in (1 - \frac{1}{n}, 1]$, $\int_{1-\frac{1}{n}}^{x} \phi^* \geq 0 = \int_1^{x+\frac{1}{n}} \phi^*$. So we have, $\int_x^{x+\frac{1}{n}} \phi^* = \int_x^1 \phi^* + \int_1^{x+\frac{1}{n}} \phi^* \leq \int_x^1 \phi + \int_{1-\frac{1}{n}}^{x} \phi = \int_{1-\frac{1}{n}}^1 \phi$. Thus, $\mathcal{A}^n(x) = n\int_x^{x+\frac{1}{n}} \phi^* \leq n\int_{1-\frac{1}{n}}^1 \phi = \mathcal{A}^n\big(1 - \frac{1}{n}\big)$. As $1 - \frac{1}{n} \in [0, 1 - \frac{1}{n}]$, we conclude that $s = \sup_{x\in I} \mathcal{A}^n(x)$. In particular, since $[0, 1 - \frac{1}{n}] \subset [0, 1]$, for $I \supset [0, 1 - \frac{1}{n}]$,

$$\tilde{\mathcal{A}}^n = \sup_{x\in[0,1]} \mathcal{A}^n(x) = s = \sup_{x\in I} \mathcal{A}^n(x).$$

□

Proof of Theorem 18. We show the equation holds by showing:

$$\limsup_{n\to\infty} \tilde{\mathcal{A}}^n \leq \tilde{\phi} \qquad \& \qquad \liminf_{n\to\infty} \tilde{\mathcal{A}}^n \geq \tilde{\phi}.$$

The proof of the first inequality is straightforward. In particular,

$$\limsup_{n\to\infty} \tilde{\mathcal{A}}^n = \lim_{n_0\to\infty} \sup_{n\geq n_0} \tilde{\mathcal{A}}^n \leq \lim_{n_0\to\infty} \left(\sup_{n\in\mathbb{N}^*} \tilde{\mathcal{A}}^n \right) = \tilde{\phi}.$$

We prove the second inequality by contradiction. If it does not hold, then there exists $\epsilon_0 > 0$ such that for all $N \in \mathbb{N}^*, \exists n' > N$,

$$\tilde{\mathcal{A}}^{n'} < \tilde{\phi} - \epsilon_0. \tag{6.4}$$

Since $\tilde{\phi} = \sup_{n\in\mathbb{N}^*} \tilde{\mathcal{A}}^n$, there exists $n_0 \in \mathbb{N}^*$ such that $\tilde{\mathcal{A}}^{n_0} > \tilde{\phi} - \frac{\epsilon_0}{2}$. Pick $n' > n_0$ satisfying Eq (6.4). Then there exists $k \in \mathbb{N}^*$ such that $kn_0 < n' \leq (k+1)n_0$. By Corollary 3, we have $\tilde{\mathcal{A}}^{kn_0} > \tilde{\phi} - \frac{\epsilon_0}{2}$. So,

$$\begin{aligned}
&\sup_{x\in[0,1]} \mathcal{I}\left(x, \frac{n'-kn_0}{kn_0n'}\right) = \sup_{x\in[0,1]} \mathcal{I}\left(x, \frac{1}{kn_0} - \frac{1}{n'}\right) = \sup_{x\in[0,1]} \int_x^{x+\frac{1}{kn_0}-\frac{1}{n'}} \phi^* \\
&= \sup_{x\in[0,1]} \left[\int_{x-\frac{1}{n'}}^{x+\frac{1}{kn_0}-\frac{1}{n'}} \phi^* - \int_{x-\frac{1}{n'}}^{x} \phi^* \right] = \sup_{x\in[0,1]} \left[\mathcal{I}^{kn_0}\left(x - \frac{1}{n'}\right) - \mathcal{I}^{n'}\left(x - \frac{1}{n'}\right) \right] \\
&= \sup_{x\in[-\frac{1}{n'},1-\frac{1}{n'}]} \left[\mathcal{I}^{kn_0}(x) - \mathcal{I}^{n'}(x) \right] \\
&\geq \sup_{x\in\left[-\frac{1}{n'},1-\frac{1}{n'}\right]} \mathcal{I}^{kn_0}(x) - \sup_{x\in\left[-\frac{1}{n'},1-\frac{1}{n'}\right]} \mathcal{I}^{n'}(x) && \text{(by Lemma 14)} \\
&= \tilde{\mathcal{A}}^{kn_0} - \tilde{\mathcal{A}}^{n'} && \text{(by Lemma 15)} \\
&\geq \frac{1}{kn_0} \cdot \left(\tilde{\phi} - \frac{\epsilon_0}{2}\right) - \frac{1}{n'} \cdot (\tilde{\phi} - \epsilon_0) \\
&\geq \frac{n'-kn_0}{kn_0n'}\tilde{\phi} + \frac{\epsilon_0(2kn_0-n')}{2kn_0} > \frac{n'-kn_0}{kn_0n'}\tilde{\phi}.
\end{aligned}$$

Then $\sup_{x\in[0,1]} \mathcal{A}\left(x, \frac{n'-kn_0}{kn_0n'}\right) > \tilde{\phi}$, contradicting to Corollary 4. Hence, the second in-

equality holds. So

$$\tilde{\phi} \leq \liminf_{n\to\infty} \tilde{\mathcal{A}}^n \leq \lim_{n\to\infty} \tilde{\mathcal{A}}^n \leq \limsup_{n\to\infty} \tilde{\mathcal{A}}^n \leq \tilde{\phi},$$

which completes the proof. □

Theorem 19. *For arbitrary CALA taking actions based on Continuous Probability Distributions (CPDs), its reward probability is upper bounded by* $\tilde{\phi}$.

Proof. Suppose that, in some iteration, a CALA takes an action according to the CPD defined by a density function $f(x)$. Since $\lim_{t\to\infty} t\int_x^{x+1/t} \phi^* = \phi^*(x)$ almost everywhere [15, Corollary 3.33], the reward probability Rw of the current iteration satisfies

$$\begin{aligned} Rw =& \int_{\mathbb{R}} \phi^*(x) f(x) = \int_0^1 \left[\lim_{t\to\infty} t\int_x^{x+1/t} \phi^*\right] f(x) \\ =& \int_0^1 \left[\lim_{t\to\infty} \mathcal{A}^t(x)\right] f(x) \\ \leq& \int_0^1 \left[\lim_{t\to\infty} \tilde{\mathcal{A}}^t\right] f(x) = \int_0^1 \tilde{\phi}\cdot f(x) \qquad \text{(by Theorem 18)} \\ =& \tilde{\phi}\cdot\int_0^1 f(x) \leq \tilde{\phi} \end{aligned}$$

Thence, its reward probability is always upper bounded by $\tilde{\phi}$. □

6.3 Generalized-LSCALA

In this section, we propose a new CALA algorithm named Generalized-LSCALA (G-LSCALA). Compared to the LSCALA algorithm introduced in Chapter 3, the analysis we make in Section 6.4 shows that it can handle more general reward functions. The analysis on this algorithm also implies that its long-term reward probability approaches to the ORP $\tilde{\phi}$ in an ϵ-optimal manner, which thus gives a constructive proof that $\tilde{\phi}$ is

the least upper bound of all action probability distribution-based CALA algorithms.

Without loss of generality, assume that $[0,1]$ is the action space of the environment whose reward function is $\phi : [0,1] \to [0,1]$. For $m, n \in \mathbb{N}^*$, define $\theta_i^{mn} = \frac{i}{mn}$ for $i \in \mathbb{Z}$ and a subset of the action space

$$\Theta^{mn} = \{\theta_i^{mn} : i = 0, 1, \cdots, mn\}.$$

Then G-LSCALA algorithm works as follows. Let **id** be a parameter whose initial value is randomly picked from Θ^{mn}. In a reinforcement-learning iteration, assume that $\mathbf{id} = \theta_i^{mn}$. The LA takes an action λ sampled from the uniform distribution $\mathbf{U}(\mathbf{id} - 1/n, \mathbf{id} + 1/n)$. If $\lambda \notin [0,1]$, it gets a penalty without consulting the environment. Otherwise, it has probability $\phi(\lambda)$ to get a reward and $1 - \phi(\lambda)$ a penalty. The LA updates **id** according to Eq (6.5) and repeats from taking an action. The pseudocode of the algorithm is given in Algorithm 2.

$$\mathbf{id} = \begin{cases} \theta_{i-1}^{mn} & \text{if a reward and } \lambda \le \theta_i^{mn}, \\ \theta_{i+1}^{mn} & \text{if a reward and } \lambda > \theta_i^{mn}, \\ \theta_i^{mn} & \text{else (a penalty).} \end{cases} \tag{6.5}$$

Remark 6. *In G-LSCALA algorithm,* $\mathbf{id} \in \Theta^{mn}$ *always. In an iteration, if* $\mathbf{id} \in \Theta^{mn} \setminus \{\theta_0^{mn}, \theta_{mn}^{mn}\}$ *initially, then the updated* **id** *is in* Θ^{mn}*. Otherwise, if* $\mathbf{id} = \theta_0^{mn} = 0$*, the LA cannot get a reward by taking* $\lambda \le 0$ *(*$\lambda = 0$ *has zero probability). So the updated* **id** *is either* θ_0^{mn} *or* θ_1^{mn} *and thus in* Θ^{mn}*. Similar conclusion can be drawn if* $\mathbf{id} = \theta_{mn}^{mn} = 1$*. Therefore, since the initial value of* **id** *is in* Θ^{mn}*,* $\mathbf{id} \in \Theta^{mn}$ *forever.*

As **id** is at the middle of the sample space, we name it kernel and thus Θ^{mn} kernel space. Most of the analysis in the sequel is made with fixed m and n. For the sake of

Algorithm 2: Generalized-LSCALA (G-LSCALA)

Input: m, n and ϕ.

Method:

```
pick i ∈ {0, 1, ··· , mn} randomly;
while True do
    λ ← a sample from U(θ_i^{mn} − 1/n, θ_i^{mn} + 1/n);
    if λ ∈ [0, 1] then
        response ← response from the environment given input λ;
        if (response = Reward) then
            if λ ≤ θ_i^{mn} then
                i ← i − 1
            else
                i ← i + 1
            end if
        end if
    end if
end while
```

conciseness, θ_i^{mn} is written as θ_i if the choices of m and n are clear.

6.4 Analysis

The main goal of the section is to show G-LSCALA algorithm is ϵ-optimal. Namely, when m and n approach to infinity, the reward probability in a long-term execution approaches to its theoretical maximum, the ORP defined in Section 6.2. As a corollary, we also prove that it is the least upper bound of the reward probability that the CALA, taking actions based on some CPD, can possibly achieve.

Our discussion is made as follows. In Section 6.4.1, we show that the transition of the kernel **id** can be modelled by a Markov chain. Also, a closed-form expression for its stationary distribution is derived. Then the ϵ-optimality of the algorithm is proved in Section 6.4.2. The analysis is made under the following assumption.

Assumption 6. *The reward function ϕ satisfies two requirements as follows.*

Req. *1* ϕ *is lower bounded by some* $\underset{\sim}{\phi} > 0$*, and*

Req. *2* ϕ *is of bounded variation.*

We show that **Req.** 1 is not necessary in Section 6.5 by implementing a modified G-LSCALA algorithm using a similar idea applied in Section 4.2.

Req. 2 essentially requires that ϕ cannot fluctuate arbitrarily. Readers who are interested in its rigorous definition may consult Ref. [26]. A function of bounded variation can be written as the difference of two bounded increasing functions [26, Theorem 2.18], which are (Riemann) integrable. As a result, ϕ is integrable, and thus, the definition of the ORP $\tilde{\phi}$ is applicable.

The analysis in the sequel except Theorem 23 only needs the integrability of ϕ. So merely considering ϕ as an integrable function does not affect the understanding of most of the analysis. In Section 6.6.2, we demonstrate some typical reward functions complying with **Req.** 2 and thus compatible with our algorithm.

In G-LSCALA algorithm, if $\lambda \notin [0, 1]$, the LA gets a penalty without consulting the environment. Equivalently, the LA always consults an environment that accepts all $\lambda \in \mathbb{R}$ and gives responses according to the auxiliary function ϕ^* defined in Eq (6.1). To avoid arduous discussions on the boundary cases, we mainly use ϕ^* in the sequel. Note that the behaviours of the LA in these two cases are exactly the same as long as the initial value of $\mathbf{id}$ is in Θ^{mn} (Remark 6).

6.4.1 Transition of the Kernel

We first show that the transition of the kernel $\mathbf{id}$ is a Markov chain. Then we give a closed-form expression for its stationary distribution. In order to keep our discussion concise, when the choices of m and n are clear, $\mathcal{I}^n(\theta_i^{mn})$ is abbreviated to $\mathcal{I}_i$. Let $\mathbf{id}(t)$ denote the value of $\mathbf{id}$ in the t-th iteration before the update. Assume that $\mathbf{id}(t) = \theta_i$.

Let Λ denote the random variable of the action taken by the LA . If $\Lambda = \lambda$, the response $\mathcal{R}(\lambda)$ of the environment follows a Bernoulli distribution with success probability $\phi(\lambda)$. Let $\mathcal{R}(\lambda) = 0$ indicate a penalty and $\mathcal{R}(\lambda) = 1$ a reward. For $\mathbf{id}(t+1)$, the probability that it equals θ_{i-1} is

$$\begin{aligned} &Pr\left[\mathbf{id}(t+1) = \theta_{i-1} | \mathbf{id}(t) = \theta_i\right] = Pr\left[\mathcal{R}(\Lambda) = 1 \wedge \Lambda \leq \theta_i | \mathbf{id}(t) = \theta_i\right] \\ &= \int_{\theta_i - \frac{1}{n}}^{\theta_i} Pr\left[\Lambda = \lambda | \mathbf{id}(t) = \theta_i\right] Pr\left[\mathcal{R}(\lambda) = 1 | \Lambda = \lambda \wedge \mathbf{id}(t) = \theta_i\right] d\lambda \\ &= \int_{\theta_{i-m}}^{\theta_{i-m}+\frac{1}{n}} \frac{n}{2} \cdot \phi^*(\lambda)\, d\lambda = \frac{n}{2} \cdot \int_{\theta_{i-m}}^{\theta_{i-m}+\frac{1}{n}} \phi^* = \frac{n}{2} \cdot \mathcal{I}_{i-m}. \end{aligned}$$

Similarly, we have,

$$\begin{aligned} Pr\left[\mathbf{id}(t+1) = \theta_{i+1} | \mathbf{id}(t) = \theta_i\right] &= Pr\left[\mathcal{R}(\Lambda) = 1 \wedge \Lambda > \theta_i | \mathbf{id}(t) = \theta_i\right] \\ &= \int_{\theta_i}^{\theta_i + \frac{1}{n}} \frac{n}{2} \phi^*(\lambda)\, d\lambda = \frac{n}{2} \cdot \mathcal{I}_i. \end{aligned}$$

So the probability of a reward (denoted by $\mathcal{K}_i$) given $\mathbf{id}(t) = \theta_i$ is the sum,

$$\mathcal{K}_i = Pr\left[\mathcal{R}(\Lambda) = 1 | \mathbf{id}(t) = \theta_i\right] = \int_{\theta_i - \frac{1}{n}}^{\theta_i + \frac{1}{n}} \frac{n}{2} \phi^* = \frac{n}{2} \cdot (\mathcal{I}_{i-m} + \mathcal{I}_i), \tag{6.6}$$

and the one of a penalty (also $\mathbf{id}(t+1) = \mathbf{id}(t)$) is the complement,

$$Pr\left[\mathbf{id}(t+1) = \theta_i | \mathbf{id}(t) = \theta_i\right] = 1 - \frac{n}{2} \cdot (\mathcal{I}_{i-m} + \mathcal{I}_i).$$

Since $\mathbf{id}(t+1)$ follows a distribution fully determined by $\mathbf{id}(t)$, the transition of $\mathbf{id}$ is a Markov chain. Figure 6.1 represents the process with the transition probabilities among $\theta_i \in \Theta^{mn}$, where $r_i = \frac{n}{2} \cdot \mathcal{I}_i$, $l_i = \frac{n}{2} \cdot \mathcal{I}_{i-m}$ and $s_i = 1 - \frac{n}{2} \cdot (\mathcal{I}_{i-m} + \mathcal{I}_i)$.

Let $\pi(t)$ denote the distribution of $\mathbf{id}(t)$ and $\pi^* = \lim_{t\to\infty} \pi(t)$ its stationary distri-

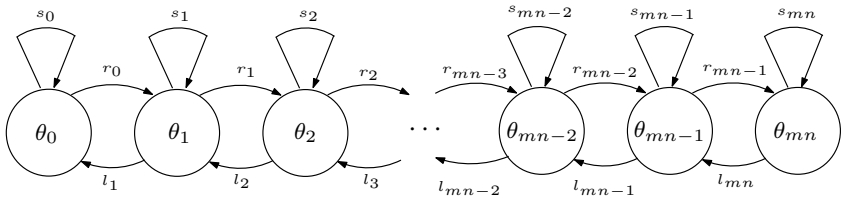


Figure 6.1: The transition diagram among θ_i, where $r_i = \frac{n}{2} \cdot \mathcal{I}_i$, $l_i = \frac{n}{2} \cdot \mathcal{I}_{i-m}$ and $s_i = 1 - \frac{n}{2} \cdot (\mathcal{I}_{i-m} + \mathcal{I}_i)$.

bution (we show the existence, uniqueness and independence from $\pi(1)$ soon). Besides, let $\pi_i(t)$ denote the probability that $\mathbf{id} = \theta_i$ in $\pi(t)$ and similarly π_i^* the one that $\mathbf{id} = \theta_i$ in π^*. Then we have,

Lemma 16. *As $t \to \infty$, $\pi(t)$ approaches to a unique stationary distribution π^* regardless of the initial distribution $\pi(1)$. Besides, for $i = 0, 1, \cdots, mn-1$,*

$$\pi_{i+1}^* = \pi_i^* \cdot \frac{\mathcal{I}_i}{\mathcal{I}_{i+1-m}}. \tag{6.7}$$

Proof. In Figure 6.1, it is possible to get to any states from any states. So the Markov chain is irreducible by definition. And thus, the existence and uniqueness of the stationary distribution π^* follows [27, Corollary 1.17]. Since each state has a loop, the Markov chain is aperiodic. Hence, $\lim_{t\to\infty} \pi(t) = \pi^*$ [27, Theorem 4.9]. Assume that the stationary distribution is reached. For $i = 0, 1, \cdots, mn-1$, the rate from states i to $i+1$ equals the reversed one (otherwise, $Pr[\mathbf{id} \le \theta_i]$ and $Pr[\mathbf{id} \ge \theta_{i+1}]$ change, which conflicts with the assumption). That is, $\pi_i^* \cdot r_i = \pi_{i+1}^* \cdot l_{i+1}$. So

$$\frac{\pi_{i+1}^*}{\pi_i^*} = \frac{r_i}{l_{i+1}} = \frac{(n/2) \cdot \mathcal{I}_i}{(n/2) \cdot \mathcal{I}_{i+1-m}} = \frac{\mathcal{I}_i}{\mathcal{I}_{i+1-m}},$$

and Eq (6.7) follows. □

Combined with the fact that $\sum_{i=0}^{mn} \pi_i^* = 1$, there is a closed-form expression for π_i^* as

follows.

Theorem 20. *For $i = 0, 1, \cdots, mn$,*

$$\pi_i^* = \pi_0^* \cdot \frac{\prod_{k=i-m+1}^{i-1} \mathcal{I}_k}{\prod_{k=1-m}^{-1} \mathcal{I}_k}, \tag{6.8}$$

where

$$\pi_0^* = \frac{\prod_{k=1-m}^{-1} \mathcal{I}_k}{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i-1} \mathcal{I}_k}. \tag{6.9}$$

Proof. For $i = 0$, Eq (6.8) holds trivially. Assume Eq (6.8) holds when $i = j$. Then for $i = j + 1$, by Lemma 16,

$$\pi_{j+1}^* = \pi_j^* \cdot \frac{\mathcal{I}_j}{\mathcal{I}_{j+1-m}} = \pi_0^* \cdot \frac{\prod_{k=j-m+1}^{j-1} \mathcal{I}_k}{\prod_{k=1-m}^{-1} \mathcal{I}_k} \cdot \frac{\mathcal{I}_j}{\mathcal{I}_{j+1-m}} = \pi_0^* \cdot \frac{\prod_{k=j-m+2}^{j} \mathcal{I}_k}{\prod_{k=1-m}^{-1} \mathcal{I}_k},$$

which is also true. Hence, Eq (6.8) holds for $i = 0, 1, \cdots, mn$. As $\sum_{i=0}^{mn} \pi_i^* = 1$,

$$1 = \sum_{i=0}^{mn} \pi_i^* = \sum_{i=0}^{mn} \left(\pi_0^* \cdot \frac{\prod_{k=i-m+1}^{i-1} \mathcal{I}_k}{\prod_{k=1-m}^{-1} \mathcal{I}_k} \right) = \pi_0^* \cdot \frac{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i-1} \mathcal{I}_k}{\prod_{k=1-m}^{-1} \mathcal{I}_k},$$

and Eq (6.9) follows. □

6.4.2 Gross Reward Probability and ϵ-optimallity

We call the overall reward probability in an iteration the gross reward probability. Recall Eq (6.6). For fixed m and n, the gross reward probability in the t-th iteration is

$$Rw_n(m, t) = \sum_{i=0}^{mn} \pi_i(t) \cdot Pr\left[\mathcal{R}(\Lambda) = 1 | \mathbf{id}(t) = \theta_i\right] = \sum_{i=0}^{mn} \pi_i(t) \cdot \mathcal{K}_i. \tag{6.10}$$

And correspondingly, when π^* is reached, the gross reward probability becomes

$$Rw_n(m) = \sum_{i=0}^{mn} \pi_i^* \cdot \mathcal{K}_i. \tag{6.11}$$

Since $Rw_n(m)$ is the probability that the LA gets a reward in a long-term execution, we call it the long-term gross reward probability in the sequel. Since the actions taken by the LA implementing G-LSCALA algorithm follows some CPD in each iteration, by Theorem 19 we have

Theorem 21. *For $m, n, t \in \mathbb{N}^*$, $Rw_n(m,t) \leq \tilde{\phi}$. Therefore, $Rw_n(m) \leq \tilde{\phi}$.*

In the rest of this section, we show that when $n, m \to \infty$, $Rw_n(m) \geq \tilde{\phi}$ and thus, the ϵ-optimality follows. As a corollary, the proof follows that $\tilde{\phi}$ is the least upper bound of the reward probability that the CALA, taking actions by CPDs, can possibly achieve. Let $\mathscr{I}^n(x)$ denote the normalized $\mathcal{I}^n(x)$. That is,

$$\mathscr{I}^n(x) = \frac{\mathcal{I}^n(x)}{\tilde{\mathcal{I}}^n},$$

where $\tilde{\mathcal{I}}^n = \sup_{x\in\mathbb{R}} \mathcal{I}^n(x) = n\tilde{\mathcal{A}}^n$. Similarly, $\mathscr{I}^n(\theta_i)$ is abbreviated to $\mathscr{I}_i$ whenever there is no ambiguity. Namely, $\mathscr{I}_i = \frac{\mathcal{I}_i}{\tilde{\mathcal{I}}^n}$. By Eqs (6.6) and (6.11) and Theorem 20, the

long-term gross reward probability can be written as

$$\begin{aligned}
Rw_n(m) &= \sum_{i=0}^{mn} \pi_i^* \cdot \mathcal{K}_i \\
&= \sum_{i=0}^{mn} \pi_i^* \cdot \frac{n}{2} \cdot (\mathcal{I}_{i-m} + \mathcal{I}_i) \\
&= \frac{n}{2} \cdot \sum_{i=0}^{mn} \pi_i^* \cdot (\mathcal{I}_{i-m} + \mathcal{I}_i) \\
&= \frac{n}{2} \cdot \sum_{i=0}^{mn} \pi_0^* \cdot \frac{\prod_{k=i-m+1}^{i-1} \mathcal{I}_k}{\prod_{k=1-m}^{-1} \mathcal{I}_k} \cdot (\mathcal{I}_{i-m} + \mathcal{I}_i) \\
&= \frac{n}{2} \cdot \frac{\prod_{k=1-m}^{-1} \mathcal{I}_k}{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i-1} \mathcal{I}_k} \cdot \sum_{i=0}^{mn} \frac{\prod_{k=i-m+1}^{i-1} \mathcal{I}_k (\mathcal{I}_{i-m} + \mathcal{I}_i)}{\prod_{k=1-M}^{-1} \mathcal{I}_k} \\
&= \frac{n}{2} \cdot \frac{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i-1} \mathcal{I}_k (\mathcal{I}_{i-m} + \mathcal{I}_i)}{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i-1} \mathcal{I}_k} \\
&= \frac{n}{2} \cdot \tilde{\mathcal{I}}^n \cdot \frac{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i-1} \left(\frac{\mathcal{I}_k}{\tilde{\mathcal{I}}^n}\right) \left(\frac{\mathcal{I}_{i-m}}{\tilde{\mathcal{I}}^n} + \frac{\mathcal{I}_i}{\tilde{\mathcal{I}}^n}\right)}{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i-1} \frac{\mathcal{I}_k}{\tilde{\mathcal{I}}^n}} \\
&= \frac{\tilde{\mathcal{A}}^n}{2} \frac{\sum_{i=0}^{mn} \prod_{k=i-m}^{i-1} \mathscr{I}_k}{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i-1} \mathscr{I}_k} + \frac{\tilde{\mathcal{A}}^n}{2} \frac{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i} \mathscr{I}_k}{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i-1} \mathscr{I}_k} \\
&= \mathbb{A}(n,m) + \mathbb{B}(n,m), \qquad (6.12)
\end{aligned}$$

where

$$\mathbb{A}(n,m) = \frac{\tilde{\mathcal{A}}^n}{2} \frac{\sum_{i=0}^{mn} \prod_{k=i-m}^{i-1} \mathscr{I}_k}{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i-1} \mathscr{I}_k} \qquad (6.13)$$

and

$$\mathbb{B}(n,m) = \frac{\tilde{\mathcal{A}}^n}{2} \frac{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i} \mathscr{I}_k}{\sum_{i=0}^{mn} \prod_{k=i-m+1}^{i-1} \mathscr{I}_k}. \qquad (6.14)$$

We will show that $\lim_{n\to\infty} \lim_{m\to\infty} \mathbb{A}(n,m) \geq \frac{\tilde{\phi}}{2}$. Then by symmetry, so is $\mathbb{B}(n,m)$. Thus, the ϵ-optimality follows. Notice that $\mathbb{A}(n,m)$ has two types of products, $\prod_{k=i-m}^{i-1} \mathscr{I}_k$ and $\prod_{k=i-m+1}^{i-1} \mathscr{I}_k$. They are either upper bounded or approximated by some exponential functions. We give a rigorous proof for the first type (Thoerems 22 and 23). The proof for the second type is quite similar and so omitted. All the results regarding them are

summarized in Theorem 24.

Let

$$\mathcal{P}_n^m(i) = \prod_{k=i-m}^{i-1} \mathscr{I}_k.$$

The next theorem shows, for $\theta_i \notin \left(\frac{1}{2n}, 1-\frac{1}{2n}\right)$, $\mathcal{P}_n^m(i)$ is upper bounded by an exponential function.

Theorem 22. *There exists $\omega < 1$ such that for $\theta_i \in [0, \frac{1}{2n}] \cup [1-\frac{1}{2n}, 1]$, $\mathcal{P}_n^m(i) \leq \omega^m$.*

Lemma 17. *For $a, b \in [0,1]$, $a \cdot b \leq \frac{a+b}{2}$. In particular, for $k_1, k_2 \in \mathbb{Z}$, we have*

$$\mathscr{I}_{k_1}\mathscr{I}_{k_2} \leq \frac{1}{2}(\mathscr{I}_{k_1} + \mathscr{I}_{k_2}).$$

Proof. Obviously, $ab \leq \frac{a^2+b^2}{2} \leq \frac{a+b}{2}$. Notice that for all $k \in \mathbb{Z}$, $\mathscr{I}_k \leq 1$. So, in particular, we have $\mathscr{I}_{k_1}\mathscr{I}_{k_2} \leq \frac{1}{2}(\mathscr{I}_{k_1} + \mathscr{I}_{k_2})$. □

Proof of Theorem 22. We show the inequality holds for $\theta_i \in [0, \frac{1}{2n}]$, then the case that $\theta_i \in [1-\frac{1}{2n}, 1]$ follows by symmetry. For $-\frac{1}{n} \leq \theta_k \leq 0$ (namely, $-m \leq k \leq 0$),

$$\begin{aligned}\mathscr{I}_k = \frac{\mathcal{I}_k}{\tilde{\mathcal{I}}^n} = \frac{\int_{\theta_k}^{\theta_k+\frac{1}{n}} \phi^*}{\tilde{\mathcal{I}}^n} = \frac{\int_0^{\theta_k+\frac{1}{n}} \phi^*}{\tilde{\mathcal{I}}^n} = \frac{\int_0^{\frac{k}{mn}+\frac{1}{n}} \phi^*}{\tilde{\mathcal{I}}^n} \\ = \frac{\int_0^{\frac{1}{n}} \phi^* - \int_{\frac{k}{mn}+\frac{1}{n}}^{\frac{1}{n}} \phi^*}{\tilde{\mathcal{I}}^n} \leq 1 - \frac{-\frac{k}{mn} \cdot \underset{\sim}{\phi}}{1/n} = 1 + \frac{k}{m} \cdot \underset{\sim}{\phi}.\end{aligned}$$

Note that $0 \le \theta_i \le \frac{1}{2n}$ implies $0 \le i \le \frac{m}{2}$. Combining with Lemma 17, we have

$$\begin{aligned}
\mathcal{P}_n^m(i) &= \prod_{k=i-m}^{i-1} \mathscr{I}_k \le \prod_{k=i-m}^{0} \mathscr{I}_k \le \prod_{k=i-m}^{i-m/2} \mathscr{I}_k \\
&= \prod_{k=i-m}^{i-3m/4} \mathscr{I}_k \cdot \mathscr{I}_{(2i-3m/2)-k} \\
&\le \prod_{k=i-m}^{i-3m/4} \frac{1}{2}(\mathscr{I}_k + \mathscr{I}_{(2i-3m/2)-k}) \qquad \text{(by Lemma 17)} \\
&\le \prod_{k=i-m}^{i-3m/4} \frac{1}{2}\left(1 + \frac{k}{m}\underset{\sim}{\phi} + 1 + \frac{(2i-3m/2)-k}{m}\underset{\sim}{\phi}\right) \\
&= \prod_{k=i-m}^{i-3m/4} \left(1 + \frac{2i-3m/2}{2m}\underset{\sim}{\phi}\right) \le \left(1 + \frac{m-3m/2}{2m}\underset{\sim}{\phi}\right)^{\frac{m}{4}} = \left[\left(1 - \frac{1}{4}\underset{\sim}{\phi}\right)^{\frac{1}{4}}\right]^m.
\end{aligned}$$

By choosing $\omega = \left(1 - \frac{1}{4}\underset{\sim}{\phi}\right)^{1/4}$ which is strictly less than one, the proof is completed. □

Let $\mathscr{A}^n(x)$ denote the average value of $\ln \mathscr{I}^n$ upon $\left(x, x + \frac{1}{n}\right)$. In particular,

$$\mathscr{A}^n(x) = n \int_{x-1/n}^{x} \ln \mathscr{I}^n.$$

Furthermore, set

$$\mathscr{A}_m^n(i) = \mathscr{A}^n\left(\theta_i - \frac{1}{2mn}\right),$$

where the term $-\frac{1}{2mn}$ is a correction that ensures the gap defined in Eq (6.17) is $o(m)$.

Theorem 23 shows that when m is sufficiently large, for all $\theta_i^{mn} \in \left(\frac{1}{2n}, 1 - \frac{1}{2n}\right)$, $\mathcal{P}_n^m(i)$ can be approximated by $\mathscr{A}_m^n(i)$. More accurately, $\mathcal{P}_n^m(i) \to \exp\left(m \cdot \mathscr{A}_m^n(i)\right)$ uniformly for all $\theta_i^{mn} \in \left(\frac{1}{2n}, 1 - \frac{1}{2n}\right)$ as $m \to \infty$.

Theorem 23. *For a fixed n, there exists $K > 0$ such that for all $m > 0$, if $\theta_i^{mn} \in$*

$\left(\frac{1}{2n}, 1-\frac{1}{2n}\right)$,

$$\exp\left(-\frac{K}{m}\right) \leq \frac{\mathcal{P}_n^m(i)}{\exp\left(m\mathscr{A}_m^n(i)\right)} \leq \exp\left(\frac{K}{m}\right). \tag{6.15}$$

Proof. Notice that

$$\begin{aligned}\left|\ln\frac{\mathcal{P}_n^m(i)}{\exp\left(m\mathscr{A}_m^n(i)\right)}\right| &= \left|\ln\mathcal{P}_n^m(i) - \ln\exp\left(m\mathscr{A}_m^n(i)\right)\right| \\ &= \left|\sum_{k=i-m}^{i-1}\ln\mathscr{I}_k - mn\int_{\theta_{i-m}^{mn}-(2mn)^{-1}}^{\theta_i^{mn}-(2mn)^{-1}}\ln\mathscr{I}^n\right| \\ &= mn\left|\sum_{k=i-m}^{i-1}\frac{1}{mn}\ln\mathscr{I}_k - \int_{\theta_{i-m}^{mn}-(2mn)^{-1}}^{\theta_i^{mn}-(2mn)^{-1}}\ln\mathscr{I}^n\right|. \end{aligned}\tag{6.16}$$

Let $\mathcal{G}(m,n)$ denote the gap

$$\left|\sum_{k=i-m}^{i-1}\frac{1}{mn}\ln\mathscr{I}_k - \int_{\theta_{i-m}^{mn}-(2mn)^{-1}}^{\theta_i^{mn}-(2mn)^{-1}}\ln\mathscr{I}^n\right|. \tag{6.17}$$

Chui has proved that

$$\mathcal{G}(m,n) \leq \frac{K_1\cdot V_{I_i}\left(\left[\ln\mathscr{I}^n\right]'\right)}{m^2},$$

where K_1 is a positive constant and $V_{I_i}\left(\left[\ln\mathscr{I}^n\right]'\right)$ is the total variation of $\left[\ln\mathscr{I}^n\right]'$ on $I_i = \left(\theta_{i-m}^{mn}-\frac{1}{2mn}, \theta_i^{mn}-\frac{1}{2mn}\right)$ [10, Theorem 1.d]. So to prove the theorem, we only need to show $V_{I_i}\left(\left[\ln\mathscr{I}^n\right]'\right)$ is upper bounded by a constant for all $\theta_i^{mn} \in \left(\frac{1}{2n}, 1-\frac{1}{2n}\right)$.

Since the variation of a function upon an interval cannot exceed the one on its super-interval [26, Proposition 2.6.ii], it is sufficient to show that the total variation of $\left[\ln\mathscr{I}^n\right]'$ is bounded on $\left[-\frac{1}{2n}, 1-\frac{1}{2n}\right]$. Notice that $\left[\ln\mathscr{I}^n(x)\right]' = \frac{\phi^*\left(x+\frac{1}{n}\right)-\phi^*(x)}{\mathcal{I}^n(x)}$ almost everywhere [15, Corollary 3.33]. Since $\phi^*(x)$ is of bounded variation on $\left[-\frac{1}{2n}, 1+\frac{1}{2n}\right]$, so is the difference $\phi^*\left(x+\frac{1}{n}\right)-\phi^*(x)$ on $\left[-\frac{1}{2n}, 1-\frac{1}{2n}\right]$ [26, Exercise 2.5.i]. As $\mathcal{I}^n(x) = \int_x^{x+1/n}\phi^*$ on $\left[-\frac{1}{2n}, 1-\frac{1}{2n}\right]$, it is of bounded variation [15, Corollary 3.33].

Combined with the fact that $\mathcal{I}^n(x) \geq \frac{\phi}{2n} > 0$ for all $x \in \left[-\frac{1}{2n}, 1-\frac{1}{2n}\right]$, the variation of the quotient $[\ln \mathscr{I}^n(x)]' = \frac{\phi^*\left(x+\frac{1}{n}\right)-\phi^*(x)}{R^{(n)}(x)}$ on $\left[-\frac{1}{2n}, 1-\frac{1}{2n}\right]$ is bounded [26, Exercise 2.5.iii]. Let $K_2 = V_{\left[-\frac{1}{2n},1-\frac{1}{2n}\right]}\left([\ln \mathscr{I}^n]'\right)$. Then $\mathcal{G}(m,n) < \frac{K_1K_2}{m^2}$. And thus, Eq (6.16) becomes

$$\left|\ln \frac{\mathcal{P}_n^m(i)}{\exp\left(m \cdot \mathscr{A}_m^n(i)\right)}\right| = mn \cdot \mathcal{G}(m,n) < \frac{nK_1K_2}{m}.$$

Set $K = nK_1K_2$ and take the exponential on both sides. Eq (6.15) follows. □

Theorems 22 and 23 provide a systematical way to control or approximate $\prod_{k=i-m}^{i-1} \mathscr{I}_k$ for $i = 0, 1, \cdots, mn$. The trick is also applicable to $\prod_{k=i-m+1}^{i-1} \mathscr{I}_k$. As the proof is almost the same, we omit it and simply state the results. The results regarding $\prod_{k=i-m}^{i-1} \mathscr{I}_k$ are also restated.

Theorem 24. *For $n \in \mathbb{N}^*$, there exist $K > 0$ and $\omega < 1$ such that for all $m > 0$, if $\theta_i^{mn} \in (\frac{1}{2n}, 1-\frac{1}{2n})$,*

$$\exp\left(-\frac{K}{m}\right) \leq \frac{\prod_{k=i-m}^{i-1} \mathscr{I}_k}{\exp\left(m \cdot \mathscr{A}_m^n(i)\right)} \leq \exp\left(\frac{K}{m}\right), \text{ and} \tag{6.18}$$

$$\exp\left(-\frac{K}{m}\right) \leq \frac{\prod_{k=i-m+1}^{i-1} \mathscr{I}_k}{\exp\left((m-1) \cdot \mathscr{B}_m^n(i)\right)} \leq \exp\left(\frac{K}{m}\right), \tag{6.19}$$

where

$$\mathscr{A}_m^n(i) = \mathscr{A}^n\left(\theta_i^{mn} - (2mn)^{-1}\right) = n\int_{\theta_{i-m}^{mn}-(2mn)^{-1}}^{\theta_i^{mn}-(2mn)^{-1}} \ln \mathscr{I}^n$$

and

$$\mathscr{B}_m^n(i) = \frac{mn}{m-1}\int_{\theta_{i-m+1}^{mn}-(2mn)^{-1}}^{\theta_i^{mn}-(2mn)^{-1}} \ln \mathscr{I}^n.$$

If $\theta_i^{mn} \in [0, \frac{1}{2n}] \cup [1 - \frac{1}{2n}, 1]$, *we have*

$$\prod_{k=i-m}^{i-1} \mathscr{I}_k \leq \omega^m, \text{ and} \tag{6.20}$$

$$\prod_{k=i-m+1}^{i-1} \mathscr{I}_k \leq \omega^m. \tag{6.21}$$

Remark 7. *Eq (6.18) and (6.19) share* K *by choosing the maximum of their exclusive ones. Similarly, Eq (6.20) and (6.21) share* ω.

In Theorem 24, we handle the products $\prod_{k=i-m}^{i-1} \mathscr{I}_k$ and $\prod_{k=i-m+1}^{i-1} \mathscr{I}_k$ based on the location of θ_i. If we define

$$\mathcal{S}_{mn} = \left\{ i = 0, 1, \cdots, mn : \theta_i^{mn} \notin \left(\frac{1}{2n}, 1 - \frac{1}{2n} \right) \right\}$$

and

$$\tilde{\mathcal{S}}_{mn} = \left\{ i = 0, 1, \cdots, mn : \theta_i^{mn} \in \left(\frac{1}{2n}, 1 - \frac{1}{2n} \right) \right\},$$

according to Theorem 24, $\mathbb{A}(n, m)$ can be written as

$$\mathbb{A}(n, m) = \frac{\tilde{\mathcal{A}}^n}{2} \cdot \frac{\sum_{k \in \mathcal{S}_{mn}} \mathcal{U}\left(\omega^m\right) + \sum_{k \in \tilde{\mathcal{S}}_{mn}} \exp\left[m \mathscr{A}_m^n(k)\right]}{\sum_{k \in \mathcal{S}_{mn}} \mathcal{U}\left(\omega^m\right) + \sum_{k \in \tilde{\mathcal{S}}_{mn}} \exp\left[(m-1) \mathscr{B}_m^n(k)\right]}, \tag{6.22}$$

where $\mathcal{U}\left(\omega^m\right)$ denotes a term upper bounded by ω^m. Before analyzing how $\mathbb{A}(n, m)$, we will further simplify the expression by getting rid of $\mathscr{B}_m^n(k)$. This can be done by constructing a relation between $\mathscr{A}_m^n(k)$ and $\mathscr{B}_m^n(k)$. Ahead of this, we first discuss the value space of $\exp \mathscr{A}^n(x)$.

Theorem 25. *For a fixed* n *and* $x \in (\frac{1}{2n}, 1 - \frac{1}{2n})$, $\frac{\phi}{2} \leq \exp \mathscr{A}^n(x) \leq 1$.

Proof. For all $x \in (-\frac{1}{2n}, 1 - \frac{1}{2n})$, $\mathscr{I}^n(x) = \frac{\mathcal{I}^n(x)}{\tilde{\mathcal{I}}^n} \geq \frac{\frac{1}{2n} \cdot \phi}{\frac{1}{n}} = \frac{\phi}{2}$. By definition, $\mathscr{I}^n(x) \leq 1$. So, to sum up, $\frac{\phi}{2} \leq \mathscr{I}^n(x) \leq 1$. As $\exp \mathscr{A}^n(x)$ is defined in terms of $\mathscr{I}^n(x)$, for

$x \in (\frac{1}{2n}, 1-\frac{1}{2n})$, we have

$$\begin{aligned}\exp \mathscr{A}^n(x) &= \exp\left[n\int_{x-1/n}^{x} \ln \mathscr{I}^n\right] \geq \exp\left[n\int_{x-1/n}^{x} \ln\left(\inf_{x\in I} \mathscr{I}^n(x)\right)\right] \\ &= \inf_{x\in(-\frac{1}{2n},1-\frac{1}{2n})} \mathscr{I}^n(x) \geq \frac{\phi}{2},\end{aligned}$$

and

$$\exp \mathscr{A}^n(x) = \exp\left[n\int_{x-1/n}^{x} \ln \mathscr{I}^n\right] \leq \exp\left[n\int_{x-1/n}^{x} \ln 1\right] = \exp(\ln 1) = 1,$$

which together complete the proof. □

The next theorem constructs a relation between $\mathscr{A}_m^n(k)$ and $\mathscr{B}_m^n(k)$.

Theorem 26. *For a fixed n and a sufficiently large m, there exists a function $\mathcal{E}$ defined on $(0,1]$ such that for $\theta_i^{mn} \in (\frac{1}{2n}, 1-\frac{1}{2n})$,*

$$\exp\left[(m-1)\mathscr{B}_m^n(i)\right] \leq \exp\left[(m-1)\mathscr{A}_m^n(i)\right] \cdot \mathcal{E}(\mathscr{A}_m^n(i)), \tag{6.23}$$

where $\mathcal{E}(x) \leq \frac{2}{\phi}$ and $\lim_{x\to 1^-} \mathcal{E}(x) = \mathcal{E}(1) = 1$.

Remark 8. $\lim_{x\to 1^-} \mathcal{E}(x) = 1$ *means $\mathcal{E}(x)$ approaches to one as x approaches to one from left.*

Lemma 18. *For $c > 0$, function*

$$f(y) = \left[\frac{1}{2}(y+1)\right]^{\frac{1-y}{2c}}$$

defined on $(-1,1]$ is increasing and reaches its global maximum one at $x = 1$. Also, it has an increasing inverse function $f^{-1} : (0,1] \to (-1,1]$ such that

$$\lim_{y\to 1^-} f^{-1}(y) = f^{-1}(1) = 1.$$

Proof. Considering that ln is strictly increasing, to show f is increasing, we can show that $\ln f$ is increasing instead. In particular, we show its derivative $[\ln f(y)]'$ is non-negative on $(-1, 1]$. It is easy to check that

$$[\ln f(y)]' = \left[\frac{1-y}{2c} \cdot \ln\left(\frac{1}{2}(y+1)\right)\right]' = -\frac{1}{2c}\ln\left[\frac{1}{2}(y+1)\right] + \frac{1-y}{2c}\frac{1}{(y+1)}.$$

For the first term, since $y \in (-1, 1]$, $\frac{1}{2}(y+1) \leq 1$. So $\ln\left(\frac{1}{2}(y+1)\right) \leq 0$ and thus $-\ln\left(\frac{1}{2}(y+1)\right) \geq 0$. Besides, it is easy to check that the second term $\frac{1-y}{2c}\frac{1}{(y+1)}$ is also not less than zero upon $y \in (-1, 1]$. Therefore, we conclude that $[\ln f(y)]' \geq 0$. Hence, f is increasing.

The range of f can be easily found, which is $(0, 1]$. As the inverse function of a monotonic function always exists, f has an increasing inverse f^{-1} defined on $(0, 1]$. Obviously, f is continuous. Then, so is f^{-1}. Combined with the fact that $f(1) = 1$, $\lim_{y\to 1^-} f^{-1}(y) = f^{-1}(1) = 1$. □

Proof of Theorem 26. We first show that the difference between two consecutive $\mathscr{I}_k$ is roughly bounded by $\frac{1}{m}$. In particular,

$$|\mathscr{I}_k - \mathscr{I}_{k+1}| = \left|\frac{\int_{\theta_k^{mn}}^{\theta_k^{mn}+\frac{1}{n}} \phi^*}{\frac{1}{n}\tilde{\mathcal{A}}^n} - \frac{\int_{\theta_{k+1}^{mn}}^{\theta_{k+1}^{mn}+\frac{1}{n}} \phi^*}{\frac{1}{n}\tilde{\mathcal{A}}^n}\right| = \frac{\left|\int_{\theta_k^{mn}}^{\theta_{k+1}^{mn}} \phi^* - \int_{\theta_k^{mn}+\frac{1}{n}}^{\theta_{k+1}^{mn}+\frac{1}{n}} \phi^*\right|}{\frac{1}{n}\tilde{\mathcal{A}}^n} \leq \frac{\frac{1}{mn}\cdot\tilde{\phi}}{\frac{1}{n}\tilde{\mathcal{A}}^n} = \frac{\tilde{\phi}}{\tilde{\mathcal{A}}^n}\cdot\frac{1}{m}.$$

Let $c = \frac{\tilde{\phi}}{\tilde{\mathcal{A}}^n}$. Then $\prod_{k=i-m}^{i-1}\mathscr{I}_k$ is upper bounded by a function of $\mathscr{I}_{i-m}$. In particular,

according to Lemma 17,

$$\prod_{k=i-m}^{i-1} \mathscr{I}_k \leq \underbrace{\mathscr{I}_{i-m} \cdot \left(\mathscr{I}_{i-m} + \frac{c}{m}\right) \cdot \left(\mathscr{I}_{i-m} + \frac{2c}{m}\right) \cdots s}_{\frac{s-\mathscr{I}_{i-m}}{c/m}+1 \text{ terms}}$$
$$\leq \left(\frac{1}{2}(\mathscr{I}_{i-m} + s)\right)^{\frac{s-\mathscr{I}_{i-m}}{2c/m}} \leq \left(\frac{1}{2}(\mathscr{I}_{i-m} + 1)\right)^{\frac{1-\mathscr{I}_{i-m}}{2c/m}}, \tag{6.24}$$

where $s = \sup\{\mathscr{I}_{i-m} + \frac{kc}{m} \leq 1 : k \in \mathbb{N}^*\}$. By rearranging the left side of Eq (6.18), we have

$$\exp[m \cdot \mathscr{A}_m^n(i)] \leq \exp\left(\frac{K}{m}\right) \prod_{k=i-m}^{i-1} \mathscr{I}_k.$$

Combining it with Eq (6.24), we get

$$\exp[m \cdot \mathscr{A}_m^n(i)] \leq \exp\left(\frac{K}{m}\right) \prod_{k=i-m}^{i-1} \mathscr{I}_k \leq \exp\left(\frac{K}{m}\right) \left(\frac{1}{2}(\mathscr{I}_{i-m} + 1)\right)^{\frac{1-\mathscr{I}_{i-m}}{2c/m}}.$$

When m is sufficiently large,

$$\exp[\mathscr{A}_m^n(i)] \leq \exp\left(\frac{K}{m^2}\right) \cdot \left(\frac{1}{2}(\mathscr{I}_{i-m} + 1)\right)^{\frac{1-\mathscr{I}_{i-m}}{2c}}$$
$$\approx \left(\frac{1}{2}(\mathscr{I}_{i-m} + 1)\right)^{\frac{1-\mathscr{I}_{i-m}}{2c}} = f(\mathscr{I}_{i-m}), \tag{6.25}$$

where f is defined in Lemma 18. Theorem 25 shows $x \in (\frac{1}{2n}, 1 - \frac{1}{2n})$ implies $\exp \mathscr{A}^n(x) \in [\frac{\phi}{2}, 1]$. So for $\theta_i^{mn} \in (\frac{1}{2n}, 1 - \frac{1}{2n})$, $\exp[\mathscr{A}_m^n(i)] \in [\frac{\tilde{\phi}}{2}, 1]$ and thus in the domain of f^{-1}. So we can take f^{-1} on the both sides of Eq (6.25) and get

$$f^{-1}\left(\exp[\mathscr{A}_m^n(i)]\right) \leq f^{-1}(f(\mathscr{I}_{i-m})) = \mathscr{I}_{i-m}.$$

Let $g(x) = \max\{f^{-1}(x), \frac{\phi}{2}\}$. Since $\lim_{x \to 1^-} f^{-1}(x) = f^{-1}(1) = 1 > \frac{\phi}{2}$, then $\lim_{x \to 1^-} g(x) =$

$g(1) = 1$. Combined with the fact that for all $x \in (-\frac{1}{2n}, 1 - \frac{1}{2n})$, $\mathscr{I}^n(x) = \frac{\mathcal{I}^n(x)}{\tilde{\mathcal{I}}^n} \geq \frac{\frac{1}{2n} \cdot \underset{\sim}{\phi}}{\frac{1}{n}} = \frac{\underset{\sim}{\phi}}{2}$, neither $\frac{\underset{\sim}{\phi}}{2}$ nor $f^{-1}\left(\exp[\mathscr{A}_m^n(i)]\right)$ is greater than $\mathscr{I}_{i-m}$. So $0 < \frac{\underset{\sim}{\phi}}{2} \leq g(\exp[\mathscr{A}_m^n(i)]) \leq \mathscr{I}_{i-m}$. By Eqs (6.18) and (6.19),

$$\frac{\exp[(m-1)\cdot\mathscr{B}_m^n(i)]}{\exp[(m-1)\cdot\mathscr{A}_m^n(i)]} \cdot \frac{\mathscr{I}_{i-m}}{\exp[\mathscr{A}_m^n(i)]} = \frac{\exp[(m-1)\cdot\mathscr{B}_m^n(i)] \cdot \prod_{k=i-m}^{i-1} \mathscr{I}_k}{\exp[m\cdot\mathscr{A}_m^n(i)] \prod_{k=i-m+1}^{i-1} \mathscr{I}_k} \leq \exp\left(\frac{2K}{m}\right).$$

Then

$$\frac{\exp[(m-1)\cdot\mathscr{B}_m^n(i)]}{\exp[(m-1)\cdot\mathscr{A}_m^n(i)]} \leq \exp\left(\frac{2K}{m}\right) \cdot \frac{\exp[\mathscr{A}_m^n(i)]}{\mathscr{I}_{i-m}} \leq \exp\left(\frac{2K}{m}\right) \cdot \frac{\exp[\mathscr{A}_m^n(i)]}{g\left(\exp[\mathscr{A}_m^n(i)]\right)}. \quad (6.26)$$

Let $\mathcal{E}(x) = \frac{x}{g(x)}$ on $(0, 1]$. Then

$$\lim_{x\to 1^-} \mathcal{E}(x) = \frac{1}{\lim_{x\to 1^-} g(x)} = \frac{1}{g(1)} = 1 = \mathcal{E}(1) \qquad \& \qquad \mathcal{E}(x) \leq \frac{1}{\underset{\sim}{\phi}/2} = \frac{2}{\underset{\sim}{\phi}}.$$

When m is sufficiently large, Eq (6.26) reduces to

$$\frac{\exp[(m-1)\cdot\mathscr{B}_m^n(i)]}{\exp[(m-1)\cdot\mathscr{A}_m^n(i)]} \leq \frac{\exp[\mathscr{A}_m^n(i)]}{g\left(\exp[\mathscr{A}_m^n(i)]\right)} = \mathcal{E}(\mathscr{A}_m^n(i)).$$

Thus, Eq (6.23) follows. □

Recall Eq (6.22). By Theorem 26, we have

$$\mathbb{A}(n,m) \geq \frac{\tilde{\mathcal{A}}^n}{2} \cdot \frac{\sum_{k\in\mathcal{S}_{mn}} \mathcal{U}\left(\omega^m\right) + \sum_{k\in\tilde{\mathcal{S}}_{mn}} \exp\left[m\mathscr{A}_m^n(k)\right]}{\sum_{k\in\mathcal{S}_{mn}} \mathcal{U}\left(\omega^m\right) + \sum_{k\in\tilde{\mathcal{S}}_{mn}} \exp\left[(m-1)\mathscr{A}_m^n(k)\right] \cdot \mathcal{E}(\mathscr{A}_m^n(k))}. \quad (6.27)$$

The following theorem indicates that $\sup_{x\in[0,1]} \exp \mathscr{A}^n(x)$ is lower bounded and arbitrarily close to one as n goes to infinity.

Theorem 27. *For all $\epsilon > 0$. There exists some $N > 0$ such that for $n > N$,*

$$\tilde{\mathcal{A}}^n = n\tilde{\mathcal{I}}^n \geq \tilde{\phi} - \epsilon \qquad \& \qquad \sup_{x\in[0,1]} \exp \mathscr{A}^n(x) \geq \frac{\tilde{\phi} - \epsilon}{\tilde{\phi}}.$$

Lemma 19. *For any functions g_1 and g_2 defined on $[0, 1]$, we have*

$$\sup_{x\in[0,1]} |g_1 - g_2| \geq \left| \sup_{x\in[0,1]} g_1 - \sup_{x\in[0,1]} g_2 \right|. \tag{6.28}$$

Proof. Notice that $\sup_{x\in[0,1]} |f| \geq \sup_{x\in[0,1]} f$ for any f defined on $[0, 1]$. Combining with Lemma 14, we have

$$\sup_{x\in[0,1]} |g_1 - g_2| \geq \sup_{x\in[0,1]} (g_1 - g_2) \geq \sup_{x\in[0,1]} g_1 - \sup_{x\in[0,1]} g_2.$$

By symmetry, we also have

$$\sup_{x\in[0,1]} |g_1 - g_2| \geq \sup_{x\in[0,1]} g_2 - \sup_{x\in[0,1]} g_1.$$

Thus, Eq (6.28) follows. □

Lemma 20. *Pick $\epsilon > 0$. There exists $N > 0$ such that for all $n > N$, $\sup_{x\in[0,1]} \frac{n}{2} \cdot \mathcal{I}\left(x, \frac{2}{n}\right) > \tilde{\phi} - \epsilon$.*

Proof. If n is even, then $n = 2k$ for $k \in \mathbb{N}$. So $\frac{n}{2} = k$. By Theorem 18, there exists $N_1 \in \mathbb{N}$ that $k > N_1$ implies

$$\sup_{x\in[0,1]} k\mathcal{I}\left(x, \frac{1}{k}\right) = \sup_{x\in[0,1]} k\mathcal{I}^k(x) = \tilde{\mathcal{A}}^k > \tilde{\phi} - \frac{\epsilon}{2}. \tag{6.29}$$

By definition, $\tilde{\phi} \geq \tilde{\mathcal{A}}^k$. So $\left|\tilde{\mathcal{A}}^k - \tilde{\phi}\right| < \frac{\epsilon}{2}$. If n is odd, then $n = 2k+1$. By Lemma 19,

$$\begin{aligned}
&\left|\sup_{x\in[0,1]} \frac{2k+1}{2} \cdot \mathcal{I}\left(x, \frac{2}{2k+1}\right) - \tilde{\mathcal{A}}^k\right| \\
\leq &\sup_{x\in[0,1]} \left|\frac{2k+1}{2}\mathcal{I}\left(x, \frac{2}{2k+1}\right) - k\mathcal{I}\left(x, \frac{1}{k}\right)\right| \\
= &\sup_{x\in[0,1]} \left|\left[\frac{2k+1}{2}\mathcal{I}\left(x, \frac{2}{2k+1}\right) - k\mathcal{I}\left(x, \frac{2}{2k+1}\right)\right] + \left[k\mathcal{I}\left(x, \frac{2}{2k+1}\right) - k\mathcal{I}\left(x, \frac{1}{k}\right)\right]\right| \\
\leq &\sup_{x\in[0,1]} \left[\left|\frac{1}{2} \cdot \mathcal{I}\left(x, \frac{2}{2k+1}\right)\right| + \left|k\int_{x+\frac{1}{k}}^{x+\frac{2}{2k+1}} \phi^*\right|\right] \\
\leq &\frac{1}{2} \cdot \frac{2}{2k+1} + k \cdot \left|\frac{2}{2k+1} - \frac{1}{k}\right| = \frac{2}{2k+1}.
\end{aligned}$$

So there exists $N_2 \in \mathbb{N}$ such that $\forall k \geq N_2$, $\left|\sup_{x\in[0,1]} \frac{2k+1}{2} \cdot \mathcal{I}\left(x, \frac{2}{2k+1}\right) - \tilde{\mathcal{A}}^k\right| < \frac{\epsilon}{2}$. Then for $k > \max\{N_1, N_2\}$,

$$\begin{aligned}
&\left|\sup_{x\in[0,1]} \frac{2k+1}{2} \cdot \mathcal{I}\left(x, \frac{2}{2k+1}\right) - \tilde{\phi}\right| \\
= &\left|\left[\sup_{x\in[0,1]} \frac{2k+1}{2} \cdot \mathcal{I}\left(x, \frac{2}{2k+1}\right) - \tilde{\mathcal{A}}^k\right] + \left[\tilde{\mathcal{A}}^k - \tilde{\phi}\right]\right| \\
\leq &\left|\sup_{x\in[0,1]} \frac{2k+1}{2} \cdot \mathcal{I}\left(x, \frac{2}{2k+1}\right) - \tilde{\mathcal{A}}^k\right| + \left|\tilde{\mathcal{A}}^k - \tilde{\phi}\right| \\
\leq &\frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

So we conclude that

$$\sup_{x\in[0,1]} \frac{2k+1}{2} \cdot \mathcal{I}\left(x, \frac{2}{2k+1}\right) \geq \tilde{\phi} - \epsilon. \tag{6.30}$$

Combined with Eq (6.29) for n is even, the theorem follows by choosing $N = 2\max\{N_1, N_2\}$.

□

Lemma 21. *Pick $\epsilon > 0$. Suppose that for some m and n, there exists $i \in \mathbb{Z}$ such that $\frac{n}{2} \cdot \mathcal{I}\left((\theta_{i-m}, \theta_{i+m})\right) \geq \tilde{\phi} - \frac{\epsilon}{2}$. Then for $\theta_{i-m} \leq \theta_j \leq \theta_i$, we have $n\mathcal{I}_j \geq \tilde{\phi} - \epsilon$.*

Proof. Figure 6.2 demostrates the relations among θ_{i-m}, θ_j, θ_i, θ_{j+m} and θ_{i+m}.

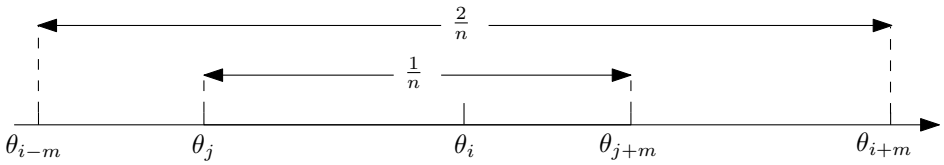


Figure 6.2: Relations among θ_{i-m}, θ_j, θ_i, θ_{j+m} and θ_{i+m}.

Assume the opposite is true. There exists some θ_j such that $n\mathcal{I}_j < \tilde{\phi} - \epsilon$. As

$$\begin{aligned}\mathcal{I}\left((\theta_{i-m},\theta_j)\cup(\theta_{j+m},\theta_{i+m})\right) &= \left(|\theta_{i-m}-\theta_j|+|\theta_{j+m}-\theta_{i+m}|\right)\cdot\mathcal{A}((\theta_{i-m},\theta_j)\cup(\theta_{j+m},\theta_{i+m})) \\ &= \frac{1}{n}\cdot\mathcal{A}((\theta_{i-m},\theta_j)\cup(\theta_{j+m},\theta_{i+m}))\,,\end{aligned}$$

then

$$\begin{aligned}\mathcal{A}((\theta_{i-m},\theta_j)\cup(\theta_{j+m},\theta_{i+m})) &= n\mathcal{I}\left((\theta_{i-m},\theta_j)\cup(\theta_{j+m},\theta_{i+m})\right) \\ &= n\left[\mathcal{I}\left((\theta_{i-m},\theta_{i+m})\right)-\mathcal{I}\left((\theta_j,\theta_{j+m})\right)\right] \\ &= n\mathcal{I}\left((\theta_{i-m},\theta_{i+m})\right)-n\mathcal{I}_j \geq 2\left(\tilde{\phi}-\frac{\epsilon}{2}\right)-(\tilde{\phi}-\epsilon)=\tilde{\phi}.\end{aligned}$$

Lemma 13 shows at least one of $\mathcal{A}((\theta_{i-m},\theta_j))$ and $\mathcal{A}((\theta_{j+m},\theta_{i+m}))$ is greater than $\tilde{\phi}$, which contradicts to Corollary 4. □

Proof of Theorem 27. Pick $\epsilon > 0$ arbitrary. By Lemma 20, there exists $N \in \mathbb{N}$ such that for all $n > N$,

$$\sup_{x\in[0,1]}\frac{n}{2}\cdot\mathcal{I}\left(x,\frac{2}{n}\right) > \tilde{\phi}-\frac{\epsilon}{4}.$$

Pick such n. Since $\frac{n}{2}\cdot\mathcal{I}\left(x,\frac{2}{n}\right)$ is continuous, there is $x_0 \in [0,1]$ that gives the supremum and some $\delta > 0$ such that for $x \in (x_0-\delta, x_0+\delta)$, $\frac{n}{2}\cdot\mathcal{I}\left(x,\frac{2}{n}\right) > \tilde{\phi}-\frac{\epsilon}{2}$. Choose a big enough m so that there is $\theta^{mn}_{i-m} \in \Theta^{mn}$ contained by $(x_0-\delta, x_0+\delta)$. Then we have $\frac{n}{2}\cdot\mathcal{I}\left(\theta^{mn}_{i-m},\frac{2}{n}\right) > \tilde{\phi}-\frac{\epsilon}{2}$. Namely, $\frac{n}{2}\cdot\mathcal{I}\left((\theta^{mn}_{i-m},\theta^{mn}_{i+m})\right) > \tilde{\phi}-\frac{\epsilon}{2}$. According to

Lemma 21, for $\theta_{i-m}^{mn} \le \theta_j^{mn} \le \theta_i^{mn}$, $n\mathcal{I}_j \ge \tilde{\phi} - \epsilon$, and thus the first equation follows. Also, $\mathscr{I}_j = \frac{\mathcal{I}_j}{\bar{\mathcal{I}}^n} \ge \frac{(1/n)(\tilde{\phi}-\epsilon)}{(1/n)\cdot\tilde{\phi}} = \frac{\tilde{\phi}-\epsilon}{\tilde{\phi}}$. When m is sufficiently large, according to Theorem 24, $\exp\left(m \cdot \mathscr{A}_m^n(i)\right) = \prod_{j=i-m}^{i-1} \mathscr{I}_j \ge \left(\frac{\tilde{\phi}-\epsilon}{\tilde{\phi}}\right)^m$. Therefore,

$$\sup_{x\in[0,1]} \exp \mathscr{A}^n(x) \ge \exp \mathscr{A}_m^n(i) \ge \frac{\tilde{\phi}-\epsilon}{\tilde{\phi}}.$$

□

The next theorem completes most of the the work for the proof that G-LSCALA is ϵ-optimal.

Theorem 28. $\lim_{n\to\infty}\lim_{m\to\infty} \mathbb{A}(n,m) \ge \frac{\tilde{\phi}}{2}$ *and* $\lim_{n\to\infty}\lim_{m\to\infty} \mathbb{B}(n,m) \ge \frac{\tilde{\phi}}{2}$.

Lemma 22. *Suppose that there is a series of functions* $\{g_i(k)\}_{i\ge1}$ *defined on* $\mathbb{N}^*$ *such that for all* i, $g_i(k) \le p^k$. *Then for* $\epsilon > 0$ *and* $t_1, t_2, t_3 \in \mathbb{N}$, *we have*

$$\sum_{i=1}^{t_1 k} t_2 \cdot g_i(k - t_3) = o((p+\epsilon)^k).$$

Proof.

$$\lim_{k\to\infty} \frac{\sum_{i=1}^{t_1 k} t_2 \cdot g_i(k-t_3)}{(p+\epsilon)^k} \le \lim_{k\to\infty} \frac{(t_1 k) t_2 p^{k-t_3}}{(p+\epsilon)^k}$$
$$= \lim_{k\to\infty} \frac{(t_1 k) t_2 p^{-t_3}}{((p+\epsilon)/p)^k} = \lim_{k\to\infty} \frac{t_1 t_2 p^{-t_3}}{k((p+\epsilon)/p)^{k-1}} = 0.$$

Note that second last equation is due to L'Hospital's rule. □

Proof of Theorem 28. Recall Eq (6.27). Let

$$h(m) = \frac{\sum_{k\in\mathcal{S}_{mn}} \mathcal{U}\left(\omega^m\right) + \sum_{k\in\tilde{\mathcal{S}}_{mn}} \exp\left[m\mathscr{A}_m^n(k)\right]}{\sum_{k\in\mathcal{S}_{mn}} \mathcal{U}\left(\omega^m\right) + \sum_{k\in\tilde{\mathcal{S}}_{mn}} \exp\left[(m-1)\mathscr{A}_m^n(k)\right] \cdot \mathcal{E}(\mathscr{A}_m^n(k))}.$$

Then Eq (6.27) can be written as $\mathbb{A}(n,m) \geq \frac{\tilde{A}^n}{2} h(m)$. Pick $0 < \epsilon < 1-\omega$. By Theorem 26, as $\lim_{x\to 1^-} \mathcal{E}(x) = 1$, there exists $\delta > 0$ such that $0 < 1-x < \delta$ implies $|\mathcal{E}(x) - 1| < \epsilon$. Since $\mathcal{E}(1) = 1$,

$$0 \leq 1 - x < \delta \Longrightarrow \mathcal{E}(x) < 1 + \epsilon. \tag{6.31}$$

Pick $0 < \epsilon' < \min(\epsilon, \delta)$. Partition $\tilde{\mathcal{S}}_{mn}$ into three parts:

$$\begin{aligned}
\underset{\sim}{T}_{\epsilon'}(m) &= \left\{ i \in \tilde{\mathcal{S}}_{mn} : \exp \mathscr{A}_m^n(i) \leq 1 - \epsilon' \right\}, \\
\tilde{\underset{\sim}{T}}_{\epsilon'}(m) &= \left\{ i \in \tilde{\mathcal{S}}_{mn} : 1 - \epsilon' < \exp \mathscr{A}_m^n(i) \leq 1 - \frac{\epsilon'}{2} \right\}, \text{ and} \\
\tilde{T}_{\epsilon'}(m) &= \left\{ i \in \tilde{\mathcal{S}}_{mn} : 1 - \frac{\epsilon'}{2} < \exp \mathscr{A}_m^n(i) \right\}.
\end{aligned}$$

Then $h(m)$ can be written as

$$\begin{aligned}
h(m) = &\left(\sum_{k \in \underset{\sim}{\mathcal{S}}_{mn}} \mathcal{U}(\omega^m) + \sum_{k \in \underset{\sim}{T}_{\epsilon'}(m)} \exp\left[m\mathscr{A}_m^n(k)\right] + \sum_{k \in \tilde{\underset{\sim}{T}}_{\epsilon'}(m)} \exp\left[m\mathscr{A}_m^n(k)\right] + \sum_{k \in \tilde{T}_{\epsilon'}(m)} \exp\left[m\mathscr{A}_m^n(k)\right] \right) \Bigg/ \\
&\left(\sum_{k \in \underset{\sim}{\mathcal{S}}_{mn}} \mathcal{U}(\omega^m) + \sum_{k \in \underset{\sim}{T}_{\epsilon'}(m)} \exp\left[(m-1)\mathscr{A}_m^n(k)\right] \mathcal{E}(\exp \mathscr{A}_m^n(k)) + \right. \\
&\left. \sum_{k \in \tilde{\underset{\sim}{T}}_{\epsilon'}(m)} \exp\left[(m-1)\mathscr{A}_m^n(k)\right] \mathcal{E}(\exp \mathscr{A}_m^n(k)) + \sum_{k \in \tilde{T}_{\epsilon'}(m)} \exp\left[(m-1)\mathscr{A}_m^n(k)\right] \mathcal{E}(\exp \mathscr{A}_m^n(k)) \right) \\
= &\left(\sum_{k \in \underset{\sim}{\mathcal{S}}_{mn} \cup \underset{\sim}{T}_{\epsilon'}(m)} \mathcal{U}(\omega^m) + \sum_{k \in \tilde{\underset{\sim}{T}}_{\epsilon'}(m)} \exp\left[m\mathscr{A}_m^n(k)\right] + \sum_{k \in \tilde{T}_{\epsilon'}(m)} \exp\left[m\mathscr{A}_m^n(k)\right] \right) \Bigg/ \\
&\left(\sum_{k \in \underset{\sim}{\mathcal{S}}_{mn}} \mathcal{U}(\omega^m) + \sum_{k \in \underset{\sim}{T}_{\epsilon'}(m)} \exp\left[(m-1)\mathscr{A}_m^n(k)\right] \mathcal{E}(\exp \mathscr{A}_m^n(k)) + \right. \\
&\left. \sum_{k \in \tilde{\underset{\sim}{T}}_{\epsilon'}(m)} \exp\left[(m-1)\mathscr{A}_m^n(k)\right] \mathcal{E}(\exp \mathscr{A}_m^n(k)) + \sum_{k \in \tilde{T}_{\epsilon'}(m)} \exp\left[(m-1)\mathscr{A}_m^n(k)\right] \mathcal{E}(\exp \mathscr{A}_m^n(k)) \right)
\end{aligned}$$

Note that $\omega < 1 - \epsilon \leq 1 - \epsilon'$. According to Lemma 22, those terms which are sums of

$\mathcal{U}(\omega^m)$ are $o\left((1-\frac{\epsilon'}{2})^m\right)$. Recall Theorem 26, $\mathcal{E}(x) \leq \frac{2}{\tilde{\phi}}$. So

$$\sum_{k\in\tilde{T}_{\epsilon'}(m)} \exp\left[(m-1)\mathscr{A}_m^n(k)\right]\mathcal{E}(\exp\mathscr{A}_m^n(k)) \leq \frac{2}{\tilde{\phi}}\cdot\sum_{k\in\tilde{T}_{\epsilon'}(m)} \exp\left[(m-1)\mathscr{A}_m^n(k)\right],$$

which is also $o\left((1-\frac{\epsilon'}{2})^m\right)$ by Lemma 22. Therefore, $h(m)$ can be written as

$$h(m) = \frac{o\left((1-\frac{\epsilon'}{2})^m\right) + \sum_{k\in\tilde{T}_{\epsilon'}(m)\cup\tilde{T}_{\epsilon'}(m)} \exp\left[m\mathscr{A}_m^n(k)\right]}{o\left((1-\frac{\epsilon'}{2})^m\right) + \sum_{k\in\tilde{T}_{\epsilon'}(m)\cup\tilde{T}_{\epsilon'}(m)} \exp\left[(m-1)\mathscr{A}_m^n(k)\right]\mathcal{E}(\exp\mathscr{A}_m^n(k))}.$$

According to Theorem 27, there exists $N > 0$ such that, for all $n > N$, $\tilde{\mathcal{A}}^n > \tilde{\phi} - \frac{\epsilon'\tilde{\phi}}{2} > \tilde{\phi} - \frac{\epsilon'}{2} > \tilde{\phi} - \epsilon$ and $\sup_{x\in[0,1]} \exp\mathscr{A}^n(x) \geq \frac{\tilde{\phi}-\epsilon'\tilde{\phi}/2}{\tilde{\phi}} > 1 - \frac{\epsilon'}{2}$. For such n, since $\mathscr{A}^n(x)$ is continuous, there exists $x_0 \in [0,1]$ that $\mathscr{A}^n(x_0) = \sup_{x\in[0,1]} \mathscr{A}^n(x)$. As $m \to \infty$, $\inf\{|\theta_i^{mn} - x_0| : \theta_i^{mn} \in \Theta^{mn}\} \to 0$. So there exists $M > 0$ such that for all $m > M$,

$$\sup\{\exp\mathscr{A}_m^n(i) : \theta_i^{mn} \in \Theta^{mn}\} = \sup_{x\in[0,1]} \exp\mathscr{A}^n(x) > 1 - \frac{\epsilon'}{2}.$$

Therefore, for $n > N$ and $m > M$, $\tilde{T}_{\epsilon'}(m)$ is not empty. By picking such $n > N$, we have

$$\lim_{m\to\infty} h(m) = \lim_{m\to\infty} \frac{\sum_{k\in\tilde{T}_{\epsilon'}(m)\cup\tilde{T}_{\epsilon'}(m)} \exp\left[m\mathscr{A}_m^n(k)\right]}{\sum_{k\in\tilde{T}_{\epsilon'}(m)\cup\tilde{T}_{\epsilon'}(m)} \exp\left[(m-1)\mathscr{A}_m^n(k)\right]\mathcal{E}(\exp\mathscr{A}_m^n(k))}.$$

Notice that for $k \in \tilde{T}_{\epsilon'}(m)\cup\tilde{T}_{\epsilon'}(m)$, $\exp\mathscr{A}_m^n(k) > 1-\epsilon' \geq 1-\delta$. Besides, by Theorem 25, $\exp\mathscr{A}_m^n(k) \leq 1$. So we conclude that $0 \leq 1 - \exp\mathscr{A}_m^n(k) < \delta$, and thus $\mathcal{E}(\exp\mathscr{A}_m^n(k)) < 1+\epsilon$ by Eq (6.31). Then

$$\lim_{m\to\infty} h(m) > \lim_{m\to\infty} \frac{1-\epsilon'}{1+\epsilon}\cdot\frac{\sum_{k\in\tilde{T}_{\epsilon'}(m)\cup\tilde{T}_{\epsilon'}(m)} \exp\left[(m-1)\mathscr{A}_m^n(k)\right]}{\sum_{k\in\tilde{T}_{\epsilon'}(m)\cup\tilde{T}_{\epsilon'}(m)} \exp\left[(m-1)\mathscr{A}_m^n(k)\right]} = \frac{1-\epsilon'}{1+\epsilon} > \frac{1-\epsilon}{1+\epsilon}.$$

So

$$\lim_{n\to\infty}\lim_{m\to\infty} \mathbb{A}(n,m) \geq \lim_{n\to\infty}\lim_{m\to\infty} \frac{\tilde{\mathcal{A}}^n}{2} h(m) > \frac{\tilde{\phi}-\epsilon}{2}\cdot\frac{1-\epsilon}{1+\epsilon}.$$

Since ϵ can be arbitrarily small, we conclude that

$$\lim_{n\to\infty}\lim_{m\to\infty}\mathbb{A}(n,m)\geq\frac{\tilde{\phi}}{2}.$$

By symmetry, we also have

$$\lim_{n\to\infty}\lim_{m\to\infty}\mathbb{B}(n,m)\geq\frac{\tilde{\phi}}{2}.$$

□

Theorem 29 (ϵ-optimal)**.** $\lim_{n\to\infty}\lim_{m\to\infty}Rw_n(m)=\tilde{\phi}$.

Proof. According to Theorem 21, $Rw_n(m)\leq\tilde{\phi}$. Besides,

$$\begin{aligned}\lim_{n\to\infty}\lim_{m\to\infty}Rw_n(m)&=\lim_{n\to\infty}\lim_{m\to\infty}\mathbb{A}(n,m)+\lim_{n\to\infty}\lim_{m\to\infty}\mathbb{B}(n,m) && \text{(by Eq (6.12))}\\ &\geq\frac{\tilde{\phi}}{2}+\frac{\tilde{\phi}}{2}=\tilde{\phi}. && \text{(by Theorem 28)}\end{aligned}$$

So we have

$$\tilde{\phi}\leq\lim_{n\to\infty}\lim_{m\to\infty}Rw_n(m)\leq\tilde{\phi},$$

which implies that

$$\lim_{n\to\infty}\lim_{m\to\infty}Rw_n(m)=\tilde{\phi}.$$

□

Notice that G-LSCALA algorithm uses CPDs to take actions. Since we have showed that the LA implementing G-LSCALA algorithm has the long-term reward probability arbitrarily close to $\tilde{\phi}$ by pushing m and n to infinity, combining with Theorem 19, we conclude that

Corollary 5. *For CALA taking actions according to CPDs, $\tilde{\phi}$ is the least upper bound of the reward probability they can possibly achieve.*

6.5 Generalization

In this section, we propose a MG-LSCALA algorithm that is ϵ-optimal without assuming that the reward function satisfies **Req.** 1.

The modification is made as follows. Given a reward function ϕ that only satisfies **Req.** 2, we can adapt it to satisfy **Req.** 1 by adding a subsistent reward probability $c \in (0, 1)$. In more details, when the LA gets a penalty by taking $\lambda \in [0, 1]$, it still has probability c to consider it as a reward. Then the reward probability by taking λ becomes

$$\psi(\lambda) = \phi(\lambda) + c \cdot (1 - \phi(\lambda)) = (1 - c) \cdot \phi(\lambda) + c, \tag{6.32}$$

which is lower bounded by $c > 0$. In the sequel, for a LA running the modified algorithm, we call the probability of a reward given by the environment the objective reward probability (specified by ϕ) and the one (added a subsistent reward probability c) determining the updates of the kernel the subjective reward probability (specified by ψ). Let $\tilde{\psi}$ denote the ORP of ψ.

Remark 9. *The objective reward probability is the one we care about and thus the one we try to maximize. In contrast, the subjective reward probability is merely an auxiliary concept that allows us to use the results having been proved in Sections 6.2-6.4.2.*

Let LA_m denote the LA implementing the modified algorithm and interacting with ϕ. LA_o implements the original algorithm and functioning with ψ. All characteristics of LA_m and LA_o are same because the transitions of their kernels are both guided by ψ (the reward probability of LA_o is the subjective reward probability of LA_m). So we can investigate LA_m through LA_o whose properties have been studied.

The subjective reward probability ψ is always lower bounded by c and thus satisfies **Req.** 1. Any constant function has variation zero on any interval [26, Proposition 2.10].

Thus, if ϕ is of bounded variation on $[0,1]$, so is $\psi = (1-c)\phi + c$ [26, Exercise 2.5]. Hereafter, we assume that ϕ satisfies **Req.** 2 only. Then ψ satisfies Assumption 6, and LA_o can be characterized by the theorems proved in Sections 6.2-6.4.2.

Let $Rw_n^o(m)$ and $Rw_n^s(m)$ denote the long-term gross objective and subjective reward probabilities of LA_m, respectively. Then $Rw_n^s(m)$ is also the long-term gross reward probability of LA_o. Our main goal is to show that the modified algorithm is also ϵ-optimal. Namely,

$$\lim_{n\to\infty}\lim_{m\to\infty} Rw_n^o(m) = \tilde{\phi}. \tag{6.33}$$

Recall Eq (6.6). For LA_m, $\mathcal{K}_i = \int_{\theta_i-\frac{1}{n}}^{\theta_i+\frac{1}{n}} \frac{n}{2}\phi^*$ is the objective reward probability when $\mathbf{id} = \theta_i$. Let $\mathscr{K}_i = \int_{\theta_i-\frac{1}{n}}^{\theta_i+\frac{1}{n}} \frac{n}{2}\psi^*$ be its subjective counterpart. For both LA_m and LA_o, let π^{c*} denote the stationary distributions of $\mathbf{id}$ and π_i^{c*} the probability that $\mathbf{id} = \theta_i$ in it. Let $\mathcal{T}_c(x) = (1-c)x + c$. Then Eq (6.32) can be written as

$$\psi(\lambda) = \mathcal{T}_c(\phi(\lambda)). \tag{6.34}$$

Recall Eq (6.1). We extend the domain of ψ to $\mathbb{R}$ in the same way and get ψ^*. If $\lambda \in [0,1]$, we have $\psi^*(\lambda) = \psi(\lambda) = \mathcal{T}_c(\phi(\lambda)) = \mathcal{T}_c(\phi^*(\lambda))$. Otherwise, we have $\psi^*(\lambda) = 0 < c = \mathcal{T}_c(\phi^*(\lambda))$. To sum up, $\psi^*(\lambda) \le \mathcal{T}_c(\phi^*(\lambda))$, and thus

Lemma 23. $\mathscr{K}_i \le (1-c)\mathcal{K}_i + c$.

Proof. $\mathscr{K}_i = \int_{\theta_i-\frac{1}{n}}^{\theta_i+\frac{1}{n}} \frac{n}{2}\psi^* \le \int_{\theta_i-\frac{1}{n}}^{\theta_i+\frac{1}{n}} \frac{n}{2}\mathcal{T}_c(\phi^*(\lambda)) = (1-c)\left[\int_{\theta_i-\frac{1}{n}}^{\theta_i+\frac{1}{n}} \frac{n}{2}\phi^*\right] + c = (1-c)\mathcal{K}_i + c.$ □

Since $\mathcal{T}_c$ is trivially bijective, then its inverse $\mathcal{T}_c^{-1}$ exists. Besides, as $\mathcal{T}_c$ is strictly increasing , so is $\mathcal{T}_c^{-1}$. Then we have,

Theorem 30. $Rw_n^o(m) \ge \mathcal{T}_c^{-1}(Rw_n^s(m))$.

Proof. Recall Eq (6.11), $Rw_n(m) = \sum_{i=0}^{mn} \pi_i^* \mathcal{K}_i$. Correspondingly, we have

$$Rw_n^o(m) = \sum_{i=0}^{mn} \pi_i^{c*} \mathcal{K}_i \tag{6.35}$$

and

$$Rw_n^s(m) = \sum_{i=0}^{mn} \pi_i^{c*} \mathscr{K}_i.$$

Then

$$\begin{aligned} Rw_n^s(m) &= \sum_{i=0}^{mn} \pi_i^{c*} \mathscr{K}_i \leq \sum_{i=0}^{mn} \pi_i^{c*} \left[(1-c)\mathcal{K}_i + c\right] && \text{(Lemma 23)} \\ &= (1-c)\left[\sum_{i=0}^{mn} \pi_i^{c*} \mathcal{K}_i\right] + c = (1-c) Rw_n^o(m) + c = \mathcal{T}_c\left(Rw_n^o(m)\right). \end{aligned}$$

Therefore, $Rw_n^o(m) = \mathcal{T}_c^{-1}\left(\mathcal{T}_c\left(Rw_n^o(m)\right)\right) \geq \mathcal{T}_c^{-1}\left(Rw_n^s(m)\right)$. □

Also, we have

Theorem 31. $\tilde{\phi} = \mathcal{T}_c^{-1}\left(\tilde{\psi}\right)$.

Proof. Recall Eq (6.32). We have

$$\begin{aligned} \tilde{\psi} &= \lim_{n\to\infty} \sup_{x\in[0,1]} n \int_x^{x+\frac{1}{n}} \psi^* && \text{(by Theorem 18)} \\ &= \lim_{n\to\infty} \sup_{x\in[0,1-\frac{1}{n}]} n \int_x^{x+\frac{1}{n}} \psi && \text{(by Lemma 15)} \\ &= \lim_{n\to\infty} \sup_{x\in[0,1-\frac{1}{n}]} n \int_x^{x+\frac{1}{n}} (1-c)\phi + c && \text{(by Eq (6.32))} \\ &= (1-c)\left[\lim_{n\to\infty} \sup_{x\in[0,1-\frac{1}{n}]} n \int_x^{x+\frac{1}{n}} \phi\right] + c \\ &= (1-c)\left[\lim_{n\to\infty} \sup_{x\in[0,1]} n \int_x^{x+\frac{1}{n}} \phi^*\right] + c && \text{(by Lemma 15)} \\ &= (1-c)\tilde{\phi} + c = \mathcal{T}_c\left(\tilde{\phi}\right). && \text{(by Theorem 18)} \end{aligned}$$

Therefore, $\tilde{\phi} = \mathcal{T}_c^{-1}\left(\mathcal{T}_c\left(\tilde{\phi}\right)\right) = \mathcal{T}_c^{-1}\left(\tilde{\psi}\right)$. □

All the theorems discussed in Sections 6.2-6.4.2 apply to LA_o. Particularly, Theorem 29 indicates that its long-term gross reward probability satisfies

$$\lim_{n\to\infty}\lim_{m\to\infty} Rw_n^s(m) = \tilde{\psi}. \tag{6.36}$$

Since $\mathcal{T}_c$ is linear and thus continuous, so is $\mathcal{T}_c^{-1}$. Hence [1, Theorem 4.3.2.iii],

$$\lim_{n\to\infty}\lim_{m\to\infty} \mathcal{T}_c^{-1}\left(Rw_n^s(m)\right) = \lim_{Rw_n^s(m)\to\tilde{\psi}} \mathcal{T}_c^{-1}\left(Rw_n^s(m)\right) = \mathcal{T}_c^{-1}\left(\tilde{\psi}\right).$$

Combined with Theorems 30 and 31, the objective reward probability of LA_m satisfies

$$\lim_{n\to\infty}\lim_{m\to\infty} Rw_n^o(m) \geq \lim_{n\to\infty}\lim_{m\to\infty} \mathcal{T}_c^{-1}\left(Rw_n^s(m)\right) = \mathcal{T}_c^{-1}\left(\tilde{\psi}\right) = \tilde{\phi}.$$

Applying Theorem 21, we have

$$\tilde{\phi} \leq \lim_{n\to\infty}\lim_{m\to\infty} Rw_n^o(m) \leq \tilde{\phi}.$$

Thus, Eq (6.33) follows, and the modified algorithm is also ϵ-optimal.

6.6 Examples and Comments

In this section, we first make some comments on ORP. Then we provide some typical examples compatible with (modified) G-LSCALA algorithm.

6.6.1 Comments on ORP

The ORP $\tilde{\phi}$ may not coincide with the global maximum

$$\mathcal{M}_\phi = \sup_{x\in[0,1]} \phi(x).$$

The definition of ORP requires the existence of some interval I such that for almost all $\lambda \in I$, $\phi(\lambda)$ is arbitrarily close $\tilde{\phi}$. In other words, while the global maximum focus on a single-point value, the ORP considers the average function value on an arbitrarily short interval. Figure 6.3 plots a typical case that $\tilde{\phi} \neq \mathcal{M}_\phi$.

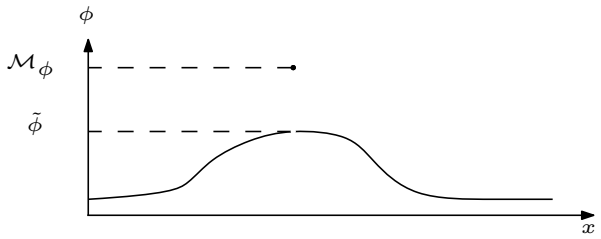


Figure 6.3: A typical case that $\tilde{\phi} \neq \mathcal{M}_\phi$.

The next theorem shows that if ϕ is continuous at the global maximum point, then $\tilde{\phi} = \mathcal{M}_\phi$.

Theorem 32. *Suppose that ϕ has a global maximum point at which ϕ is continuous. Then $\tilde{\phi} = \mathcal{M}_\phi$.*

Proof. It is trivial that $\tilde{\phi} \leq \mathcal{M}_\phi$. So we only need to show $\mathcal{M}_\phi \leq \tilde{\phi}$ as well. Pick $\epsilon > 0$ arbitrary. Let x_0 be the global maximum point. Since ϕ is continuous at x_0, there exists $\delta > 0$ such that $\forall x \in (x_0 - \delta, x_0 + \delta)$, $\mathcal{M}_\phi - \phi(x) < \epsilon$. Equivalently, $\mathcal{M}_\phi - \epsilon < \phi(x)$. Pick $n \in \mathbb{N}^*$ such that $\frac{1}{n} < \delta$, then

$$\mathcal{A}^n(x_0) = n \int_{x_0}^{x_0+\frac{1}{n}} \phi^* = n \int_{x_0}^{x_0+\frac{1}{n}} \phi > n \cdot \left(\frac{1}{n} \cdot (\mathcal{M}_\phi - \epsilon)\right) = \mathcal{M}_\phi - \epsilon.$$

Therefore, $\tilde{\phi} \geq \mathcal{A}^n(x_0) > \mathcal{M}_\phi - \epsilon$. Since ϵ can be arbitrarily small, so $\tilde{\phi} \geq \mathcal{M}_\phi$. □

6.6.2 Examples

We show some typical reward functions, in practice, of bounded variation and thus compatible with (modified) G-LSCALA algorithm.

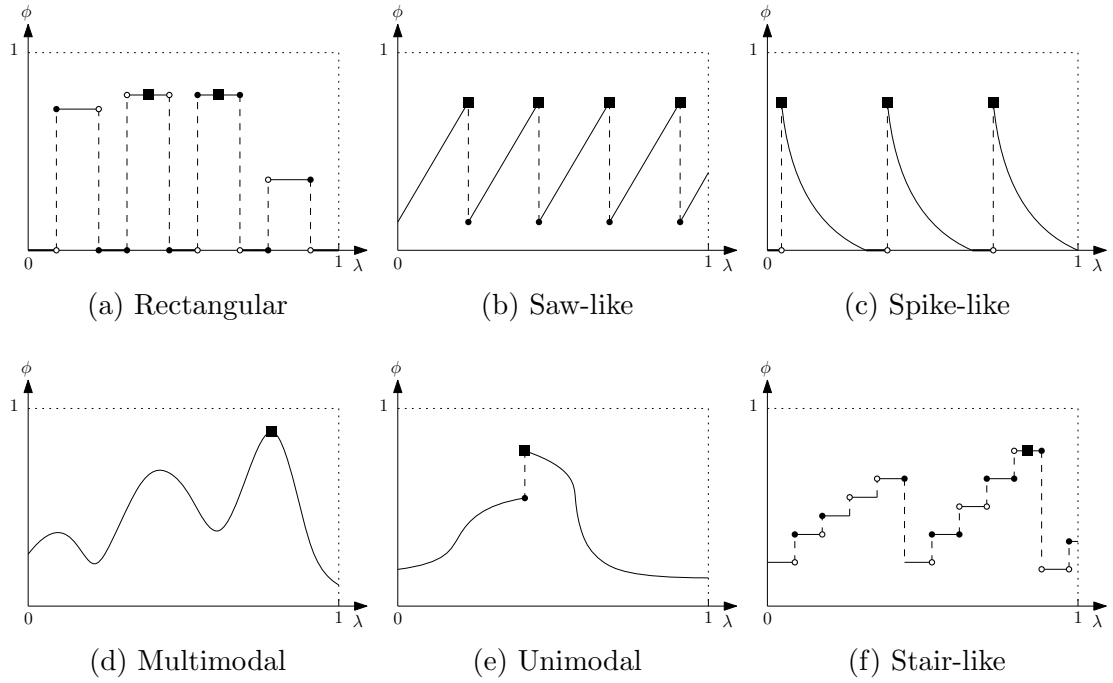
If ϕ is a monotone, then its variation can be easily calculated as follows.

Theorem 33 (Proposition 2.10. [26]). *Let $I \subset \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ a monotone function. Then for every interval $J \subset I$, the variation $V_J(f) = \sup_J f - \inf_J f$.*

For a reward function ϕ, $\sup_{x\in[0,1]} \phi \leq 1$ and $\inf_{x\in[0,1]} \phi \geq 0$. So the variation of a monotone reward function is not greater than one and thus bounded. Therefore, MG-LSCALA algorithm applies to all monotone reward functions. The next theorem shows that piecewise monotone functions are sufficient.

Theorem 34 (Proposition 2.6.iii [26]). *Let $I \subset \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ a function. If $c \in I$, then $V_{I\cup(-\infty,c]}(f) + V_{I\cup[c,\infty)}(f) = V_I(f)$.*

Theorem 34 indicates that the variation is addable over the mutually exclusive intervals. Then, suppose that ϕ can be partitioned into k monotone functions. Its variation is not greater than $k \cdot 1 = k$ and thus bounded. Hence, the modified algorithm can also apply. Figure 6.4 plots some typical piecewise-monotone reward functions whose optimal reward probabilities are marked by solid black squares. The original G-LSCALA algorithm can handle all the functions except the ones in Figures 6.4a and 6.4c as they have minimal function value zero.



(a) Rectangular (b) Saw-like (c) Spike-like

(d) Multimodal (e) Unimodal (f) Stair-like

Figure 6.4: Typical reward functions of bounded variation and thus compatible with our algorithm. The optimal reward probabilities are marked by solid black squares.

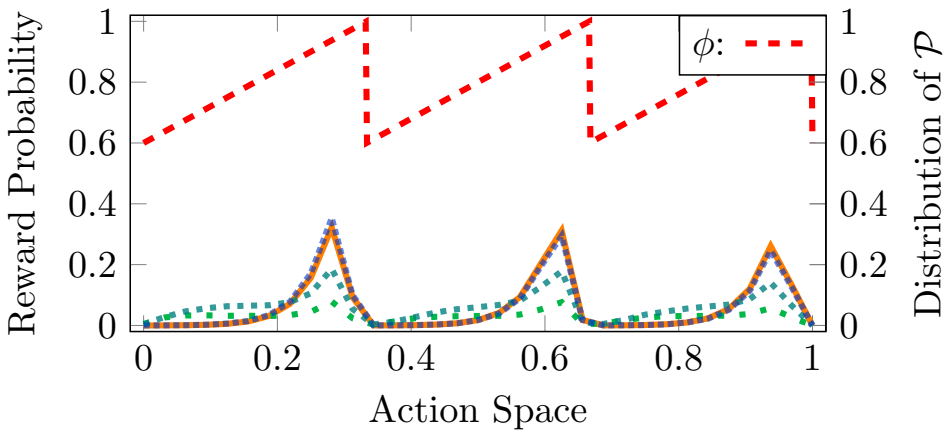
6.7 Experimental Results

Although a considerable number of simulations has been done for different combinations of ϕ, m and n, we only demonstrate a few examples here for the sake of brevity. The experiments are conducted by choosing relatively small m and n so that readers can verify our results within a reasonable amount of time. Better results can always be made by increasing m and n at the cost of the convergence speed of the distribution of **id**.

In the first example, the environment gives responses according to the reward function,

$$\phi(x) = 1.2x - 0.4 \cdot \lfloor 3x \rfloor + 0.6. \tag{6.37}$$

The function is saw-like and plotted in Figure 6.5 by the red dashed line. The analysis



Distr. of $\mathcal{P}$				
# iterations:	1×10^3	2×10^3	2×10^4	∞
Reward Rate:	83.88%	85.77%	91.20%	91.33%

Figure 6.5: The estimated density curve of **id** and the reward rate as a function of the number of iterations. The estimation is made by running 150,000 LA implementing the original G-LSCALA algorithm with $n = 30$, $m = 20$ and the reward function defined by Eq (6.37).

in Section 6.6.2 has shown that the reward function is of bounded variation. Besides, the figure shows ϕ is lower bounded by 0.6 and so applicable with the G-LSCALA algorithm. The function's ORP $\tilde{\phi}$ coincides with its global maximum, which is 1.

It is worth to note that the function fails Assumption 3 as it is not continuous at its global maximum point. So the function cannot be handled by the LSCALA or E-LSCALA algorithms introduced in Sections 3.1 and 4.2.

Figure 6.5 plots the results of the simulation where $n = 30$ and $m = 20$. We run 150,000 LA implementing the original G-LSCALA algorithm in parallel and plot the estimated density curve of **id** and the reward rate as a function of the number of iterations. The results of the infinite number of iterations are theoretical and calculated by Theorem 20 and Eq (6.11). More theoretical results for other combinations of m and n are provided in Table 6.1.

Remark 10. *In Figure 6.5, the density curve of the kernels are scaled for a better*

Table 6.1: The $Rw_n(m)$ of the LA implementing the G-LSCALA algorithm. The reward function ϕ is defined by Eq (6.37).

n \ m	20	32	128	512
32	91.33%	92.72%	94.36%	94.67%
128	94.24%	95.88%	97.98%	98.48%
512	94.99%	96.70%	98.91%	99.47%
1024	95.11%	96.83%	99.07%	99.64%

representation. So readers may not expect that the area below the density curve is one. We apply the same treatment in Figure 6.6.

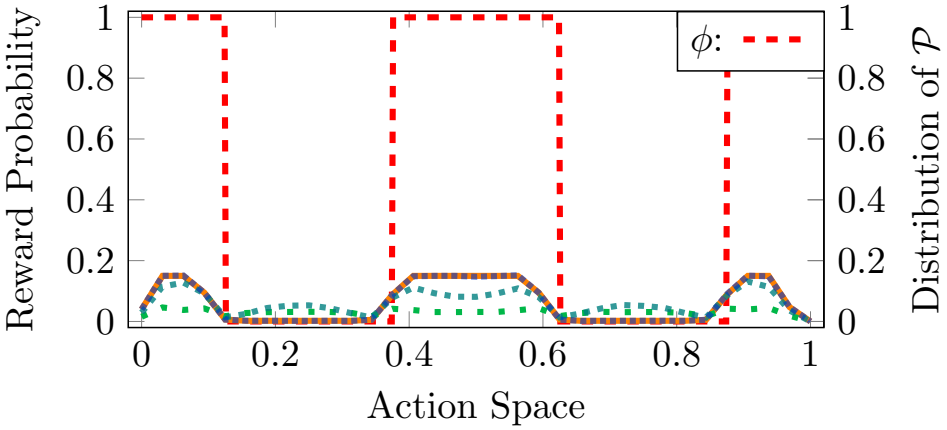
The second experiment is implemented with a square-wave reward function defined by

$$\phi(x) = 2 \cdot \lfloor 2x + 0.25 \rfloor - \lfloor 4x + 0.5 \rfloor + 1. \tag{6.38}$$

We plot the function in Figure 6.6. And it is easy to check that $\tilde{\phi} = 1$. The minimum of the function is zero and so fails **Req.** 1. So the function cannot be dealt with by Algorithm G-LSCALA. Notice that due to the discontinuities near the global maximum points, the function cannot be handled by Algorithms LSCALA or E-LSCALA, either. As a result, the MG-LSCALA algorithm is the only choice. By running $150,000$ LA implementing the algorithm with $n = 32$, $m = 20$ and $c = 0.8$, we get the simulation results presented in Figure 6.6.

We also plot the estimated density curves of **id** and the corresponding reward rate for different number of iterations. The values for infinite number of iterations are calculated by applying the closed-form expression from Theorem 20 and Eq (6.35). It is worth to note that the computation of the stationary distribution of **id** is based on the subjective reward function $\psi = 0.2 \cdot \phi + 0.8$ (see Eq (6.32)). More theoretical results for other choices of m and n are listed in Table 6.2.

From both experiments, we can observe that the density curve of **id** and the reward



Distr. of $\mathcal{P}$				
# iterations:	2×10^2	2×10^3	6×10^6	∞
Reward Rate:	54.69%	70.02%	94.22%	94.28%

Figure 6.6: The estimated density curve of **id** and the reward rate as a function of the number of iterations. The estimation is made by running 150,000 LA implementing the MG-LSCALA algorithm with $n = 32$, $m = 20$ and $c = 0.8$.

Table 6.2: The $Rw_n^o(m)$ of the LA implementing the MG-LSCALA algorithm with $c = 0.5$. The reward function ϕ is defined by Eq (6.38).

n \ m	20	32	128	512
32	94.28%	97.11%	99.23%	99.80%
128	97.65%	99.33%	99.85%	99.96%
512	98.36%	99.76%	99.97%	99.99%
1024	98.47%	99.84%	99.98%	100.00%

rate converge to their theoretical counterparts, which effectively corroborate our analysis about the distribution of **id** and $Rw_n(m)$ for a fixed m and n. Tables 6.1 and 6.2 clearly show that $Rw_n(m)$ approaches to $\tilde{\phi}$ when m and n increase. This trend confirms our statement that both the G-LSCALA and MG-LSCALA algorithms are ϵ-optimal.

6.8 Summary

In this chapter, we have discussed the performance limit of the CALA that takes actions according to some CPD. Given an integrable reward function ϕ, we have defined an associated index, $\tilde{\phi}$, named ORP. We have proved that the index is the least upper bound of the long-term reward probability that an action probability distribution-based CALA algorithm can possibly achieve. We also have discussed the relation between the ORP and the reward function's global maximum points, which have been heavily used in the ϵ-optimality proofs of all classical LA problems.

Our proof is constructive. In particular, we have proposed a CALA algorithm, named G-LSCALA, that converges to $\tilde{\phi}$ in an ϵ-optimal manner. Compared to other LSCALA-series algorithms reported in Chapters 3 and 4, the assumptions over the reward function are further relaxed. We also have shown that one assumption can be dropped by applying a modified algorithm (MG-LSCALA).

We have shown some representative experimental results at the end of this chapter. The results coincide with our expectations based on our theoretical analysis and thus confirm our claims.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

The focus of the thesis is on designing LA whose action space is an interval. We proposed a series of LSCALA algorithms that attempted to resolve numerous issues related to the design of any CALA. All the algorithms have been proved to be ϵ-optimal with the conditions listed in Table 7.1.

This thesis pioneers the concept where the uni-modal functional form for the Environment's infinite reward probabilities do not obey a well-established functional form. To the best of our knowledge, these algorithms are the first four that are proven ϵ-optimal for such arbitrary functions. Since our analysis, in fact, does not impose any constraints upon the shape of the reward function, our algorithm can also work with, but not limited to, multimodal functions. By investigating the problem from a different perspective, we have argued that the solution proposed is closely related to the family of "Stochastic Point Location" problems involving either discretized steps or d-ary parallel machines.

Based on the original LSCALA algorithm introduced in Chapter 3, we also proposed a maximum point search algorithm named CALA-GMS. The algorithm can find the global

Table 7.1: The LSCALA algorithms introduced in the thesis with the assumptions under which the algorithm is ϵ-optimal.

Algorithms:	Assumptions under which the algorithm is ϵ-optimal
LSCALA:	SP, FC
M-LSCALA:	FC
G-LSCALA:	SP, BV
MG-LSCALA:	BV

SP: The reward function is strictly positive.

FC: The set of local maxima can be written as a finite union of intervals, which could possibly be degenerate. Besides, the reward function is continuous near the global maximum points.

BV: The reward function is of bounded variation.

maximum points of a real valued function by taking the sampled function values which may be contaminated by noise. All the claims we made are based the rigorous theoretical analysis and corroborated by experimental results.

Moreover, we investigated the limit of the performance that a CALA algorithm could possible achieve. Our analysis applies for all the CALA algorithms which take actions based on some CPD. In particular, given a reward function, we defined an index associated to it, and we have proved that it is the least upper bound of the long-term reward probability that an action probability distribution-based CALA can achieve. Apart from this, we have proved that the G-LSCALA and MG-LSCALA algorithms meet $\tilde{\phi}$ in an ϵ-optimal manner.

The thesis also highlights the niche contributions within the broader field of learning theory, and to the best of our knowledge, there are no comparable results reported in the literature.

7.2 Future Work

The future work that we anticipate involves the following areas:

- **Absorbing and Ergodic LA**: According to their Markovian representation, automata fall into two categories: Ergodic automata and automata possessing absorbing barriers. The latter automata get locked into a barrier state – sometimes after even a *finite* number of iterations. Many families of automata that posses absorbing barriers have been reported [22, 30, 31, 33, 35, 45]. This current paper has dealt with ergodic solutions to the infinite-action problem domain. The problem of designing absorbing LA to achieve the same remains open.

- **Estimator-based and Pursuit-based LA**: The fastest LA to-date are those which also involve the estimates of the reward probabilities, and these are referred to as Estimator Algorithms (EAs). They work with a noticeably different paradigm, namely one in which the phases of "Exploration" and "Exploitation" are continuously and constantly interleaved. During each learning cycle, these algorithms incorporate an estimation phase, in which they update the estimates of the reward probabilities of the various actions, thus maintaining the so-called "Estimator" vector. These LA, pioneered by Thathachar and Sastry [54], render the learning process to be more goal-directed, and the probability updating of the action probability vector involves an updating function and the "Estimator" vector. This also leads to the families of *Pursuit* Algorithms (PAs) ([2, 37, 63–65]). Since these algorithms consider both the *short-term* responses of the Environment and the *long-term* reward probability estimates in formulating the action probability updating rules, they outperforms traditional VSSA schemes in accuracy. This, in turn, also leads to a much faster convergence – almost an order of magnitude faster

than the continuous and discretized VSSA. While the current paper deals with traditional VSSA to work in infinite-action Environments, the task of designing analogous Estimator-based and Pursuit-based LA achieve the same remains open.

- **Bayesian Estimator-based LA**: More recently, PAs that use the Bayesian estimates (instead of the ML estimates) in the "Estimator" vector have also been designed and analyzed. They probably constitute the fastest LA to-date [65, 66]. The task of designing Bayesian Estimator-based LA to learn in similar Environments with infinite-actions is unresolved.

References

[1] S. Abbott. *Understanding Analysis*. Springer-Verlag, New York, USA, 2015.

[2] M. Agache and B. J. Oommen. Generalized pursuit learning schemes: new families of continuous and discretized learning automata. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 32(6):738–749, 2002.

[3] A. F. Atlassis, N. H. Loukas, and A. V. Vasilakos. The use of learning algorithms in ATM networks call admission control problem: A methodology. *Computer Networks*, 34(3):341–353, 2000.

[4] A. F. Atlassis and A. V. Vasilakos. The use of reinforcement learning algorithms in traffic control of high speed networks. *Advances in Computational Intelligence and Learning*, 18:353–369, 2002.

[5] A. Baddeley and R. Turner. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42, 2005.

[6] M. Barzohar and D. B. Cooper. Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):707–722, 1996.

[7] M. L. Brandeau and S. S. Chiu. An overview of representative problems in location research. *Management Science*, 35(6):645–674, 1989.

[8] P. Bremaud. *Markov chains : Gibbs fields, Monte Carlo simulation, and queues.* Springer, New York, USA, 1999.

[9] H. Hartenstein C. Bettstetter and X. Perez-Costa. Stochastic properties of the random waypoint mobility model. *Wireless Networks*, 10(5):555–567, 2004.

[10] C. K. Chui. Concerning rates of convergence of riemann sums. *Journal of Approximation Theory*, 4(3):279–287, 1971.

[11] J. J. Collins, C. C. Chow, and T. T. Imhoff. Aperiodic stochastic resonance in excitable systems. *Physical Review E*, 52(4):R3321–R3324, 1995.

[12] R. L. Cook. Stochastic sampling in computer graphics. *ACM Transactions on Graphics*, 5(1):51–72, 1986.

[13] J. P. Cusumano and B. W. Kimble. A stochastic interrogation method for experimental measurements of global dynamics and basin evolution: application to a two-well oscillator. *Nonlinear Dynamics*, 8(2):213–235, 1995.

[14] Simon Du, Jason Lee, Yuandong Tian, Aarti Singh, and Barnabas Poczos. Gradient descent learns one-hidden-layer CNN: Don't be afraid of spurious local minima. In *International Conference on Machine Learning (ICML)*, pages 1339–1348, 2018.

[15] G. B. Folland. *Real analysis: modern techniques and their applications.* Wiley, New York, USA, 1999.

[16] Rong Ge, Jason D Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2973–2981, 2016.

[17] Elad Hazan, Kfir Y. Levy, and Shai Shalev-Shwartz. On graduated optimization for stochastic non-convex problems. In *International Conference on Machine Learning (ICML)*, pages 1833–1841, 2016.

[18] M. N. Howell, G. P. Frost, T. J. Gordon, and Q. H. Wu. Continuous action reinforcement learning applied to vehicle suspension control. *Mechatronics*, 7(3):263–276, 1997.

[19] M. N. Howell and T. J. Gordon. Continuous action reinforcement learning automata and their application to adaptive digital filter design. *Engineering Applications of Artificial Intelligence*, 14(5):549–561, 2001.

[20] D. Huang and W. Jiang. A general cpl-ads methodology for fixing dynamic parameters in dual environments. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, 42(5):1489–1500, 2012.

[21] J. Kabudian, M. R. Meybodi, and M. M. Homayounpour. Applying continuous action reinforcement learning automata (CARLA) to global training of hidden markov models. In *Proceedings of ITCC'04*, pages 638–642, 2004.

[22] S. Lakshmivarahan. Learning algorithms theory and applications. chapter 2. Springer-Verlag, New York, 1981.

[23] S. Lakshmivarahan and M. A. L. Thathachar. Absolutely expedient learning algorithms for stochastic automata. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(3):281–286, 1973.

[24] J. K. Lanctot and B. J. Oommen. On discretizing estimator-based learning algorithms. In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, volume 2, pages 1417–1422, 1991.

[25] J. K. Lanctot and B. J. Oommen. Discretized estimator learning automata. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 22(6):1473–1483, 1992.

[26] G. Leoni. *A First Course in Sobolev Spaces*. American Mathematical Society, Providence, USA, 2009.

[27] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, USA, 2008.

[28] M. R. Meybodi and H. Beigy. New learning automata based algorithms for adaptation of backpropagation algorithm parameters. *International Journal of Neural Systems*, 12(1):45–67, 2002.

[29] S. Misra and B. J. Oommen. GPSPA: A new adaptive algorithm for maintaining shortest path routing trees in stochastic networks. *International Journal of Communication Systems*, 17(10):963–984, 2004.

[30] K. Najim and A. S. Poznyak. *Learning Automata: Theory and Applications*. Pergamon Press, Oxford, UK, 1994.

[31] K. S. Narendra and M. A. L. Thathachar. *Learning Automata: An Introduction*. Prentice Hall, Englewood Cliffs, USA, 1989.

[32] J. R. Norris. *Markov chains*. Cambridge University Press, Cambridge, United Kingdom, 1997.

[33] M. S. Obaidat, G. I. Papadimitriou, and A. S. Pomportsis. Learning automata: theory, paradigms, and applications. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 32(6):706–709, 2002.

[34] M. S. Obaidat, G. I. Papadimitriou, A. S. Pomportsis, and H. S. Laskaridis. Learning automata-based bus arbitration for shared-medium ATM switches. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 32(6):815–820, 2002.

[35] B. J. Oommen. Absorbing and ergodic discretized two-action learning automata. *IEEE Transactions on Systems, Man, and Cybernetics*, 16(2):282–296, 1986.

[36] B. J. Oommen. Stochastic searching on the line and its applications to parameter learning in nonlinear optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 27(4):733–739, 1997.

[37] B. J. Oommen and M. Agache. Continuous and discretized pursuit learning schemes: various algorithms and their comparison. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 31(3):277–287, 2001.

[38] B. J. Oommen and T. D. S. Croix. Discretized pursuit learning automata. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(4):195–208, 1990.

[39] B. J. Oommen and T. D. S. Croix. Graph partitioning using learning automata. *IEEE Transactions on Computers*, 45(2):195–208, 1996.

[40] B. J. Oommen and T. D. S. Croix. String taxonomy using learning automata. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 27(2):354–365, 1997.

[41] B. J. Oommen and G. Raghunath. Automata learning and intelligent tertiary searching for stochastic point location. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 28(6):947–954, 1998.

[42] B. J. Oommen, G. Raghunath, and B. Kuipers. Parameter learning from stochastic

teachers and stochastic compulsive liars. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 36(4):820–834, 2006.

[43] M. Paola. Digital simulation of wind field velocity. *Journal of Wind Engineering and Industrial Aerodynamics*, 74-76:91–109, 1998.

[44] G. I. Papadimitriou and A. S. Pomportsis. Learning-automata-based TDMA protocols for broadcast communication systems with bursty traffic. *IEEE Communication Letters*, 4(3):107–109, 2000.

[45] A. S. Poznyak and K. Najim. *Learning Automata and Stochastic Optimization*. Springer-Verlag, London, UK, 1997.

[46] A. Rodríguez, R. Grau, and A. Nowé. Continuous action reinforcement learning automata - performance and convergence. In *Proceedings of ICAART*, pages 473–478, 2011.

[47] B. S. Rowlingson and P. J. Diggle. Splancs: Spatial point pattern analysis code in s-plus. *Computers & Geosciences*, 19(5):627–655, 1993.

[48] G. Santharam, P. Sastry, and M. Thathachar. Continuous action set learning automata for stochastic optimization. *Journal of the Franklin Institute*, 331(5):607–628, 1994.

[49] F. Seredynski. Distributed scheduling using simple learning machines. *European Journal of Operational Research*, 107(2):401–413, 1998.

[50] Umut Simsekli, Cagatay Yildiz, Than Huy Nguyen, Taylan Cemgil, and Gael Richard. Asynchronous stochastic quasi-Newton MCMC for non-convex optimization. In *International Conference on Machine Learning (ICML)*, pages 4674–4683, 2018.

[51] Houshang H. Sohrab. *Basic Real Analysis*. Birkhauser, New York, USA, 2014.

[52] T. Tao, H. Ge, G. Cai, and S. Li. Adaptive step searching for solving stochastic point location problem. In *Proceedings of ICIC*, pages 192–198, 2013.

[53] M. A. L. Thathachar and B. J. Oommen. Discretized reward-inaction learning automata. In *Journal of Cybernetics and Information Science*, pages 24–29, 1979.

[54] M. A. L. Thathachar and P. S. Sastry. Estimator algorithms for learning automata. In *Proceedings of the Platinum Jubilee Conference on Systems and Signal Processing*, 1986.

[55] M. A. L. Thathachar and P. S. Sastry. *Networks of Learning Automata: Techniques for Online Stochastic Optimization*. Springer, New York, USA, 2004.

[56] M. L. Tsetlin. Finite automata and models of simple forms of behaviour. *Russian Mathematical Surveys*, 18(4):1–27, 1963.

[57] C. Unsal, P. Kachroo, and J. S. Bay. Simulation study of multiple intelligent vehicle control using stochastic learning automata. *Transactions of the Society for Computer Simulation International*, 14(4):193–210, 1997.

[58] C. Unsal, P. Kachroo, and J. S. Bay. Multiple stochastic learning automata for vehicle path control in an automated highway system. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 29(1):120–128, 1999.

[59] A. V. Vasilakos, M. P. Saltouros, A. F. Atlassis, and W. Pedrycz. Optimizing QoS routing in hierarchical ATM networks using computational intelligence techniques. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, 33(3):297–312, 2003.

[60] A. Yazidi, O.-C. Granmo, and B. J. Oommen. Service selection in stochastic environments: a learning-automaton based solution. *Applied Intelligence*, 36(3):617–637, 2012.

[61] A. Yazidi, O.-C. Granmo, B. J. Oommen, and M. Goodwin. A novel strategy for solving the stochastic point location problem using a hierarchical searching scheme. *IEEE Transactions on Cybernetics*, 44(11):2202–2220, 2014.

[62] J. Zhang, Y. Wang, C. Wang, and M. Zhou. Symmetrical hierarchical stochastic searching on the line in informative and deceptive environments. *IEEE Transactions on Cybernetics*, 47(3):626–635, 2017.

[63] X. Zhang, O.-C. Granmo, and B. J. Oommen. The bayesian pursuit algorithm: a new family of estimator learning automata. In *Proceedings of IEA/AIE*, pages 522–531, 2011.

[64] X. Zhang, O.-C. Granmo, and B. J. Oommen. Discretized Bayesian pursuit - a new scheme for reinforcement learning. In *Proceedings of IEA/AIE*, pages 784–793, 2012.

[65] X. Zhang, O.-C. Granmo, and B. J. Oommen. On incorporating the paradigms of discretization and Bayesian estimation to create a new family of pursuit learning automata. *Applied Intelligence*, 39(4):782–792, 2013.

[66] X. Zhang, B. J. Oommen, and O.-C. Granmo. The design of absorbing Bayesian pursuit algorithms and the formal analyses of their ϵ-optimality. *Pattern Analysis and Applications*, 20(3):797–808, 2017.