DISSERTATION

DECISION AND LEARNING IN LARGE NETWORKS

Submitted by

Zhenliang Zhang

Department of Electrical and Computer Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2013

Doctoral Committee:

Advisor: Ali Pezeshki
Co-Advisor: Edwin K. P. Chong

Haonan Wang
Rockey J. Luo

ABSTRACT

DECISION AND LEARNING IN LARGE NETWORKS

We consider two topics in this thesis: 1) learning in feedforward and hierarchical networks; and 2) string submodularity in optimal control problems.

In the first topic, we consider a binary hypothesis testing problem and an associated network that attempts jointly to solve the problem. Each agent in the network takes a private signal of the underlying truth, observes the past actions of his neighboring agents, and makes a decision to optimize an objective function (e.g., probability of error). We are interested in the following questions:

- Will the agents asymptotically learn the underlying truth? More specifically, will the overall decision converges (in probability) to the underlying truth as the number of agents goes to infinity?

- If so, how fast is the convergence with respect to the number of agents?

To answer these questions, we investigate two types of networks: Feedforward network and hierarchical tree network, which arise naturally in social and technological networks. Moreover, we investigate the following three parameters: 1. memory size; 2. private signal 'strength;' 3. communication noisiness. We establish conditions on these parameters such that the agents asymptotically learn the underlying truth. Moreover, we study the relationship between the convergence rates and these parameters.

First, we consider the feedforward network, consisting of a large number of nodes, which sequentially make decisions between two given hypotheses. Each node takes a private signal of the underlying truth, observes the decisions from some immediate predecessors, and makes

a decision between the given hypotheses. We consider two classes of broadcast failures: 1) each node broadcasts a decision to the other nodes, subject to random erasure in the form of a binary erasure channel; 2) each node broadcasts a randomly flipped decision to the other nodes in the form of a binary symmetric channel. We are interested in whether there exists a decision strategy consisting of a sequence of likelihood ratio tests such that the node decisions converge in probability to the underlying truth. In both cases, we show that if each node only learns from a bounded number of immediate predecessors, then there does not exist a decision strategy such that the decisions converge in probability to the underlying truth. However, in case 1, we show that if each node learns from an unboundedly growing number of predecessors, then the decisions converge in probability to the underlying truth, even when the erasure probabilities converge to 1. We also derive the convergence rate of the error probability. In case 2, we show that if each node learns from all of its previous predecessors, then the decisions converge in probability to the underlying truth when the flipping probabilities of the binary symmetric channels are bounded away from $1/2$. In the case where the flipping probabilities converge to $1/2$, we derive a necessary condition on the convergence rate of the flipping probabilities such that the decisions still converge to the underlying truth. We also explicitly characterize the relationship between the convergence rate of the error probability and the convergence rate of the flipping probabilities.

Second, we consider the hypothesis testing problem in the context of balanced binary relay trees, where the leaves (and only the leaves) of the tree correspond to N identical and independent sensors. The root of the tree represents a fusion center that makes the overall decision. Each of the other nodes in the tree is a relay node that combines two binary messages to form a single output binary message. In this way, the information from the sensors is aggregated into the fusion center via the relay nodes. We consider the case where the fusion rules at all nonleaf nodes are the Bayesian likelihood ratio tests. In this case, we describe the evolution of the Type I and Type II error probabilities of the binary data as it propagates from the leaves towards the root. Tight upper and lower bounds for the total error probability at the fusion center as functions of N are derived. These bounds characterize the decay rate of the total error probability to 0 with respect to N, even if the individual sensors have error probabilities that converge to $1/2$. We further investigate this problem in the case where nodes and links fail with certain probabilities. Naturally, the asymptotic decay rate of the total error probability is not larger than that in the non-failure case. However, we derive an explicit necessary and sufficient condition on the decay rate of the local failure probabilities (combination of node and link failure probabilities at each level) such that the decay rate of the total error probability in the failure case is the same as that of the non-failure case. We also consider a more general M-ary relay tree configuration, where each non-leaf node in the tree has M child nodes. We derive upper and lower bounds for the Type I and Type II error probabilities associated with this decision with respect to the number of sensors, which in turn characterize the converge rates of the Type I, Type II, and total error probabilities. We also provide a message-passing scheme involving non-binary message alphabets and characterize the exponent of the error probability with respect to the message alphabet size.

In the second topic, we extend the notion of submodularity to optimal control problems. More precisely, we introduce the notion of *string submodularity* in the problem of maximizing an objective function defined on a set of strings subject to a string length constraint. We show that the greedy strategy achieves a $(1 - e^{-1})$-approximation of the optimal strategy. Moreover, we can improve this approximation by introducing additional constraints on curvature, namely, *total backward curvature, total forward curvature*, and *elemental forward curvature*. We show that if the objective function has total backward curvature σ, then the greedy strategy achieves at least a $\frac{1}{\sigma}(1 - e^{-\sigma})$-approximation of the optimal strategy. If the objective function has total forward curvature ϵ, then the greedy strategy achieves at least a $(1 - \epsilon)$-approximation of the optimal strategy. Moreover, we consider a generalization of the diminishing-return property by defining the elemental forward curvature. We also introduce the notion of *string-matroid* and consider the problem of maximizing the objective function subject to a string-matroid constraint. We investigate three applications of string submodular functions with curvature constraints: 1) designing a string of fusion rules in balanced binary relay trees such that the reduction in the error probability is maximized; 2) choosing a string of actions to maximize the expected fraction of accomplished tasks; and 3) designing a string of measurement matrices such that the information gain is maximized.

ACKNOWLEDGEMENTS

My graduate study at Colorado State University is such a wonderful period of my life. I felt very fortunate to have met and worked with so many incredible people. Foremost, I would like to express my sincere gratitude to my advisors Profs. Ali Pezeshki and Edwin K. P. Chong for the continuous supports of my Ph.D study and research, for their patience, motivations, enthusiasms, and immense knowledge. I have learned so many things from them: Not only on how to do good research but also on how to be a rightful person with positive attitude and good work ethic.

I would like to thank the rest of my thesis committee: Profs. Haonan Wang and Rockey J. Luo, for their encouragement, insightful comments, and hard questions.

My sincere thanks go to Prof. William Moran and Dr. Stephen Howard, for several fruitful discussions and insightful comments on my research and papers. My sincere thanks also go to Prof. Louis Scharf, for teaching me the first course in statistical signal processing and giving insightful comments on my research work.

I am grateful to my colleagues and friends, Wenbing Dang, Ramin Zahedi, Yang Zhang, Shankar Ragi, Zheng Wang, Yajing liu, Vidarshana Bandara, Lucas Krakow, Mark Ritschard, Pooria Pakrooh, and Somayeh Hosseini, for all the help and discussions.

I must acknowledge my great friends Zihui Wang, Feng Chen, Yiyan Sun, Lei Lu, Shasha Liu, Xiaoran Wang, and many others who have helped me in so many ways and made my life so enjoyable.

At last, but at most, I would like to thank my wife Yuan Wang. Her support, encouragement, quiet patience, and unwavering love were undeniably the bedrock upon which the past nine years of my life have been built. I am also grateful to my parents Xuedong Zhang and Hongxin Yu, and my sister Han Zhang, for unconditionally supporting and encouraging me throughout my life. I owe them all I am, and all I ever will be.

DEDICATION

To my family.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

As we engage more with online social networks, our opinions get influenced by what we learn from our friends. Meanwhile, there is always the danger of a "*herding*" mentality — when people simply follow a group consensus. Can the network learn the true state of the world and maximize the "wisdom of the crowd?" Moreover, many modern technological systems are networked systems. Networked systems are informationally decentralized, comprise many nodes carrying disparate information, and are subject to constraints on power, communication, and computation. A typical question is how to efficiently aggregate disparate information from a networked system to jointly achieve an overall objective such as detection, target tracking, etc.

We consider a binary hypothesis testing problem and an associated social network that attempts (jointly) to solve the problem. The network consists of a set of agents with interconnections among them. Each of the agents makes a measurement of the underlying true hypothesis, observes the past actions of his neighboring agents, and makes a decision to optimize an objective function (e.g., probability of error). In this thesis, we are interested in the following questions: Will the agents *asymptotically learn* the underlying true hypothesis? More specifically, will the overall network decision converges in probability to the correct decision as the network size (number of agents) increases? If so, how fast is the convergence with respect to the network size? In general, the answers to these questions depend on the social network structure. There are two structures primarily studied in the previous literature.

- Feedforward structure: Each of a set of agents makes a decision in sequence based on its private measurement and the decisions of some or all previous agents. For example, we usually decide on which restaurant to dine in or which movie to go to based on our own taste and how popular they appear to be with previous patrons. Investors often behave similarly in asset markets.

- Hierarchical tree structure: Each of a set of agent makes a decision based on its private measurement and the decisions of its descendent agents in the tree. This structure is common in enterprises, military hierarchies, political structures, online social networks, and even engineering systems (e.g., sensor networks).

The problem of social learning as described above is closely related to the decentralized detection (also known as distributed detection) problem. The latter concerns decision making in a sensor network, where each of the sensors is allowed to transmit a summarized message of its measurement (using a compression function) to an overall decision maker (usually called the fusion center). The goal typically is to characterize the optimal compression functions such that the error probability associated with the detection decision at the fusion center is minimized. However, this problem becomes intractable as the network structure gets complicated. Much of the recent work studies the decentralized detection problems in the asymptotic regime, focusing on the problems of the convergence and convergence rate of the error probability.

1.1 Related Work

The literature on social learning is vast spanning various disciplines including signal processing, game theory, information theory, economics, biology, physics, computer science, and statistics. Here we only review the relevant asymptotic learning results in the two aforementioned network structures.

1.1.1 Feedforward Structure

Consider a large number of nodes, which sequentially make decisions about the underlying truth θ, which equals to one of two given hypotheses H_0 and H_1. At stage k, node a_k takes a measurement X_k (called its *private signal*), receives the decisions of $m_k < k$ immediate predecessors, and makes a binary decision $d_k = 0$ or 1 about the prevailing hypothesis H_0 or

H_1, respectively. It then broadcasts a decision to its successors. Note that m_k is often referred to as the *memory size*.

The research on our problem begins with a seminal paper by Cover [1], which considers the case where each node only observes the decision from its immediate previous node, i.e., $m_k = 1$ for all k. This structure is also known as a *serial network* or *tandem network* and has been studied extensively in [1]–[12]. We use $\mathbb{P}_j$ and π_j to denote the probability measure and the prior probability associated with H_j, $j = 0, 1$, respectively. Cover [1] shows that if the (log)-likelihood ratio for each private signal X_k is bounded almost surely, then using a sequence of likelihood ratio tests the (Bayesian) error probability

$$\mathbb{P}_e^k = \pi_0 \mathbb{P}_0(d_k = 1) + \pi_1 \mathbb{P}_1(d_k = 0)$$

does not converge in probability to 0 as $k \to \infty$. Conversely, if the likelihood ratio is unbounded, then the error probability converges to 0. In the case of unbounded likelihood ratios for the private signals, Veeravalli [8] shows that the error probability converges sub-exponentially with respect to the number k of nodes in the case where the private signals are independent and follow identical Gaussian distribution. Tay *et al.* [9] show that the convergence of error probability is in general sub-exponential and derive a lower bound for the convergence rate of the error probability in the tandem network. Lobel *et al.* [10] derive a lower bound for the convergence rate in the case where each node learns randomly from one previous node (not necessarily its immediate predecessor). In the case of bounded likelihood ratios, Drakopoulos *et al.* [11] provide a non-Bayesian decision strategy, which leads to the convergence of the error probability.

Another extreme scenario is that each node can observe *all* the previous decisions; i.e., $m_k = k - 1$ for all k. This scenario was first studied in the context of social learning [13], [14], where each node uses the Bayesian likelihood ratio test to make its decision. In the case of bounded likelihood ratios for the private signals, the authors of [13] and [14] show that the error probability does not converge to 0, which results in arriving at the wrong decision with positive probability. In the case of unbounded likelihood ratios for the private signals, Smith and Sorensen [15] study this problem using martingales and show that the error probability converges to 0. Krishnamurthy [16], [17] studies this problem from the perspective of quickest time change detection. Acemoglu *et al.* [18] show that the nodes can asymptotically learn the underlying truth in more general network structures.

Most previous work including those reviewed above assume that the nodes and links are perfect. We study the sequential hypothesis testing problem when broadcasts are subject to random erasure or random flipping.

1.1.2 Hierarchical Tree Structure

In many relevant situations, the social network structure is very complicated, wherein each individual makes its decision not by learning from all the past agent decisions, but from only a subset of agents that are directly connected to this individual. For complex network structures, Acemoglu *et al.* [18] provide some sufficient conditions for agents to learn asymptotically from a Bayesian perspective. Jadbabaie *et al.* [19] study the social learning problem from a non-Bayesian perspective. Cattivelli and Sayed [20] study this problem using a diffusion approach. However, analyzing the convergence rate on learning for complex structures remains largely open.

Recent studies suggest that social networks often exhibit hierarchical structures [21]–[31]. These structures naturally arise from the concept of social hierarchy, which has been observed and extensively studied in fish, birds, and mammals [21]. Hierarchical structures can also be observed in networks of human societies [22]; for example, in enterprise organizations, military hierarchies, political structures [25], and even online social networks [29].

In the special case where the tree height is 1, this structure is usually referred as the *star configuration* [32]–[50]. This structure has also been intensively investigated in the context of decentralized detection in sensor networks. The main idea is as follows: Consider a hypothesis testing problem in a network of sensors under two scenarios: *centralized* and *decentralized*.

Under the centralized network scenario, all sensors send their raw measurements to the fusion center, which makes a decision based on these measurements. In the decentralized network scenario, because of the recourse and communication constraint, sensors can only send summaries (e.g., single-bit messages) of their measurements and observations to the fusion center. The fusion center then makes a decision between the given hypotheses. In a decentralized network, information is summarized into smaller messages using quantizing functions. Evidently, the decentralized network cannot perform better than the centralized network. It gains because of its limited use of resources and bandwidth; through transmission of summarized information it is more practical and efficient. A fundamental question is how to quantize the measurements at the sensors and fuse the messages at the fusion center so that the fusion center makes the best decision, in the sense of minimizing an objective function. For example, under the Neyman-Pearson criterion, the objective is to minimize the probability of missed detection with an upper bound constraint on the probability of false alarm. Under the Bayesian criterion, the objective is to minimize the total error probability. A typical result is that under the assumption of (conditionally) independent sensor observations, likelihood ratio quantizers are optimal. Another venue is to study how fast the error probability decays with respect to the number of sensors in a large-scale network. A well-known result is *Stein's Lemma*, which states that under the Newman-Pearson criterion, the decay rate of the error probability in the parallel architecture is exponential.

Tree networks with bounded height (greater than 1) are considered in [50]–[59]. In a tree network, measurements are summarized by leaf agents into smaller messages and sent to their parent agents, each of which fuses all the messages it receives with its own measurement (if any) and then forwards the new message to its parent agent at the next level. This process takes place throughout the tree, culminating at the root (also known as the fusion center) where an overall decision is made. In this way, information from each agent is aggregated at the root via a multihop path. Note that the information is 'degraded' along the path. Therefore, the convergence rate for tree networks cannot be better than that of the star configuration.

More specifically, under the Neyman-Pearson criterion, the optimal error exponent is as good as that of the parallel configuration under certain conditions. For example, for a bounded-height tree network with $\lim_{\tau_N \to \infty} \ell_N/\tau_N = 1$, where τ_N denotes the total number of agents and ℓ_N denotes the number of leaf agents, the optimal error exponent is the same as that of the parallel configuration [52]. under the Bayesian criterion, the error probability converges exponentially fast to 0 with an error exponent that is worse than the one associated with the star configuration [55].

The variation of detection performance with increasing tree height is still largely unexplored. If only the leaf nodes have sensors making observations, and all other nodes simply fuse the messages received and forward the new messages to their parents, the tree network is known as a relay tree. The balanced binary relay tree has been addressed in [60], in which it is assumed that the leaf nodes are independent sensors with identical Type I error probability (also known as the probability of false alarm, denoted by α_0) and identical Type II error probability (also known as the probability of missed detection, denoted by β_0). It is shown there that if the sensor error probabilities satisfy the condition $\alpha_0 + \beta_0 < 1$, then both the Type I and Type II error probabilities at the fusion center converge to 0 as the the number N of leaf nodes goes to infinity. If $\alpha_0 + \beta_0 > 1$, then both the Type I and Type II error probabilities converge to 1, which means that if we flip the decision at the fusion center, then the Type I and Type II error probabilities converge to 0. Because of this symmetry, it suffices to consider the case where $\alpha_0 + \beta_0 < 1$. If $\alpha_0 + \beta_0 = 1$, then the Type I and II error probabilities add up to 1 at each node of the tree. In consequence, this case is not of interest.

Kanoria and Montanari [61] provide an upper bound for the convergence rate of the error probability in M-ary relay trees (directed trees where each nonleaf node has indegree M and outdegree 1), with any combination of fusion rules for all nonleaf agents. Their result gives an upper bound on the rate at which an agent can learn from others in a social network. To elaborate further, the authors of [61] provide the following upper bound for the convergence rate of the error probability P_N at the fusion center (with respect to the number N of leaf nodes) with any combination of fusion rules:

$$\log_2 P_N^{-1} = O(N^{\log_M \frac{M+1}{2}}). \tag{1.1}$$

They also provide the following asymptotic lower bound for the convergence rate in the case of majority dominance rule with random tie-breaking: $\log_2 P_N^{-1} = \Omega(N^{\log_M \lfloor \frac{M+1}{2} \rfloor})$. In the case where M is odd, the majority dominance rule achieves the upper bound in (1.1), which shows that the bound is the optimal convergence rate. However, in the case where M is even, there exists a gap between these two bounds because of the floor function in the second bound. In this case, [61] leaves two questions open:

Q1. Does the majority dominance rule achieve the upper bound in (1.1)?

Q2. Do there exist other strategies that achieve the upper bound in (1.1)?

In this thesis, for the case where M is even, we answer the first question definitively by showing that the majority dominance rule does *not* achieve the upper bound in (1.1). For the second question, we provide a strategy that is closer to achieving the upper bound in (1.1) than the majority dominance rule.

The result in this thesis also differs from (and complements) [61] in a number of other ways. For example, our analysis also includes non-asymptotic results. Moreover, we also consider the Bayesian likelihood ratio test[1] (the fusion rule for *Bayesian learning*) as an alternative fusion rule, not considered in [61]. These differences should become clear as we clarify the contributions of this thesis in the next section.

In the study of social networks, M-ary relay trees arise naturally. First, as pointed out before, many organizational structures are well described in this way. Also, it is well-known

[1]By the Bayesian likelihood ratio test, we mean a likelihood ratio test in which the threshold is given by the ratio of the prior probabilities.

that many real-world social networks, including email networks [62] and the Internet [63], are *scale-free networks*; i.e., the probability $P(\ell)$ that ℓ links are connected to a node is $P(\ell) \sim c\ell^{-\gamma}$, where c is a normalization constant and the parameter $\gamma \in (2, 3)$. In other words, the number of links does not depend on the network size and is bounded with high probability. Moreover, Newman *et al.* [64] show that the average degree in a social network is bounded or grows very slowly as the network size increases. Therefore, to study the learning problem in social networks, it is reasonable to assume that each nonleaf node in the tree has a finite number of child nodes, in which case the tree height grows unboundedly as the number of agents goes to infinity.

1.2 Our Contributions

In this thesis, we study the distributed hypothesis testing problem in the context of feedforward networks and tree networks with unbounded heights. The organization of the thesis is as follows:

- In Chapter 2, we assume that each node uses a likelihood ratio test to generate its binary decision. We call the sequence of likelihood ratio tests a *decision strategy*. We want to know whether or not there exists a decision strategy such that the node decisions converge in probability to the underlying true hypothesis. We consider two classes of broadcast failures: 1) random erasure and 2) random flipping. For case 1, we show that if each node can only learn from a bounded number of immediate predecessors, i.e., there exists a constant C such that $m_k \leq C$ for all k, then for any decision strategy, the error probability cannot converge to 0. We also show that if $m_k \to \infty$ as $k \to \infty$, then there exists a decision strategy such that the error probability converges to 0, even if the erasure probability converges to 1 (given that the convergence of the erasure probability is slower than a certain rate). In the case where an agent learns from all its predecessors, the convergence rate of the error probability is $\Theta(1/\sqrt{k})$. More specifically, we show that if the memory size $m_k = \Theta(k^\sigma)$, $\sigma \leq 1$, then the error probability decreases as $\Theta(1/k^{\min(\sigma, 1/2)})$.

For case 2, we show that if each node can only learn from a bounded number of immediate predecessors, then for any decision strategy, the error probability cannot converge to 0. We also show that if each node can learn from *all* the previous nodes, i.e., $m_k = k-1$, then the error probability converges to 0 using the myopic decision strategy when the flipping probabilities are bounded away from $1/2$. In this case, we show that the error probability converges to 0 as $\Omega(1/k^2)$. In the case where the flipping probability converges to $1/2$, we derive a necessary condition on the convergence rate of the flipping probability (i.e., how fast it must converge) such that the error probability converges to 0. More specifically, we show that if there exists $p > 1$ such that the flipping probability converges to $1/2$ as $O(1/k(\log k)^p)$, then it is impossible that the error probability converges to 0. Therefore, only if the flipping probability converges as $\Omega(1/k(\log k)^p)$ for some $p \leq 1$ can we hope for asymptotic learning. Under this condition, we characterize explicitly the relationship between the convergence rate of the flipping probability and the convergence rate of the error probability.

- In Chapter 3, we study the detection performance for balanced binary relay trees. We derive explicit upper and lower bounds for the total error probability at the fusion center as functions of the number of leaf nodes. These bounds characterize the asymptotic convergence rate for the total error probability as the number of leaf nodes goes to infinity. We also show that the total error probability converges to 0 even if the leaf nodes are asymptotically crummy; i.e., the sum of Types I and II error probabilities goes to 1 as the number of leaf nodes goes to infinity, provided that the rate of crumminess is not sufficiently fast.

- In Chapter 4, we investigate the detection performance in balanced binary relay trees where nodes and links fail with certain probabilities. We show that the asymptotic decay rate of the total error probability is not larger than that in the non-failure case. We show, however, that if the given failure probabilities decrease to 0 sufficiently quickly as the nodes get closer to the fusion center, then the scaling law of the decay rate for the total error probability at the fusion center remains the same as that of the non-failure case. Conversely, if the given failure probabilities do not decrease to 0 sufficiently quickly, then the scaling law of the decay rate is strictly smaller than that of the non-failure case.

- In Chapter 5, we consider a more general M-ary relay tree configuration, where each non-leaf node in the tree has M child nodes. We consider two fusion rules: Majority dominance and the Bayesian likelihood ratio test. We derive upper and lower bounds for the Type I and Type II error probabilities with respect to the number of leaf agents, which in turn characterize the converge rates of the Type I, Type II, and total error probabilities. We also provide a message-passing scheme involving non-binary message alphabets and characterize the exponent of the error probability with respect to the message alphabet size.

- In Chapter 6, we introduce the notion of *string submodularity* in the problem of maximizing an objective function defined on a set of strings subject to a string length constraint. We show that the greedy strategy achieves at least a $(1 - e^{-1})$-approximation of the optimal strategy. Moreover, we can improve this approximation by introducing additional constraints on curvature, namely, *total backward curvature, total forward curvature*, and *elemental forward curvature*. We show that if the objective function has total backward curvature σ, then the greedy strategy achieves at least a $\frac{1}{\sigma}(1 - e^{-\sigma})$-approximation of the optimal strategy. If the objective function has total forward curvature ϵ, then the greedy strategy achieves at least a $(1 - \epsilon)$-approximation of the optimal strategy. Moreover, we consider a generalization of the diminishing-return property by defining the elemental forward curvature. We also introduce the notion of *string-matroid* and consider the problem of maximizing the objective function subject to a string-matroid constraint. We investigate three applications of string submodular functions with curvature constraints: 1) designing a string of fusion rules in balanced binary relay trees such that the reduction in the error probability is maximized; 2) choosing a string of actions to maximize the expected fraction of accomplished tasks; and 3) designing a string of measurement matrices such that the information gain is maximized.

- In Chapter 7, we conclude this thesis and discuss some further research directions.

1.3 Notation

In this thesis, we use the following notation to characterize the asymptotic relationship: For positive functions f and g defined on the positive integers, if there exist positive constant c_1 such that $f(N) \geq c_1 g(N)$ for all sufficiently large N, then we write $f(N) = \Omega(g(N))$. If there exist positive constant c_2 such that $f(N) \leq c_2 g(N)$ for all sufficiently large N, then we write $f(N) = O(g(N))$. We write $f(N) = \Theta(g(N))$ if and only if $f(N) = \Omega(g(N))$ and $f(N) = O(g(N))$. For $N \to \infty$, the notation $f(N) \sim g(N)$ means that $f(N)/g(N) \to 1$, $f(N) = \omega(g(N))$ that $f(N)/g(N) \to \infty$, and $f(N) = o(g(N))$ that $f(N)/g(N) \to 0$.

CHAPTER 2

FEEDFORWARD NETWORKS

In this chapter, we study the binary hypothesis testing problem in the context of feedforward networks. Consider a large number of nodes, which make decisions and broadcast their decisions to others sequentially. We will investigate the cases where the broadcast messages are subject to random erasure or random flipping:

1) *Random erasure*: Each broadcasted decision is erased with a certain erasure probability, modeled by a binary erasure channel. If the decision broadcasted by a node is erased, then none of its successors will observe that decision. We investigate this case in Section 2.2.

2) *Random flipping*: Each broadcasted decision is flipped with a certain flipping probability, modeled by a binary symmetric channel. If the broadcasted decision of a node is flipped, then all the successors of that node observe that flipped decision. We investigate this case in Section 2.3.

2.1 Preliminary

We use $\mathbb{P}$ to denote the underlying probability measure. We use π_j to denote the prior probability (assumed nonzero), $\mathbb{P}_j$ to denote the probability measure, and $\mathbb{E}_j$ to denote the conditional expectation associated with H_j, $j = 0, 1$. Consider a large number of nodes which make decisions sequentially. As shown in Fig. 2.1, at stage k, node a_k takes a measurement X_k of the scene and makes a decision $d_k = 0$ or $d_k = 1$ about the prevailing hypothesis H_0 or H_1. It then broadcasts a potentially corrupted form $\hat{d}_k$ of that decision to its successors. Note that in case 1, if the decision is erased, it is equivalent to saying that the corrupted decision $\hat{d}_k$ is e, which is a message that carries no information and is not useful for decision-making. Inserting e in place of erased messages allows us to unify the notation for cases 1 and 2. The

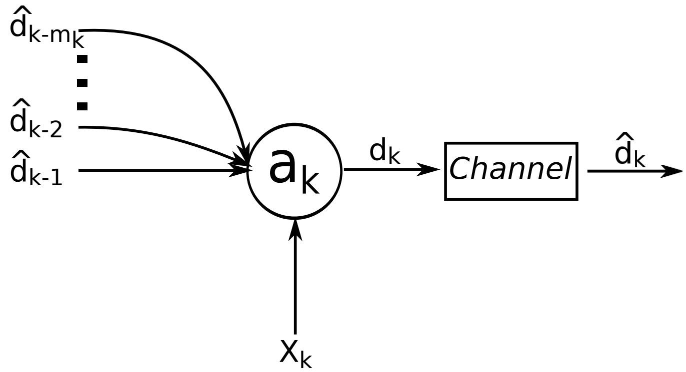


Figure 2.1: Feedforward network.

decision d_k of node a_k is made based on the private signal X_k and the sequence of corrupted decisions $\hat{D}_{m_k} = \{\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_{m_k}\}$ received from the m_k immediate predecessor nodes using a likelihood ratio test.

Our aim is to find a sequence of likelihood ratio tests such that the probability of making a wrong decision about the state of the world tends to 0 as $k \to \infty$; i.e.,

$$\lim_{k\to\infty} \mathbb{P}_e^k = \lim_{k\to\infty} (\pi_0 \mathbb{P}_0(d_k = 1) + \pi_1 \mathbb{P}_1(d_k = 0)) = 0.$$

Before proceeding, we introduce the following definitions and assumptions:

1. The private signal X_k takes values in a set S, endowed with a σ-algebra $\mathcal{S}$. We assume that X_k is independent of the broadcast history $\hat{D}_{m_k}$. Moreover, the X_ks are mutually independent and identically distributed with distribution $\mathbb{P}_j^X$, under H_j, $j = 0, 1$. (Note that $\mathbb{P}_j^X$ is a probability measure on the σ-algebra $\mathcal{S}$.) We assume that the underlying hypothesis, H_0 or H_1, does not change with k.

2. The two probability measures $\mathbb{P}_0^X$ and $\mathbb{P}_1^X$ are equivalent; i.e., they are absolutely continuous with respect to each other. In other words, if $A \in \mathcal{S}$, then $\mathbb{P}_0^X(A) = 0$ if and only if $\mathbb{P}_1^X(A) = 0$.

3. Let the likelihood ratio of a private signal $s \in S$ be

$$L_X(s) = \frac{d\mathbb{P}_1^X}{d\mathbb{P}_0^X}(s),$$

where $d\mathbb{P}_1^X/d\mathbb{P}_0^X$ denotes the Radon–Nikodym derivative (which is guaranteed to exist because of the assumption that the two measures are equivalent). We assume that the likelihood ratios for the private signals are unbounded; i.e., for any set $S' \subset S$ with probability 1 under the measure $(\mathbb{P}_0^X + \mathbb{P}_1^X)/2$, we have

$$\inf_{s \in S'} \frac{d\mathbb{P}_1^X}{d\mathbb{P}_0^X}(s) = 0$$

and

$$\sup_{s \in S'} \frac{d\mathbb{P}_1^X}{d\mathbb{P}_0^X}(s) = \infty.$$

We note that this assumption is saying that the private signals are very 'strong.' Moreover, this assumption can be significantly relaxed in the tree networks, which we will investigate in Chapters 3,4,5.

4. Suppose that θ is the underlying truth. Let $\bar{b}_k = \mathbb{P}(\theta = H_1 | X_k)$, which we call the *private belief* of a_k. By Bayes' rule, we have

$$\bar{b}_k = \left(1 + \frac{\pi_0}{\pi_1} \frac{1}{L_X(X_k)}\right)^{-1}. \tag{2.1}$$

5. Recall that node a_k observes m_k decisions $\hat{D}_{m_k}$ from its immediate predecessors. Let p_j^k be the conditional probability mass function of $\hat{D}_{m_k}$ under H_j, $j = 0, 1$. The likelihood ratio of a realization $\mathcal{D}_{m_k}$ is

$$L_D^k(\mathcal{D}_{m_k}) = \frac{p_1^k(\mathcal{D}_{m_k})}{p_0^k(\mathcal{D}_{m_k})} = \frac{\mathbb{P}_1(\hat{D}_{m_k} = \mathcal{D}_{m_k})}{\mathbb{P}_0(\hat{D}_{m_k} = \mathcal{D}_{m_k})}.$$

6. Let $b_k = \mathbb{P}(\theta = H_1 | \hat{D}_{m_k})$, which we call the *public belief* of a_k. We have

$$b_k = \left(1 + \frac{\pi_0}{\pi_1} \frac{1}{L_D^k(\hat{D}_{m_k})}\right)^{-1}. \tag{2.2}$$

7. Each node a_k makes its decision using its own measurement and the observed decisions based on a likelihood ratio test with a threshold $t_k > 0$:

$$d_k = \begin{cases} 1 & \text{if } L_X(X_k) L_D^k(\hat{D}_{m_k}) > t_k, \\ 0 & \text{if } L_X(X_k) L_D^k(\hat{D}_{m_k}) \leq t_k. \end{cases}$$

If $t_k = \pi_0/\pi_1$, then this test becomes the maximum a-posteriori probability (MAP) test, in which case the probability of error is locally minimized for node a_k. If $t_k = 1$, then the test becomes the maximum-likelihood (ML) test. If the prior probabilities are equal, then these two tests are identical. A decision strategy $\mathbb{T}$ is a sequence of likelihood ratio tests with thresholds $\{t_k\}_{k=1}^{\infty}$. Given a decision strategy, the decision sequence $\{d_k\}_{k=1}^{\infty}$ is a well-defined stochastic process.

8. We say that the system *asymptotically learns* the underlying true hypothesis with decision strategy $\mathbb{T}$ if

$$\lim_{k\to\infty} \mathbb{P}(d_k = \theta) = 1.$$

In other words, the probability of making a wrong decision goes to 0, i.e.,

$$\lim_{k\to\infty} \mathbb{P}_e^k = 0.$$

The question we are interested in is this: In each of the two classes of failures, is there a decision strategy such that the system asymptotically learns the underlying true hypothesis?

2.2 Random Erasure

In this section, we consider the sequential hypothesis testing problem in the presence of random erasures, modeled by binary erasure channels [65] (see Fig. 2.2). Recall that the binary message d_k is the input to a binary erasure channel and $\hat{d}_k$ is the output, which is either equal to d_k (no erasure) or is equal to a symbol e that represents the occurrence of an erasure. The erasure channel matrix at stage k is given by $\mathbb{P}(\hat{d}_k = i | d_k = j)$, $j = 0, 1$ and $i = j, e$. Recall that each node a_k observes m_k immediate previous broadcasted decisions. We divide our analysis into two scenarios: *A*) $\{m_k\}$ is bounded above by a positive constant; *B*) m_k goes to infinity as $k \to \infty$.

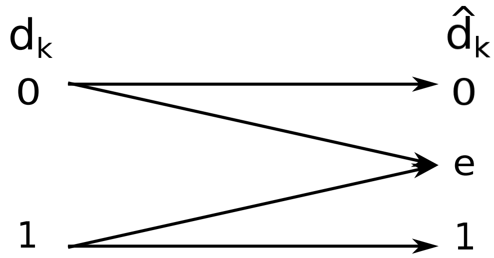


Figure 2.2: Binary erasure channel.

2.2.1 Bounded Memory

Theorem 2.2.1. *Suppose that there exists C and $\epsilon > 0$ such that for all k, $m_k \leq C$ and $\mathbb{P}(\hat{d}_k = e | d_k = j) \in [\epsilon, 1-\epsilon]$ for $j = 0, 1$. Then, there does not exist a decision strategy such that the error probability converges to 0.*

Proof. We first prove this claim for the special case of the tandem network, where $m_k = 1$ for all k. For each node a_k, with a nonzero probability $\mathbb{P}(\hat{d}_k = e | d_k = j)$, the decision $d_{k-1} = j$ of the immediate predecessor is erased and a_k makes a decision based only on its own private signal X_k. We use $\mathcal{E}_k$ to denote this event. Conditioned on $\mathcal{E}_k$, we claim that the error probability as a sequence of k,

$$
\begin{aligned}
&\mathbb{P}(d_k \neq \theta | \mathcal{E}_k) \\
&= \pi_0 \mathbb{P}_0(d_k = 1 | \mathcal{E}_k) + \pi_1 \mathbb{P}_1(d_k = 0 | \mathcal{E}_k) \\
&= \pi_0 \mathbb{P}_0(L_X(X_k) > t_k) + \pi_1 \mathbb{P}_1(L_X(X_k) \leq t_k),
\end{aligned}
$$

is bounded away from 0. We prove the above claim by contradiction. Suppose that there exists a decision strategy with threshold sequence $\{t_k\}$ such that $\mathbb{P}(d_k \neq \theta | \mathcal{E}_k) \to 0$ as $k \to \infty$. Then, we must have $\mathbb{P}_1(L_X(X_k) \leq t_k) \to 0$ because π_1 is positive. Because $\mathbb{P}_0^X$ and $\mathbb{P}_1^X$ are equivalent measures, we have $\mathbb{P}_0(L_X(X_k) \leq t_k) \to 0$. Hence we have $\mathbb{P}_0(L_X(X_k) > t_k) \to 1$. Therefore, $\mathbb{P}(d_k \neq \theta | \mathcal{E}_k)$ does not converge to 0.

We use $\mathcal{E}_k^{\mathcal{C}}$ to denote the complement event of $\mathcal{E}_k$. By the Law of Total Probability, we have

$$\begin{aligned}\mathbb{P}_e^k &= \mathbb{P}(\mathcal{E}_k)\mathbb{P}(d_k \neq \theta|\mathcal{E}_k) + \mathbb{P}(\mathcal{E}_k^{\mathcal{C}})\mathbb{P}(d_k \neq \theta|\mathcal{E}_k^{\mathcal{C}}) \\ &\geq \mathbb{P}(\mathcal{E}_k)\mathbb{P}(d_k \neq \theta|\mathcal{E}_k).\end{aligned}$$

Because $\mathbb{P}(\mathcal{E}_k) \geq \epsilon$, we conclude that the error probability does not converge to 0.

We can now generalize this proof to the case of a general bounded m_k sequence. Let $\mathcal{E}_k$ be the event that a_k receives m_k erased symbols e. Then, the probability $\mathbb{P}(\mathcal{E}_k)$ is bounded below according to

$$\mathbb{P}(\mathcal{E}_k) \geq \left(\min_{\substack{j=0,1 \\ m=k-1,\ldots,k-m_k}} \mathbb{P}(\hat{d}_m = e|d_m = j) \right)^{m_k} \geq \epsilon^{m_k}.$$

We have already shown that given this event the error probability does not converge to 0. Using the Law of Total Probability, It is easy to see that the error probability does not converge to 0. $\square$

Remark 2.2.1. We use $\mathbb{P}(\hat{d}_k = e|d_k = j) \in [\epsilon, 1-\epsilon]$ for $j = 0, 1$ to mean that the erasure probability $\mathbb{P}(\hat{d}_k = e|d_k = j)$ is bounded away from 0 and 1.

This result is straightforward to understand. If the memory sizes are bounded for all nodes, then for each node, there exists a positive probability such that all the decisions received from its immediate predecessors are erased, in which case the node has to make a decision based on its own measurement. The error probability cannot converge to 0 because of the equivalent-measure assumption.

2.2.2 Unbounded Memory

Suppose that each node a_k observes m_k immediate previous decisions. In this section, we deal with the case where m_k is unbounded.[2] More specifically, we consider the case where m_k

[2] The assumption that m_k is unbounded is not sufficiently strong to guarantee the convergence of error probability to 0. An example is that the memory size m_k equals $\sqrt{k}$ if $\sqrt{k}$ is an integer and it equals 1 otherwise. In this case, we can use a similar argument as that in the proof of Theorem 2.2.1 to show that the error probability does not converge to 0.

goes to infinity. We first consider the case where the erasure probabilities are bounded away from 1. We have the following result.

Theorem 2.2.2. *Suppose that m_k goes to infinity as $k \to \infty$ and there exists $\epsilon > 0$ such that for all $j = 0, 1$ and for all k, $\mathbb{P}(\hat{d}_k = e | d_k = j) \leq 1 - \epsilon$. Then, there exists a decision strategy such that the error probability converges to 0.*

Proof. We prove this result by constructing a certain tandem network within the original network using a *backward-searching scheme*. The scheme is the following: Consider node a_k in the original network. Let n_k be the largest integer such that each node in the sequence $\{a_{k-n_k^2}, a_{k-n_k^2-1}, \ldots, a_k\}$ of $n_k^2 + 1$ nodes has a memory size that is greater than or equal to n_k. Note that an n_k satisfying this condition is guaranteed to exist. Moreover, because m_k goes to infinity as $k \to \infty$, we have $n_k \to \infty$ as $k \to \infty$. Consider the event that a_k receives at least one decision j, which is not erased, from $\{a_{k-n_k}, \ldots, a_{k-1}\}$, its n_k immediate predecessors. The probability of this event is at least

$$1 - \max_{\substack{j=0,1 \\ m=k-n_k,\ldots,k-1}} \mathbb{P}(\hat{d}_m = e | d_m = j)^{n_k},$$

which is bounded below by $1 - (1 - \epsilon)^{n_k}$ by the assumption on the erasure probabilities. We denote the node that sends the unerased decision by a_{k_1}. Similarly, with a certain probability, a_{k_1} receives at least one decision, which is not erased, from its n_k immediate predecessors. Recursively, with a certain probability, we can construct a tandem network with length n_k using nodes from among the $n_k^2 + 1$ nodes above within the original network. Let $\mathcal{E}_k$ be the event that such a tandem network exists. The probability $\mathbb{P}(\mathcal{E}_k)$ is at least $(1 - (1 - \epsilon)^{n_k})^{n_k}$. Recall that $\lim_{k\to\infty} n_k = \infty$, which implies that

$$\lim_{k\to\infty} (1 - (1 - \epsilon)^{n_k})^{n_k} = 1.$$

Hence we have

$$\lim_{k\to\infty} \mathbb{P}(\mathcal{E}_k) = 1.$$

Conditioned on $\mathcal{E}_k$, by using the strategy $\mathbb{T}$ consisting of a sequence of likelihood ratio tests with monotone thresholds described in [1], we can get the conditional convergence of the error probability, given $\mathcal{E}_k$, to 0. We can also use the equilibrium strategy described in [10]. Therefore, by the Law of Total Probability, we have

$$\begin{aligned}
&\lim_{k\to\infty} \mathbb{P}(d_k \neq \theta) \\
&= \lim_{k\to\infty} \left(\mathbb{P}(d_k \neq \theta|\mathcal{E}_k)\mathbb{P}(\mathcal{E}_k) + \mathbb{P}(d_k \neq \theta|\mathcal{E}_k^{\mathcal{C}})(1-\mathbb{P}(\mathcal{E}_k)\right) \\
&\leq \lim_{k\to\infty} \left(\mathbb{P}(d_k \neq \theta|\mathcal{E}_k) + (1-\mathbb{P}(\mathcal{E}_k)\right) = 0. \qquad (2.3)
\end{aligned}$$

□

Note that given a strategy, the convergence rate for the error probability in this case depends on how fast $\mathbb{P}(\mathcal{E}_k)$ converges to 1 and how fast $\mathbb{P}(d_k \neq \theta|\mathcal{E}_k)$ converges to 0.

First let us consider the convergence rate of $\mathbb{P}(\mathcal{E}_k)$. Obviously this convergence rate depends on the convergence rate of n_k. Moreover, the convergence rate of n_k depends on the convergence rate of m_k. For example, if m_k goes to infinity extremely slowly, then n_k grows extremely slowly with respect to k, which means that $\mathbb{P}(\mathcal{E}_k)$ converges to 1 extremely slowly with respect to k. Next we assume that m_k increases as $\Theta(k^\sigma)$, where $\sigma \leq 1$. We first establish a relationship between the convergence rate of m_k and the convergence rate of n_k when using the backward-searching scheme.

Proposition 2.2.1. *Suppose that $m_k = \Theta(k^\sigma)$ where $\sigma \leq 1$. Then, we have*

$$n_k = \begin{cases} \Theta(\sqrt{k}) & \text{if } \sigma \geq 1/2, \\ \Theta(k^\sigma) & \text{if } \sigma < 1/2. \end{cases}$$

Proof. Suppose that we can form a tandem network with length n_k within the original network. Recall that n_k is the largest integer such that each node in the sequence $\{a_{k-n_k^2}, a_{k-n_k^2-1}, \ldots, a_k\}$ of n_k^2+1 nodes has a memory size that is greater than or equal to n_k. Therefore, the memory size $m_{k-n_k^2}$ of $a_{k-n_k^2}$ must be larger than or equal to n_k by assumption. Hence we have

$$m_{k-n_k^2} = (k-n_k^2)^\sigma \geq n_k.$$

Moreover, the memory size $m_{k-(n_k+1)^2}$ of $a_{k-(n_k+1)^2}$ must be strictly smaller than $n_k + 1$ (otherwise we can construct a tandem network with length $n_k + 1$). Hence we have

$$m_{k-(n_k+1)^2} = (k - (n_k + 1)^2)^\sigma < n_k + 1.$$

From the above two inequalities, we easily obtain the desired asymptotic rates for n_k. $\square$

Remark 2.2.2. Note that if $\sigma < 1/2$, then the scaling law of n_k is identical to that of m_k: The faster the scaling of m_k, the faster the scaling of n_k also. However, for $\sigma \geq 1/2$, the scaling law of n_k "saturates" at $\sqrt{k}$, no matter how fast m_k scales.

We have derived the convergence rate for n_k. Recall that $\mathbb{P}(\mathcal{E}_k)$ converges to 1 at least in the rate of $\Theta(n_k(1-\epsilon)^{n_k})$ (by expanding the term $(1-(1-\epsilon)^{n_k})^{n_k}$ and keeping the dominating term). From this fact and Proposition 2.2.1, we derive the convergence rate for $\mathbb{P}(\mathcal{E}_k)$.

Corollary 2.2.1. *Suppose that $m_k = \Theta(k^\sigma)$ where $\sigma \leq 1$. Then, we have*

$$1 - \mathbb{P}(\mathcal{E}_k) = \begin{cases} O(\sqrt{k}(1-\epsilon)^{\sqrt{k}}) & \text{if } \sigma \geq 1/2, \\ O(k^\sigma(1-\epsilon)^{k^\sigma}) & \text{if } \sigma < 1/2. \end{cases}$$

Second, let us consider the convergence rate of $\mathbb{P}(d_k \neq \theta|\mathcal{E}_k)$. Recall that $\mathcal{E}_k$ denotes the event that a tandem network with length n_k exists. Conditioned on $\mathcal{E}_k$, if we use the the equilibrium strategy[3] described in [10], then it has been shown that the error probability converges to 0 as $\Theta(1/n_k)$, with appropriate assumptions on the distributions of the private signal. From this fact and Proposition 2.2.1, we derive the convergence rate for $\mathbb{P}(d_k \neq \theta|\mathcal{E}_k)$.

Corollary 2.2.2. *Suppose that $m_k = \Theta(k^\sigma)$ where $\sigma \leq 1$. Then, we have*

$$\mathbb{P}(d_k \neq \theta|\mathcal{E}_k) = \begin{cases} \Theta(1/\sqrt{k}) & \text{if } \sigma \geq 1/2, \\ \Theta(1/k^\sigma) & \text{if } \sigma < 1/2. \end{cases}$$

[3]Note that this equilibrium strategy is *not* the only strategy such that the error probability converges to 0 in a tandem network.

Notice that the convergence rate of $\mathbb{P}(d_k \neq \theta | \mathcal{E}_k)$ is much smaller than that of $\mathbb{P}(\mathcal{E}_k)$. Moreover by (2.3), the convergence rate of $\mathbb{P}(d_k \neq \theta)$ depends on the smaller of the convergence rates of $\mathbb{P}(d_k \neq \theta | \mathcal{E}_k)$ and $\mathbb{P}(\mathcal{E}_k)$. We derive the convergence rate for the error probability as follows.

Corollary 2.2.3. *Suppose that* $m_k = \Theta(k^\sigma)$ *where* $\sigma \leq 1$. *Then, we have*

$$\mathbb{P}(d_k \neq \theta) = \begin{cases} \Theta(1/\sqrt{k}) & \text{if } \sigma \geq 1/2, \\ \Theta(1/k^\sigma) & \text{if } \sigma < 1/2. \end{cases}$$

We have considered the situation where the erasure probabilities are bounded away from 1. Now consider the case where the erasure probability $\mathbb{P}(\hat{d}_k = e | d_k = j)$ converges to 1.

Theorem 2.2.3. *Suppose that* $\mathbb{P}(\hat{d}_k = e | d_k = j) \to 1$ *and there exists* $\epsilon > 1$ *and* $c > 0$ *such that* $\mathbb{P}(\hat{d}_k = e | d_k = j) \leq (cn_k)^{-\epsilon/n_k}$. *Then, there exists a decision strategy such that the error probability converges to 0.*

Proof. We use the scheme described in the proof of Theorem 2.2.2. The probability that a tandem network with length n_k exists is at least $(1 - ((cn_k)^{-\epsilon/n_k})^{n_k})^{n_k} = (1 - (cn_k)^{-\epsilon})^{n_k}$, which converges to 1 as $k \to \infty$. Using the same arguments as those in the proof of Theorem 2.2.2, we can show that the error probability converges to 0. □

As an example, we consider the situation where each node observes *all* the previous decisions; i.e, $m_k = k - 1$ for all k. In this case, it is easy to show that using the backward-searching scheme, with a certain probability, we can form a tandem network with length $n_k = \lfloor \sqrt{k-1} \rfloor$. Suppose that the erasure probabilities are bounded away from 1. Then, the error probability converges to 0 as $\Theta(1/\sqrt{k})$. Moreover, the error probability converges to 0 even if the erasure probability converges to 1, provided that $\mathbb{P}(\hat{d}_k = e | d_k = j) \leq (cn_k)^{-\epsilon/n_k}$.

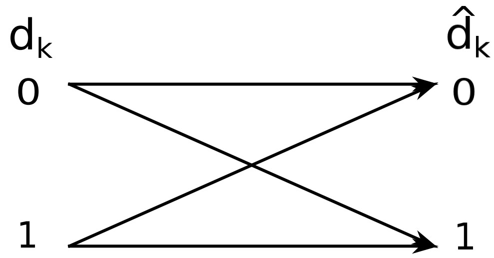


Figure 2.3: Binary symmetric channel.

2.3 Random Flipping

We study in this section the sequential hypothesis testing problem with random flipping, modeled by a binary symmetric channel [65] (see Fig. 2.3). Recall that d_k is the input to a binary symmetric channel and $\hat{d}_k$ is the output, which is either equal to d_k (no flipping) or is equal to its complement $1 - d_k$ (flipping). The channel matrix is given by $\mathbb{P}(\hat{d}_k = i | d_k = j)$, $i, j = 0, 1$. We assume that $\mathbb{P}(\hat{d}_k = 1 | d_k = 0) = \mathbb{P}(\hat{d}_k = 0 | d_k = 1) = q_k$, where q_k denotes the probability of a flip. The assumption of symmetry is for simplicity only, and all results obtained in this section can be generalized easily to a general binary communication channel with unequal flipping probabilities, i.e., $\mathbb{P}(\hat{d}_k = 1 | d_k = 0) \neq \mathbb{P}(\hat{d}_k = 0 | d_k = 1)$. We assume that each node a_k knows the probabilities of flipping associated with the corrupted decisions $\hat{D}_{m_k}$ received from its predecessors.

2.3.1 Bounded Memory

Theorem 2.3.1. *Suppose that there exists C and $\epsilon > 0$ such that for all k, $m_k \leq C$ and $q_k \in [\epsilon, 1 - \epsilon]$. Then, there does not exist a decision strategy such that the error probability converges to 0.*

Proof. We first prove this theorem in the case where each node observes the immediate previous node; i.e., $m_k = 1$ for all k. Node a_k makes a decision d_k based on its private signal X_k and the decision $\hat{d}_{k-1}$ from its immediate predecessor. Recall that $q_k = \mathbb{P}(\hat{d}_k = 1 | d_k = 0) =$

$\mathbb{P}(\hat{d}_k = 0|d_k = 1)$. The likelihood ratio test at stage k (with a threshold $t_k > 0$) is

$$d_k = \begin{cases} 1 & \text{if } L_X(X_k)L_D^k(\hat{d}_{k-1}) > t_k, \\ 0 & \text{if } L_X(X_k)L_D^k(\hat{d}_{k-1}) \le t_k, \end{cases}$$

where for each $j_{k-1} = 0, 1$

$$L_D^k(j_{k-1}) = \frac{p_1^k(j_{k-1})}{p_0^k(j_{k-1})} = \frac{\mathbb{P}_1(\hat{d}_{k-1} = j_{k-1})}{\mathbb{P}_0(\hat{d}_{k-1} = j_{k-1})},$$

and $\mathbb{P}_j(\hat{d}_{k-1} = j_{k-1})$, $j = 0, 1$ is given by

$$\begin{aligned} \mathbb{P}_j(\hat{d}_{k-1} = j_{k-1}) &= q_k(1 - \mathbb{P}_j(d_{k-1} = j_{k-1})) + (1 - q_k)\mathbb{P}_j(d_{k-1} = j_{k-1}) \\ &= q_k + (1 - 2q_k)\mathbb{P}_j(d_{k-1} = j_{k-1}). \end{aligned} \tag{2.4}$$

Let $t_k(\hat{d}_{k-1}) = t_k/L_D^k(\hat{d}_{k-1})$ be the testing threshold for $L_X(X_k)$ when $\hat{d}_{k-1}$ is received. Then, the likelihood ratio test can be rewritten as

$$d_k = \begin{cases} 1 & \text{if } L_X(X_k) > t(\hat{d}_{k-1}), \\ 0 & \text{if } L_X(X_k) \le t(\hat{d}_{k-1}). \end{cases}$$

From (2.4), we notice that $\mathbb{P}_j(\hat{d}_{k-1})$ depends linearly on $\mathbb{P}_j(d_{k-1})$. Without loss of generality, henceforth we assume that $q_k \le 1/2$.[4] It is obvious that $t_k(0) \ge t_k(1)$ because $L_D^k(j) = \mathbb{P}_1(\hat{d}_{k-1} = j)/\mathbb{P}_0(\hat{d}_{k-1} = j)$ is non-decreasing in j. Therefore, the likelihood ratio test becomes

$$d_k = \begin{cases} 1 & \text{if } L_X(X_k) > t_k(0), \\ 0 & \text{if } L_X(X_k) \le t_k(1), \\ \hat{d}_{k-1} & \text{otherwise}, \end{cases}$$

and we can write the Type I and Type II error probabilities, denoted by $\mathbb{P}_0(d_k = 1)$ and $\mathbb{P}_1(d_k = 0)$, respectively, as follows:

$$\begin{aligned} \mathbb{P}_0(d_k = 1) = \mathbb{P}_0(L_X(X_k) > t_k(0))\mathbb{P}_0(\hat{d}_{k-1} = 0) \\ + \mathbb{P}_0(L_X(X_k) > t_k(1))\mathbb{P}_0(\hat{d}_{k-1} = 1) \end{aligned}$$

[4] Note that the system is symmetric with respect to $q_k = 1/2$. For example, if the probability of flipping is 1, i.e., $q_k = 1$, then the receiver can revert the received decision back since it knows the predecessor always 'lies.'

and

$$\mathbb{P}_1(d_k = 0) = \mathbb{P}_1(L_X(X_k) \leq t_k(0))\mathbb{P}_1(\hat{d}_{k-1} = 0) \\ + \mathbb{P}_1(L_X(X_k) \leq t_k(1))\mathbb{P}_1(\hat{d}_{k-1} = 1).$$

The total error probability at stage k is

$$\begin{aligned}\mathbb{P}_e^k &= \pi_0\mathbb{P}_0(d_k = 1) + \pi_1\mathbb{P}_1(d_k = 0) \\ &= \pi_0(\mathbb{P}_0(L_X(X_k) > t_k(0)) \\ &\quad + \mathbb{P}_0(t_k(1) < L_X(X_k) \leq t_k(0))\mathbb{P}_0(\hat{d}_{k-1} = 1)) \\ &\quad + \pi_1(\mathbb{P}_1(t_k(1) < L_X(X_k) \leq t_k(0))\mathbb{P}_1(\hat{d}_{k-1} = 0) \\ &\quad + \mathbb{P}_1(L_X(X_k) \leq t_k(1))).\end{aligned}$$

We prove the claim by contradiction. Suppose that there exists a strategy such that $\mathbb{P}_e^k \to 0$ as $k \to \infty$. Then, we must have $\mathbb{P}_0(L_X(X_k) > t_k(0)) \to 0$ and $\mathbb{P}_1(L_X(X_k) \leq t_k(1)) \to 0$. Recall that $\mathbb{P}_0^X$ and $\mathbb{P}_1^X$ are equivalent measures. We have $\mathbb{P}_1(L_X(X_k) > t_k(0)) \to 0$ and $\mathbb{P}_0(L_X(X_k) \leq t_k(1)) \to 0$. These imply that $\mathbb{P}_j(t_k(1) < L_X(X_k) \leq t_k(0)) \to 1$ for $j = 0, 1$. But

$$\begin{aligned}\mathbb{P}_j(\hat{d}_{k-1} = 1 - j) &= q_k(1 - \mathbb{P}_j(d_{k-1} = 1 - j)) \\ &\quad + (1 - q_k)\mathbb{P}_j(d_{k-1} = 1 - j) \\ &= q_k + (1 - 2q_k)\mathbb{P}_j(d_{k-1} = 1 - j),\end{aligned}$$

which is bounded below by q_k. Hence $\mathbb{P}_e^k$ is also bounded below away from 0 in the asymptotic regime. This contradiction implies that $\mathbb{P}_e^k$ does not converge to 0.

We now extend the proof to the case where each node observes $m_k \geq 1$ previous decisions. The likelihood ratio test in this case is given by

$$d_k = \begin{cases} 1 & \text{if } L_X(X_k) > t(\hat{d}_{k-1}, \ldots, \hat{d}_{k-m_k}), \\ 0 & \text{if } L_X(X_k) \leq t(\hat{d}_{k-1}, \ldots, \hat{d}_{k-m_k}), \end{cases}$$

where $t(\hat{d}_{k-1},\ldots,\hat{d}_{k-m_k}) = t_k/L_D^k(\hat{d}_{k-1},\ldots,\hat{d}_{k-m_k})$ denotes the testing threshold. Note that among all possible combinations of $\{\hat{d}_{k-1},\ldots,\hat{d}_{k-m_k}\}$, it suffices to assume that the likelihood ratio in the case where each decision equals 0 (denoted by $\mathbf{0}^{m_k}$) is the smallest and that in the case where each decision equals 1 (denoted by $\mathbf{1}^{m_k}$) is the largest. Otherwise, we can always find the smallest and largest likelihood ratio. The case where the likelihood ratios for all possible combinations are equal can be excluded because it means the decisions observed have no useful information for hypothesis testing; and the node has to make a decision based on its own measurement, in which case the error probability does not converge to 0.

From these, we can define the Type I and II error probabilities:

$$
\begin{aligned}
&\mathbb{P}_0(d_k = 1) = \\
&\mathbb{P}_0(L_X(X_k) > t_k(\mathbf{0}_k^m))\mathbb{P}_0(\hat{d}_{k-1} = 0, \hat{d}_{k-2} = 0, \ldots, \hat{d}_{k-m_k} = 0) \\
&\quad + \mathbb{P}_0(L_X(X_k) > t_k(1,0,0,\ldots,0))\mathbb{P}_0(\hat{d}_{k-1} = 1, \hat{d}_{k-2} = 0, \ldots, \hat{d}_{k-m_k} = 0) + \ldots \\
&\quad + \mathbb{P}_0(L_X(X_k) > t_k(\mathbf{1}^{m_k}))\mathbb{P}_0(\hat{d}_{k-1} = 1, \hat{d}_{k-2} = 1, \ldots, \hat{d}_{k-m_k} = 1) \\
&= \mathbb{P}_0(L_X(X_k) > t_k(\mathbf{0}^{m_k})) + \mathbb{P}_0(t_k(1,0,0,\ldots,0) < L_X(X_k) \le t_k(\mathbf{0}^{m_k})) \\
&\quad \mathbb{P}_0(\hat{d}_{k-1} = 1, \hat{d}_{k-2} = 0, \ldots, \hat{d}_{k-m_k} = 0) + \ldots \\
&\quad + \mathbb{P}_0(t_k(\mathbf{1}^{m_k}) < L_X(X_k) \le t_k(\mathbf{0}^{m_k}))\mathbb{P}_0(\hat{d}_{k-1} = 1, \hat{d}_{k-2} = 1, \ldots, \hat{d}_{k-m_k} = 1)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{P}_1(d_k = 0) = \\
&\mathbb{P}_1(L_X(X_k) \le t_k(\mathbf{0}^{m_k}))\mathbb{P}_1(\hat{d}_{k-1} = 0, \hat{d}_{k-2} = 0, \ldots, \hat{d}_{k-m_k} = 0) \\
&\quad + \mathbb{P}_1(L_X(X_k) \le t_k(1,0,0,\ldots,0))\mathbb{P}_1(\hat{d}_{k-1} = 1, \hat{d}_{k-2} = 0, \ldots, \hat{d}_{k-m_k} = 0) + \ldots \\
&\quad + \mathbb{P}_1(L_X(X_k) \le t_k(\mathbf{1}^{m_k}))\mathbb{P}_1(\hat{d}_{k-1} = 1, \hat{d}_{k-2} = 1, \ldots, \hat{d}_{k-m_k} = 1) \\
&= \mathbb{P}_1(t_k(\mathbf{1}^{m_k}) < L_X(X_k) \le t_k(\mathbf{0}^{m_k}))\mathbb{P}_1(\hat{d}_{k-1} = 0, \hat{d}_{k-2} = 0, \ldots, \hat{d}_{k-m_k} = 0) \\
&\quad + \mathbb{P}_1(t_k(\mathbf{1}^{m_k}) < L_X(X_k) \le t_k(1,0,0,\ldots,0))\mathbb{P}_0(\hat{d}_{k-1} = 1, \hat{d}_{k-2} = 0, \ldots, \hat{d}_{k-m} = 0) \\
&\quad + \ldots + \mathbb{P}_1(L_X(X_k) \le t_k(\mathbf{1}^{m_k})).
\end{aligned}
$$

With the similar argument as that in the tandem network case, we have

$$\mathbb{P}_e^k = \pi_0 \mathbb{P}_0(d_k = 1) + \pi_1 \mathbb{P}_1(d_k = 0).$$

Suppose that $\mathbb{P}_e^k \to 0$ as $k \to \infty$. Then, we must have $\mathbb{P}_0(L_X(X_k) > t_k(\mathbf{0}^{m_k})) \to 0$ and $\mathbb{P}_1(L_X(X_k) \leq t_k(\mathbf{1}^{m_k})) \to 0$. Recall that $\mathbb{P}_0^X$ and $\mathbb{P}_1^X$ are equivalent measures. Hence we have $\mathbb{P}_j(t_k(\mathbf{1}^{m_k}) < L_X(X_k) \leq t_k(\mathbf{0}^{m_k})) \to 1$ for $j = 0, 1$. We have

$$\begin{aligned}
&\mathbb{P}_j(\hat{d}_{k-1} = j_{k-1}, \hat{d}_{k-2} = j_{k-2}, \ldots, \hat{d}_{k-m_k} = j_{k-m_k}) = \\
&\mathbb{P}_j(\hat{d}_{k-1} = j_{k-1} | \hat{d}_{k-2} = j_{k-2}, \ldots, \hat{d}_{k-m_k} = j_{k-m_k}) \cdot \\
&\mathbb{P}_j(\hat{d}_{k-2} = j_{k-2} | \hat{d}_{k-3} = j_{k-3}, \ldots, \hat{d}_{k-m_k} = j_{k-m_k}) \cdot \\
&\ldots \mathbb{P}_j(\hat{d}_{k-m_k+1} = j_{k-m_k+1} | \hat{d}_{k-m_k} = j_{k-m_k}) \cdot \\
&\mathbb{P}_j(\hat{d}_{k-m_k} = j_{k-m_k}).
\end{aligned}$$

We already know that $\mathbb{P}_j(\hat{d}_{k-m_k} = j_{k-m_k})$ is bounded away from 0 by q_k. Similarly, we can show

$$\begin{aligned}
&\mathbb{P}_j(\hat{d}_{k-i} = j_{k-i} | \hat{d}_{k-i-1} = j_{k-i-1}, \ldots, \hat{d}_{k-m_k} = j_{k-m_k}) \\
&= (1 - q_k)\mathbb{P}_j(d_{k-i} = j_{k-i} | \ldots, \hat{d}_{k-m_k} = j_{k-m_k}) \\
&\quad + q_k(1 - \mathbb{P}_j(d_{k-i} = j_{k-i} | \ldots, \hat{d}_{k-m_k} = j_{k-m_k})) \\
&= q_k + (1 - 2q_k)\mathbb{P}_j(d_{k-i} = j_{k-i} | \ldots, \hat{d}_{k-m_k} = j_{k-m_k}).
\end{aligned}$$

Hence $\mathbb{P}_e^k$ is also bounded below by $q_k^{m_k} \geq q_k^C$. This contradiction implies that $\mathbb{P}_e^k$ does not converge to 0 with any decision strategy.

□

2.3.2 Unbounded Memory

In this section, we consider the case where a_k can observe all its predecessors; i.e., $m_k = k - 1$. We will show that using the myopic decision strategy, the error probability converges to 0 in the presence of random flipping when the flipping probabilities are bounded away from $1/2$. In the case where the flipping probability converges to $1/2$, we derive a necessary condition on the convergence rate of the flipping probability such that the error probability converges to 0. Moreover, we precisely describe the relationship between the convergence rate of the flipping probability and the convergence rate of the error probability.

If we state the conditions on the private signal distributions in a symmetric way, then it suffices to consider the case when the true hypothesis is H_0. In this case, our aim is to show that the Type I error probability converges to 0, i.e., $\mathbb{P}_0(d_k = 1) \to 0$. We consider the myopic decision strategy; i.e., the decision made by the kth node is on the basis of the MAP test. Again, the corruption from d_k to $\hat{d}_k$ is in the form of a binary symmetric channel with flipping probability denoted by q_k. Without loss of generality, we assume that $q_k \leq 1/2$ (because of symmetry). We define the *public likelihood ratio* of $\mathcal{D}_k = (j_1, j_2, \ldots, j_k)$ to be

$$L_k(\mathcal{D}_k) = \frac{p_1^k(\mathcal{D}_k)}{p_0^k(\mathcal{D}_k)} = \frac{\mathbb{P}_1(\hat{D}_k = \mathcal{D}_k)}{\mathbb{P}_0(\hat{D}_k = \mathcal{D}_k)}.$$

We will consider two cases:

1) The flipping probabilities are bounded away from $1/2$ for all k; i.e., there exists $c > 0$ such that $q_k \leq 1/2 - c$ for all k. This ensures that the corrupted decision still contains some useful information about the true hypothesis. We call this the case of *uniformly informative nodes*.

2) The flipping probabilities q_k converge to $1/2$; i.e., $q_k \to 1/2$ as $k \to \infty$. This means that the broadcasted decisions become increasingly uninformative as we move towards the latter nodes. We call this the case of *asymptotically uninformative nodes*.

2.3.2.1 Uniformly informative nodes

We first show that the error probability converges to 0. Recall that $\bar{b}_k = \mathbb{P}(\theta = H_1 | X_k)$ denotes the private belief given by signal X_k. Let $(\mathbb{G}_0, \mathbb{G}_1)$ be the conditional distributions of the private belief $\bar{b}_k$:

$$\mathbb{G}_j(r) = \mathbb{P}_j(\bar{b}_k \leq r).$$

Note that $\mathbb{G}_j$ does not depend on k because the X_ks are identically distributed. These distributions exhibit two important properties:

a) *Proportionality*: This property is easy to get from Bayes' rule: for all $r \in (0,1)$, we have

$$\frac{d\mathbb{G}_1}{d\mathbb{G}_0}(r) = \frac{r}{1-r},$$

where $d\mathbb{G}_1/d\mathbb{G}_0$ is the Radon-Nikodym derivative of their associated probability measures.

b) *Dominance*: $\mathbb{G}_1(r) < \mathbb{G}_0(r)$ for all $r \in (0,1)$, and $\mathbb{G}_j(0) = 0$ and $\mathbb{G}_j(1) = 1$ for $j = 0, 1$. Moreover, $\mathbb{G}_1(r)/\mathbb{G}_0(r)$ is monotone non-decreasing as a function of r.

We note that the dominance property can be shown using Assumption 3) and the details of the proof is omitted.

We define an increasing sequence $\{\mathcal{F}_k\}$ of σ-algebras as follows:

$$\mathcal{F}_k = \sigma\langle X_1, X_2, \ldots, X_k; \hat{d}_1, \hat{d}_2, \ldots, \hat{d}_k\rangle.$$

Evidently $\hat{d}_k$ and $L_k(\hat{D}_k)$ are adapted to this sequence of σ-algebras. Moreover, given $\hat{D}_{k-1} = \{\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_{k-1}\}$ and X_k, the decision d_k is completely determined. Therefore, d_k is also adapted to this sequence of σ-algebras.

Lemma 2.3.1. *Under hypothesis H_0, the public likelihood ratio sequence $\{L_k(\hat{D}_k)\}$ is a martingale with respect to $\{\mathcal{F}_k\}$ and $L_k(\hat{D}_k)$ converges to a finite limit almost surely.*

Proof. The expectation of $L_{k+1}(\hat{D}_{k+1})$ conditioned on H_0 and $\mathcal{F}_k$ is

$$\begin{aligned}
\mathbb{E}_0[L_{k+1}(\hat{D}_{k+1})|\mathcal{F}_k] &= \sum_{\hat{d}_{k+1}=0,1} \mathbb{P}_0(\hat{d}_{k+1}|\mathcal{F}_k)L_{k+1}(\hat{D}_{k+1}) \\
&= \sum_{\hat{d}_{k+1}=0,1} \mathbb{P}_0(\hat{d}_{k+1}|\mathcal{F}_k)L_k(\hat{D}_k)\frac{\mathbb{P}_1(\hat{d}_{k+1}|\mathcal{F}_k)}{\mathbb{P}_0(\hat{d}_{k+1}|\mathcal{F}_k)} \\
&= L_k(\hat{D}_k) \sum_{\hat{d}_{k+1}=0,1} \mathbb{P}_0(\hat{d}_{k+1}|\mathcal{F}_k)\frac{\mathbb{P}_1(\hat{d}_{k+1}|\mathcal{F}_k)}{\mathbb{P}_0(\hat{d}_{k+1}|\mathcal{F}_k)} \\
&= L_k(\hat{D}_k).
\end{aligned}$$

Moreover, note that

$$\int |L_1(\hat{D}_1)|\, d\mathbb{P}_0 = 1 < \infty.$$

Since $L_k(\hat{D}_k)$ a non-negative martingale, by Doob's martingale convergence theorem [68], it converges almost surely to a finite limit. □

Let L_∞ be the almost sure limit of $L_k(\hat{D}_k)$ conditioned on H_0, and note that $L_\infty < \infty$ almost surely. This claim holds for both cases 1 and 2. By (2.2), we know that the public belief $b_k < 1$ almost surely. The implication is that the public belief cannot go completely wrong. Moreover, for case 1, we can show that the public likelihood ratio converges to 0 almost surely.

Lemma 2.3.2. *Suppose that the flipping probabilities are bounded away from* $1/2$*. Then under* H_0*, we have* $L_\infty = 0$ *almost surely.*

Proof. For the public likelihood ratio, we have the following recursion:

$$\begin{aligned}
L_{k+1}(\hat{D}_{k+1}) &= \frac{\mathbb{P}_1(\hat{D}_{k+1})}{\mathbb{P}_0(\hat{D}_{k+1})} \\
&= \frac{\mathbb{P}_1(\hat{d}_{k+1}|\hat{D}_k)}{\mathbb{P}_0(\hat{d}_{k+1}|\hat{D}_k)}L_k(\hat{D}_k).
\end{aligned} \tag{2.5}$$

Consider the event $A = \{L_\infty > 0\}$. On A, we have

$$\frac{\mathbb{P}_1(\hat{d}_{k+1}|\hat{D}_k)}{\mathbb{P}_0(\hat{d}_{k+1}|\hat{D}_k)} \to 1, \tag{2.6}$$

almost everywhere. Now

$$\frac{\mathbb{P}_1(\hat{d}_{k+1}|\hat{D}_k)}{\mathbb{P}_0(\hat{d}_{k+1}|\hat{D}_k)} = \frac{\sum_{d_{k+1}} \mathbb{P}_1(d_{k+1}|\hat{D}_k)\mathbb{P}(\hat{d}_{k+1}|d_{k+1})}{\sum_{d_{k+1}} \mathbb{P}_0(d_{k+1}|\hat{D}_k)\mathbb{P}(\hat{d}_{k+1}|d_{k+1})}$$
$$= \frac{\mathbb{P}_1(d_{k+1}|\hat{D}_k)(1-2q_k)+q_k}{\mathbb{P}_0(d_{k+1}|\hat{D}_k)(1-2q_k)+q_k}. \quad (2.7)$$

Equation (2.7) together with (2.6) implies

$$\frac{\mathbb{P}_1(d_{k+1}|\hat{D}_k)}{\mathbb{P}_0(d_{k+1}|\hat{D}_k)} \to 1,$$

or $\mathbb{P}_j(d_{k+1}|\hat{D}_k) \to 0$ for $j = 0, 1$, almost everywhere on A. We note that another possible situation is that there exists a subsequence of $\{\frac{\mathbb{P}_1(d_{k+1}|\hat{D}_k)}{\mathbb{P}_0(d_{k+1}|\hat{D}_k)}\}$ that converges to 1 and for its complement subsequence, we have $\mathbb{P}_j(d_{k+1}|\hat{D}_k) \to 0$ for $j = 0, 1$, almost everywhere on A. However, the proof for this situation is similar with others and it is omitted.

We will show that A has probability 0. Suppose that there exists $\omega \in A$ such that

$$\lim_{k\to\infty} \frac{\mathbb{P}_1(d_{k+1} = d_{k+1}(\omega)|\hat{D}_k = \hat{D}_k(\omega))}{\mathbb{P}_0(d_{k+1} = d_{k+1}(\omega)|\hat{D}_k = \hat{D}_k(\omega))} = 1.$$

Note that $d_{k+1}(\omega) = 0$ or 1. Without loss of generality, consider the situation where $d_{k+1}(\omega) = 0$, we have

$$\lim_{k\to\infty} \frac{\mathbb{P}_1(d_{k+1} = 0|\hat{D}_k = \hat{D}_k(\omega))}{\mathbb{P}_0(d_{k+1} = 0|\hat{D}_k = \hat{D}_k(\omega))} = 1. \quad (2.8)$$

Note that the statement $d_{k+1} = 0$ is equivalent to

$$L_X(X_{k+1})L_k(\hat{D}_k) \leq \frac{\pi_0}{\pi_1}.$$

Because of the independence between X_{k+1} and $\hat{D}_k$, we obtain

$$\mathbb{P}_j(d_{k+1} = 0|\hat{D}_k = \hat{D}_k(\omega)) = \mathbb{P}_j\left(L_X(X_{k+1})L_k(\hat{D}_k) \leq \frac{\pi_0}{\pi_1}\middle| \hat{D}_k = \hat{D}_k(\omega)\right)$$
$$= \mathbb{P}_j\left(L_X(X_{k+1})L_k(\hat{D}_k(\omega)) \leq \frac{\pi_0}{\pi_1}\right).$$

Thus (2.8) is equivalent to

$$\lim_{k\to\infty} \frac{\mathbb{P}_1(L_X(X_{k+1})L_k(\hat{D}_k(\omega)) \leq \frac{\pi_0}{\pi_1})}{\mathbb{P}_0(L_X(X_{k+1})L_k(\hat{D}_k(\omega)) \leq \frac{\pi_0}{\pi_1})} = 1. \quad (2.9)$$

By (2.1) and the definitions of $\mathbb{G}_1$ and $\mathbb{G}_0$, (2.9) is equivalent to

$$\lim_{k\to\infty} \frac{\mathbb{G}_1((1+L_k(\hat{D}_k(\omega)))^{-1})}{\mathbb{G}_0((1+L_k(\hat{D}_k(\omega))^{-1})} = 1.$$

Because $\mathbb{G}_1$ and $\mathbb{G}_0$ are right-continuous, we have $\mathbb{G}_1/\mathbb{G}_0$ is also right-continuous. Moreover, $\mathbb{G}_1/\mathbb{G}_0$ is monotone non-decreasing. Therefore, we have

$$\frac{\mathbb{G}_1((1+L_\infty(\omega))^{-1})}{\mathbb{G}_0((1+L_\infty(\omega))^{-1})} = 1.$$

However, this contradicts the dominance property (described earlier). We can use a similar argument to show that there does not exist ω such that $\mathbb{P}_j(d_{k+1} = d_{k+1}(\omega)|\hat{D}_k = \hat{D}_k(\omega)) \to 0$. Therefore, no such ω exists and this implies that $\mathbb{P}_0(A) = 0$. Hence, $\mathbb{P}_0(L_\infty = 0) = 1$. $\square$

Theorem 2.3.2. *Suppose that the flipping probabilities are bounded away from* $1/2$*. Then,* $\mathbb{P}_e^k \to 0$ *as* $k \to \infty$*.*

Proof. We know that the likelihood ratio test states that a_k decides 1 if and only if $\bar{b}_k > 1 - b_{k-1}$. The probability of deciding 1 given that H_0 is true (Type I error) is given by

$$\begin{aligned}\mathbb{P}_0(d_k = 1) &= \mathbb{P}_0(\bar{b}_k > 1 - b_{k-1}) \\ &= \mathbb{E}_0(1 - \mathbb{G}_0(1 - b_{k-1})).\end{aligned}$$

Since $L_\infty = 0$ almost surely, we have $b_k \to 0$ almost surely. We have

$$\lim_{k\to\infty} \mathbb{P}_0(d_k = 1) = \lim_{k\to\infty} \mathbb{E}_0(1 - \mathbb{G}_0(1 - b_{k-1})).$$

By the bounded convergence theorem, we have

$$\begin{aligned}\lim_{k\to\infty} \mathbb{P}_0(d_k = 1) &= 1 - \mathbb{E}_0(\lim_{k\to\infty} \mathbb{G}_0(1 - b_{k-1})) \\ &= 1 - \mathbb{G}_0(1) = 0.\end{aligned}$$

Similarly, we can prove that $\lim_{k\to\infty} \mathbb{P}_1(d_k = 0) = 0$ (i.e., Type II error probability converges to 0). Therefore, the error probability converges to 0. $\square$

Remark 2.3.1 (Additive Gaussian noise). Note that our convergence proof easily generalizes to the additive Gaussian noise scenario: Suppose that after a_k makes a decision $d_k \in \{0, 1\}$, it broadcasts the decision through a Gaussian broadcasting channel, in other words, the other nodes recieves $\hat{d}_k = F_k d_k + \mathcal{N}_k$, where $F_k \in (0, 1)$ denotes a fading coefficient and $\mathcal{N}_k$ denotes zero-mean Gaussian noise. Then, we can show that the error probability converges to 0 if F_k are bounded away from 0 and the noise variances are bounded for all k. In other words, the signal-to-noise ratios are bounded away from 0.

Now let us consider the convergence rate of the error probability. Without loss of generality, we assume that the prior probabilities are equal; i.e., $\pi_0 = \pi_1 = 1/2$. The following analysis easily generalizes to unequal prior probabilities. Recall that $b_k = \mathbb{P}(\theta = H_1 | \hat{D}_k)$ denotes the public belief. It is easy to see that the error probability converges to 0 if and only if $b_k \to 0$ almost surely given H_0 is true and $b_k \to 1$ almost surely given H_1 is true. Recall the proportionality property:

$$\frac{d\mathbb{G}_1}{d\mathbb{G}_0}(r) = \frac{r}{1-r}.$$

Moreover, we assume $\mathbb{G}_1$ and $\mathbb{G}_0$ are continuous and therefore under each of H_0 and H_1, the density of the private belief exists. By the above property, we can write these densities as follows:

$$f^1(r) = \frac{d\mathbb{G}_1}{dr}(r) = r\rho(r),$$

and

$$f^0(r) = \frac{d\mathbb{G}_0}{dr}(r) = (1-r)\rho(r),$$

where $\rho(r)$ is a non-negative function.

Without loss of generality, we assume that H_0 is the true hypothesis. Moreover, we assume that $\rho(1) > 0$ and ρ is continuous near $r = 1$. This characterizes the behavior of the tail densities. We will generalize our analysis to polynomial tail densities later, where $\rho(r) \to 0$ as $r \to 1$.

The Bayesian update of the public belief when $\hat{d}_{k+1} = 0$ is given by:

$$\begin{aligned}
b_{k+1} &= \mathbb{P}(\theta = H_1|\hat{D}_{k+1}) \\
&= \frac{\mathbb{P}_1(\hat{d}_{k+1} = 0|\hat{D}_k)b_k}{\sum_{j=0,1}\mathbb{P}_j(\hat{d}_{k+1} = 0|\hat{D}_k)\mathbb{P}(\theta = H_j|\hat{D}_k)} \\
&= \frac{(q_k + (1-2q_k)\mathbb{P}_1(d_{k+1} = 0|\hat{D}_k))b_k}{\sum_{j=0,1}(q_k + (1-2q_k)\mathbb{P}_j(d_{k+1} = 0|\hat{D}_k))\mathbb{P}(H_j|\hat{D}_k)}.
\end{aligned} \tag{2.10}$$

It is easy to show that the public belief converges to 0 in the fastest rate if $\hat{d}_k = 0$ for all k. We will establish the rate in this special case to bound the converge rate of the error probability. Notice that $\mathbb{P}(\theta = H_1|\hat{D}_k) = b_k$ and $\mathbb{P}(\theta = H_0|\hat{D}_k) = 1 - b_k$. By Lemma 2.3.2, we have $L_k(\hat{D}_k) \to 0$ almost surely, under H_0. This implies that $b_k \to 0$ almost surely. If b_k is sufficiently small, then we have

$$\begin{aligned}
\mathbb{P}_1(d_{k+1} = 0|\hat{D}_k) &= 1 - \int_{1-b_k}^{1} f^1(x)dx \\
&\simeq 1 - \rho(1)(b_k - \frac{b_k^2}{2})
\end{aligned} \tag{2.11}$$

and

$$\begin{aligned}
\mathbb{P}_0(d_{k+1} = 0|\hat{D}_k) &= 1 - \int_{1-b_k}^{1} f^0(x)dx \\
&\simeq 1 - \rho(1)\frac{b_k^2}{2}.
\end{aligned} \tag{2.12}$$

Note that $\simeq$ means asymptotically equal. We can also calculate the (conditional) Type I error probability:

$$\begin{aligned}
\mathbb{P}_0(d_{k+1} = 1|\hat{D}_k) &= 1 - \mathbb{P}_0(d_{k+1} = 1|\hat{D}_k) \\
&= \int_{1-b_k}^{1} f^0(x)dx \\
&\simeq \rho(1)\frac{b_k^2}{2}.
\end{aligned} \tag{2.13}$$

Note that (2.13) characterizes the relationship between the decay rate of Type I error probability and the decay rate of b_k. Next we derive the decay rate of b_k.

Substituting (2.11) and (2.12) into (2.10) and removing high order terms we obtain

$$b_{k+1} \simeq \frac{(1-q_k)b_k - (1-2q_k)\rho(1)b_k^2}{(1-q_k)}.$$

This implies that

$$b_{k+1} \simeq b_k \left(1 - \frac{1-2q_k}{1-q_k}\rho(1)b_k\right). \tag{2.14}$$

For any sequence that evolves according to (2.14), the following lemma characterizes the convergence rate of the sequence.

Lemma 2.3.3. *Suppose that a non-negative sequence c_k satisfies $c_{k+1} = c_k(1 - \delta c_k^n)$, where $n \geq 2$, $c_1 < 1$, and $\delta > 0$. Then, for sufficiently large k, there exists two constants C_1 and C_2 such that*

$$\frac{C_1}{(\delta k)^{1/n}} \leq c_k \leq \frac{C_2}{(\delta k)^{1/n}}.$$

This implies that $c_k \to 0$ as $k \to \infty$ and $c_k = \Theta(k^{-1/n})$.

Proof. First it is easy to see that $c_k \to 0$ because it is the only fixed point of the recursion. To show the convergence rate, we treat the recursion (2.14) as an ordinary difference equation (ODE). Therefore, we have

$$\frac{dc_k}{dk} = -\delta c_k^{n+1}.$$

The solution to this ODE is for some $C > 0$

$$c_k = \frac{C}{(\delta k)^{1/n}}.$$

Therefore, for sufficiently large k, there exists two constants C_1 and C_2 such that

$$\frac{C_1}{(\delta k)^{1/n}} \leq c_k \leq \frac{C_2}{(\delta k)^{1/n}}.$$

which implies that

$$c_k = \Theta(k^{-1/n}).$$

□

Theorem 2.3.3. *Suppose that the flipping probabilities are bounded away from* $1/2$ *and* $\rho(1)$ *is a non-negative constant. Then, the Type I error probability converges to 0 as* $\Omega(k^{-2})$.

Proof. Using (2.14) and Lemma 2.3.3, we can get the convergence rate of the public belief conditioned on event that $\hat{d}_k = 0$ for all k, in which case we have $b_k = \Theta(k^{-1})$. Recall that the public belief converges to 0 the fastest in this case among all possible outcomes. Therefore, we have $b_k = \Omega(k^{-1})$ almost surely.

Recall that $d_k = 1$ if and only if $\bar{b}_k > 1 - b_{k-1}$. Therefore, the Type I error probability is given by

$$\begin{aligned} \mathbb{P}_0(d_k = 1) &= \mathbb{P}_0(\bar{b}_k > 1 - b_{k-1}) \\ &= \mathbb{E}_0(1 - \mathbb{G}_0(1 - b_{k-1})). \end{aligned} \tag{2.15}$$

Because ρ is continuous at 1, we have if $x < 1$ is sufficiently close to 1, i.e., $1 - x$ is positive and sufficiently small, then

$$\begin{aligned} 1 - \mathbb{G}_0(x) &= \int_x^1 (1-x)\rho(x)dx \\ &\geq \frac{\rho(1)}{2} \int_x^1 (1-x)dx \\ &= \frac{\rho(1)(1-x)^2}{4}. \end{aligned} \tag{2.16}$$

From (2.15) and (2.16) and invoking Jensen's Inequality, we obtain

$$\begin{aligned} \mathbb{P}_0(d_k = 1) &\geq \frac{\rho(1)}{4}\mathbb{E}_0[b_{k-1}^2] \\ &\geq \frac{\rho(1)}{4}(\mathbb{E}_0[b_{k-1}])^2. \end{aligned} \tag{2.17}$$

Because $b_k = \Omega(k^{-1})$ almost surely, we have $\mathbb{P}_0(d_k = 1) = \Omega(k^{-2})$. □

Assume that $\rho(0) > 0$ and ρ is continuous at 0. Then, we can use the same method to calculate the decay rate of the Type II error probability, which is the same as that of the Type I error probability. Note that the decay rate of the error probability depends linearly on $(1 - 2q_k)^{-2}$.

2.3.2.2 Asymptotically uninformative nodes

In this part, we consider the case where $q_k \to 1/2$ as $k \to \infty$, which means that the broadcasted decisions become asymptotically uninformative. Let

$$Q_k = \frac{1 - 2q_k}{1 - q_k}.$$

Note that $q_k \to 1/2$ implies that $Q_k \to 0$. This parameter measures how "informative" the corrupted decision is: For example, if $q_k = 0$ (where there is no flipping), then the decision is maximally informative in terms of updating the public belief. However if $q_k = 1/2$, in which case $Q_k = 0$, then the decision is completely uninformative in terms of updating the public belief.

We will derive a necessary condition on the decay rate of Q_k to 0 for the public belief b_k to converge to 0 under H_0, which gives us a necessary condition on Q_k for asymptotic learning. For any sequence that evolve according to (2.14), the following lemma characterizes necessary and sufficient conditions such that the sequence converges to 0.

Lemma 2.3.4. *Suppose that a non-negative sequence $\{c_k\}$ follows $c_{k+1} = c_k(1 - \delta_k c_k^n)$, where $n \geq 1$, $c_1 > 0$, and $\delta_k > 0$. Then, c_k converges to 0 if and only if there exists k_0 such that $\sum_{k=k_0}^{\infty} \delta_k = \infty$.*

Proof. We will use the following claim to prove the lemma: For a non-negative sequence satisfying $c_{k+1} = c_k(1 - r_k)$, where $c_1 > 0$ and $r_k \in [0, 1)$, we have $c_k \to 0$ if and only if there exists k_0 such that $\sum_{k=k_0}^{\infty} r_k = \infty$. To show this claim, we have

$$c_{k+1} = c_1 \prod_{i=1}^{k} (1 - r_i).$$

Applying natural logarithm, we obtain

$$\ln c_{k+1} = \ln c_1 + \sum_{i=1}^{k} \ln(1 - r_i).$$

From the above equation, we have $c_k \to 0$ if and only if $\sum_{i=1}^{\infty} \ln(1-r_i) = -\infty$. In the case where there exists a subsequence of $\{r_k\}$ such that the subsequence is bounded away from 0, we have $\sum_{i=1}^{\infty} \ln(1-r_i) = -\infty$. Therefore, $c_k \to 0$ as $k \to \infty$. In the case where $r_k \to 0$, there exists k_0 such that $r_i \le -\ln(1-r_i) \le 2r_i$ for all $i \ge k_0$. Therefore, we have $c_k \to 0$ if and only if $\sum_{k=k_0}^{\infty} r_k = \infty$.

We now show the lemma. First we show that the condition is necessary. Suppose that $c_k \to 0$. Then, we have $\sum_{k=1}^{\infty} \delta_k c_k^n = \infty$. Since $c_k < 1$, we have $\sum_{k=1}^{\infty} \delta_k = \infty$. Second we show by contradiction that the condition is sufficient. Suppose that there exist k_0 such that $\sum_{k=k_0}^{\infty} \delta_k = \infty$ and c_k does not converge to 0. Since c_k is monotone decreasing, c_k must converge to a nonzero limit c. Therefore, for all k, we have $c_k \ge c$. Then, we have $c_{k+1} \le c_k(1-\delta_k c^n)$. We have

$$\sum_{k=k_0}^{\infty} \delta_k c^n = c^n \sum_{k=k_0}^{\infty} \delta_k = \infty.$$

Therefore, we have $c_k \to 0$. □

Theorem 2.3.4. *Suppose that there exists $p > 1$ such that*

$$Q_k = O\left(\frac{1}{k(\log k)^p}\right).$$

Then, the public belief converges to a nonzero limit almost surely.

Proof. Suppose that there exists $p > 1$ such that $Q_k = O\left(1/(k(\log k)^p)\right)$. Then, we have

$$\sum_{k=2}^{\infty} Q_k < \infty.$$

Therefore, by Lemma 2.3.4, b_k in (2.14) does not converge to 0. Recall that (2.14) represents the recursion of b_k conditioned on the event that the node broadcast decisions are all 0. Therefore, the public belief is the smallest among all possible outcomes. Hence, the public belief converges to a nonzero limit almost surely. □

By (2.17), it is evident that if b_k converges to a nonzero limit almost surely, then $\mathbb{P}_0(d_k = 1)$ is bounded away from 0 and $\mathbb{P}_0(d_k = 0)$ is bounded away from 1. Therefore, the system does not asymptotically learn the underlying truth. Hence Theorem 2.3.4 provides a necessary condition for asymptotically learning.

Theorem 2.3.4 also implies that for there to be a nonzero probability that the public belief converges to zero, we must have that there exists $p \leq 1$ such that $Q_k = \Omega(1/k(\log k)^p)$. If the public belief does not converge to zero, then it is impossible for there to be an eventual collective arrival at the true hypothesis. To explain this further, Let $\mathcal{H}$ denote the event that there exists a (random) k_0 such that the sequence of decisions $d_k = 0$ for all $k \geq k_0$. Occurrence of this event signifies that after a finite number of decisions, the agents arrive at the true underlying state. Such an outcome also means that, eventually, each agent's private signal is overpowered by the past collective true verdict, so that a false decision is never again declared. In the literature on social learning, this phenomenon is called *information cascade* (e.g., [66], [67]) or *herding* (e.g., [15]). We use $\mathcal{L}$ to denote the event $\{b_k \to 0\}$. Notice that $\mathcal{H}$ occurs only if $\mathcal{L}$ occurs. Hence, $\mathcal{H}$ is a subset of the event that $b_k \to 0$, i.e., $\mathcal{H} \subset \mathcal{L}$. These leads to the following corollary of Theorem 2.3.4.

Corollary 2.3.1. *If $Q_k = O(1/k(\log k)^p)$ for some $p > 1$, then $\mathbb{P}(\mathcal{H}) = 0$.*

So, by the corollary above, only if $Q_k = \Omega(1/k(\log k)^p)$ for some $p \leq 1$ can we hope for there to be a nonzero probability that $b_k \to 0$ and thus of information cascade to the truth. Even under the situation that $b_k \to 0$, i.e., conditioned on $\mathcal{L}$, we expect that the *rate* at which $b_k \to 0$ depends on the scaling law of Q_k. The following theorem relates the scaling laws of $\{Q_k\}$ with those of $\{b_k\}$ and the Type I error probability sequence $\{\mathbb{P}_0(d_k = 1)\}$.

Theorem 2.3.5. *Conditioned on $\mathcal{L}$, we have the following:*

(i) Suppose that $Q_k = \Theta(1/k^{1-p})$ where $p \in (0, 1)$. Then, $b_k = \Omega(k^{-p})$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega(k^{-2p})$.

(ii) Suppose that $Q_k = \Theta(1/k)$. Then, $b_k = \Omega(1/\log k)$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega(1/(\log k)^2)$.

(iii) Suppose that $Q_k = \Theta\left(1/(k(\log k)^p)\right)$ where $p \in (0,1)$. Then, $b_k = \Omega(1/(\log k)^q)$ almost surely, where $1/q + 1/p = 1$, and $\mathbb{P}_0(d_k = 1) = \Omega(1/(\log k)^{2q})$.

(iv) Suppose that $Q_k = \Theta\left(1/(k \log k)\right)$. Then, $b_k = \Omega(1/\log\log k)$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega(1/(\log\log k)^2)$.

Proof. (i). Suppose that $Q_k = \Theta(1/k^{1-p})$ where $p \in (0,1)$. Conditioned on $\mathcal{H}$, we have recursion (2.14) for the public belief b_k. Using this recursion, we can get similar results as those in Lemma 2.3.3, that is, there exists $C_1 > 0$ and $C_2 > 0$ such that

$$\frac{C_1}{kQ_k} \leq b_k \leq \frac{C_2}{kQ_k}. \tag{2.18}$$

Plugging in the convergence rate of Q_k in (2.18) establishes the claim.

(ii)-(iv). Suppose that $Q_k = \Theta(1/k(\log k)^p)$, where $p \in [0,1]$. Then, by (2.14), we have

$$b_{k+1} - b_k = \frac{Cb_k^2}{k(\log k)^p}$$

for some constant $C > 0$. For $p = 0$, the solution to this ODE satisfies $b_k = \Theta(1/\log k)$, which proves (ii). When $p \in (0,1)$, the solution satisfies $b_k = \Theta(1/(\log k)^q)$, where $1/q + 1/p = 1$. This establishes (iii). Finally, when $p = 1$, the solution satisfies $b_k = \Theta(1/\log\log k)$. Note that all these rates are derived conditioned on $\mathcal{H}$. By the fact that conditioned on $\mathcal{H}$, the decay rate is the fastest among all outcomes, we obtain the desired results. Having established the convergence rate of b_k, the convergence rate for the error probability in each claim follows from (2.17).

□

Note that Theorem 2.3.5 provides upper bounds for the convergence rates of the public belief and error probability. However, recall that $\mathcal{H}$ is a subset of the event that $b_k \to 0$.

Therefore, even if $b_k \to 0$ with certain probability, the probability of $\mathcal{H}$ is not guaranteed to be nonzero. Next we provide a necessary condition such that the probability of $\mathcal{H}$ is nonzero.

Theorem 2.3.6. *Suppose that there exists $p \leq 1$ such that*

$$Q_k = O\left(\frac{(p + \log k)(\log k)^{p-1}}{(k(\log k)^p)^{1/2}}\right).$$

Then, we have $\mathbb{P}(\mathcal{H}) = 0$.

Proof. We first state a key lemma which is a corollary of the Borel-Cantelli lemma [68]. Consider a probability space $(S, \mathcal{S}, \mathcal{P})$ and a sequence of events $\{\mathcal{E}_k\}$ in $\mathcal{S}$. We define the limit superior of $\{\mathcal{E}_k\}$ as follows:

$$\limsup_{k\to\infty} \mathcal{E}_k \equiv \bigcap_{k=1}^{\infty}\left(\bigcup_{n=k} \mathcal{E}_n\right).$$

Note that this is the event that infinitely many of the $\mathcal{E}_k$ occur. We use $\mathcal{E}_k^C$ to denote the complement of $\mathcal{E}_k$.

Lemma 2.3.5. *Suppose that*

$$\sum_{k=1}^{\infty} \mathcal{P}(\mathcal{E}_k | \mathcal{E}_{k-1}^C, \mathcal{E}_{k-2}^C, \ldots, \mathcal{E}_1^C) = \infty.$$

Then,

$$\mathcal{P}(\limsup_{k\to\infty} \mathcal{E}_k) = 1.$$

The proof of this lemma is omitted. Now we prove the theorem. Let $\mathcal{E}_k$ be the event that $d_k = 1$, i.e., a_k makes the wrong decision given H_0. Notice that $\mathcal{E}_k^C$ is the event that $d_k = 0$. If

$$Q_k = O\left(\frac{(p + \log k)(\log k)^{p-1}}{(k(\log k)^p)^{1/2}}\right),$$

then using the similar analysis as those in Theorem 2.3.5, we have

$$\mathbb{P}_0(\mathcal{E}_k | \mathcal{E}_{k-1}^C, \mathcal{E}_{k-2}^C, \ldots, \mathcal{E}_1^C) = \Omega\left(\frac{1}{k(\log k)^p}\right).$$

This implies that these terms are not summable, i.e., $\sum_{k=1}^{\infty} \mathbb{P}_0(\mathcal{E}_k | \mathcal{E}_{k-1}^C, \mathcal{E}_{k-2}^C, \ldots, \mathcal{E}_1^C) = \infty$. Therefore we have $\mathbb{P}_0(\limsup_{k\to\infty} \mathcal{E}_k) = 1$, which means that with probability 1, $d_k = 1$ occurs for infinitely many k. Consequentially, we have $\mathbb{P}_0(\mathcal{H}) = 0$. By symmetry, $\mathbb{P}_1(\mathcal{H}) = 0$. This concludes the proof. □

Suppose that the flipping probability converges to $1/2$ sufficiently fast. Then, even if the public belief converges to 0, its convergence rate is very small because the broadcasted decisions become uninformative in a fast rate. In this case, the private signals are capable to overcome the public belief infinitely often because of the slow convergence rate of the public belief.

2.3.2.3 Polynomial tail density

We now consider the case where the private belief has polynomial tail densities, that is, $\rho(r) \to 0$ as $r \to 1$ and there exist constants $\beta, \gamma > 0$ such that

$$\lim_{r\to 1} \frac{\rho(r)}{(1-r)^{\beta}} = \gamma. \tag{2.19}$$

Note that β denotes the leading exponent of the Taylor expansion of the density at 1. The larger the value of β, the thiner the tail density. Note that Theorem 2.3.4 (necessary condition for $\mathbb{P}(\mathcal{L}) > 0$) which was stated under the constant density assumption is also valid in the polynomial tail density case. We can use the similar analysis as before to derive the explicit relationship between the convergence rate of Q_k and the convergence rate of the public belief conditioned on $\mathcal{L}$. The following theorem establishes the scaling laws of the public belief and Type I error probability for both uniformly informative and asymptotic uninformative cases.

Theorem 2.3.7. *Consider the polynomial tail density defined in* (2.19).

1) *Uniformly informative case: Suppose that the flipping probabilities are bounded away from $1/2$. Then, we have $b_k = \Omega(k^{-1/(\beta+1)})$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega(k^{-(\beta+2)/(\beta+1)})$.*

2) Asymptotically uninformative case: Suppose that the flipping probabilities converge to $1/2$, i.e., $Q_k \to 0$. Conditioned on $\mathcal{L}$, we have

(i) if $Q_k = \Theta(1/k^{1-p})$ where $p \in (0,1)$, then $b_k = \Omega(k^{-p/(\beta+1)})$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega(k^{-(\beta+2)p/(\beta+1)})$,

(ii) if $Q_k = \Theta(1/k)$, then $b_k = \Omega((\log k)^{-1/(\beta+1)})$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega((\log k)^{-(\beta+2)/(\beta+1)})$,

(iii) if $Q_k = \Theta\left(1/(k(\log k)^p)\right)$ where $p \in (0,1)$, then $b_k = \Omega((\log k)^{-q/(\beta+1)})$ almost surely, where $1/q + 1/p = 1$, and $\mathbb{P}_0(d_k = 1) = \Omega((\log k)^{-(\beta+2)q/(\beta+1)})$,

(iv) if $Q_k = \Theta\left(1/(k \log k)\right)$, then $b_k = \Omega((\log\log k)^{-1/(\beta+1)})$ almost surely and $\mathbb{P}_0(d_k = 1) = \Omega((\log\log k)^{-(\beta+2)/(\beta+1)})$.

Proof. Proof of claim 1: If the flipping probabilities are bounded away from $1/2$, then the public belief b_k converges to 0 and conditioned on $\mathcal{H}$ we have

$$\begin{aligned}\mathbb{P}_1(d_{k+1} = 0|\hat{D}_k) &= 1 - \int_{1-b_k}^{1} f^1(x)dx \\ &\simeq 1 - \frac{\gamma}{\beta} b_k^{\beta+1}\end{aligned} \tag{2.20}$$

and

$$\begin{aligned}\mathbb{P}_0(d_{k+1} = 0|\hat{D}_k) &= 1 - \int_{1-b_k}^{1} f^0(x)dx \\ &\simeq 1 - \frac{\gamma}{\beta+1} b_k^{\beta+2}.\end{aligned} \tag{2.21}$$

We can also calculate the (conditional) Type I error probability in this case:

$$\begin{aligned}\mathbb{P}_0(d_{k+1} = 1|\hat{D}_k) &= 1 - \mathbb{P}_0(d_{k+1} = 1|\hat{D}_k) \\ &= \int_{1-b_k}^{1} f^0(x)dx \\ &\simeq \frac{\gamma}{\beta+1} b_k^{\beta+2}.\end{aligned} \tag{2.22}$$

Note that (2.22) describes the relationship between the decay rate of Type I error probability and the decay rate of b_k. Next we derive the decay rate of b_k.

By (2.20) and (2.21), we can derive the recursion for the public belief as follows:

$$b_{k+1} = b_k - \frac{\gamma}{\beta} Q_k b_k^{\beta+2}. \tag{2.23}$$

By Lemma 2.3.3, we know that $b_k \to 0$ and the decay rate is $b_k = \Theta(k^{-1/(\beta+1)})$. Recall that conditioned on the event that $\hat{d}_k = 0$ for all k, the convergence of b_k is the fastest. Therefore, we have $b_k = \Omega(k^{-1/(\beta+1)})$ almost surely. From (2.22) and invoking Jensen's Inequality, we obtain

$$\begin{aligned} \mathbb{P}_0(d_k = 1) &\geq \frac{\gamma}{\beta+1} \mathbb{E}_0[b_k^{\beta+2}] \\ &\geq \frac{\gamma}{\beta+1} (\mathbb{E}_0[b_k])^{\beta+2}. \end{aligned} \tag{2.24}$$

Because $b_k = \Omega(k^{-1/(\beta+1)})$ almost surely, we have $\mathbb{P}_0(d_k = H_1) = \Omega(k^{-(\beta+2)/(\beta+1)})$.

Proof of claim 2: Using Lemma 2.3.3, we can show that there exist two positive constants C_1 and C_2 such that

$$\frac{C_1}{(kQ_k)^{1/(\beta+1)}} \leq b_k \leq \frac{C_2}{(kQ_k)^{1/(\beta+1)}}. \tag{2.25}$$

Therefore, if $Q_k = 1/k^{1-p}$, then using (2.25) and the fact that b_k given $\mathcal{H}$ is the smallest among all possible outcomes, we have $b_k = \Omega(k^{-p/(\beta+1)})$. This establishes (i). For (ii)-(iv), we can solve the ODEs given by (2.23) and the solutions give rise to the convergence rates for b_k, which in turn characterize the convergence rates of the error probabilities.

□

Next we provide a necessary condition such that $\mathcal{H}$ has nonzero probability.

Theorem 2.3.8. *Suppose that there exists* $p \leq 1$ *such that*

$$Q_k = O\left(\frac{(p + \log k)(\log k)^{p-1}}{(k(\log k)^p)^{1/(\beta+2)}} \right).$$

Then, we have $\mathbb{P}(\mathcal{H}) = 0$.

Proof. The proof is similar to that of Theorem 2.3.6 and is omitted. $\square$

Note that as β gets larger, this necessary condition states that Q_k has to decay very slowly in order that it is possible for $\mathcal{H}$ to occur.

Similarly we can calculate the decay rate for the Type II error probability $\mathbb{P}_1(d_k = 0)$. Assume that the tail density is given by

$$\lim_{r \to 0} \frac{\rho(r)}{r^{\bar{\beta}}} = \bar{\gamma}$$

where $\bar{\beta}, \bar{\gamma} > 0$. Then, we can show that if the flipping probabilities are bounded away from $1/2$, then

$$\mathbb{P}_1(d_k = 0) = \Omega(k^{-(\bar{\beta}+2)/(\bar{\beta}+1)}).$$

The decay rate of the error probability is given by

$$\mathbb{P}_e^k = \Omega\left(k^{-(1+1/(\max(\beta,\bar{\beta})+1))}\right).$$

CHAPTER 3

BALANCED BINARY RELAY TREES

In this chapter, we study the binary hypothesis testing problem in the context of balanced binary relay trees. We derive explicit upper and lower bounds for the total error probability at the fusion center as functions of the number of leaf nodes. These characterize the decay rate of the total error probability in the asymptotic regime. We also show that the total error probability converges to 0 even if the sensors are asymptotically crummy.

3.1 Problem Formulation

We consider the problem of binary hypothesis testing between H_0 and H_1 in a balanced binary relay tree, as shown in Fig. 3.1. Leaf nodes are sensors undertaking initial and independent detections of the same event in a scene. These measurements are summarized into binary messages and forwarded to nodes at the next level. Each nonleaf node with the exception of the root, the fusion center, is a relay node, which fuses two binary messages into one new binary message and forwards the new binary message to its parent node. This process takes place at each node culminating in the fusion center, at which the final decision is made based on the information received. Only the leaves are sensors in this tree architecture.

In this configuration, the closest sensor to the fusion center is as far as it could be, in terms of the number of arcs in the path to the root. In this sense, this configuration is the worst case among all relay trees with N sensors. Moreover, in contrast to the configuration in [52] and [54] discussed earlier, in our balanced binary tree we have $\lim_{\tau_N \to \infty} \ell_N / \tau_N = 1/2$ (as opposed to 1 in [52] and [54]). Hence, the number of times that information is aggregated is essentially as large as the number of measurements (cf., [52] and [54], in which the number of measurements dominates the number of fusions). In addition, the height of the tree is $\log N$, which grows as the number of sensors increases. (Throughout this thesis, $\log$ stands for the binary logarithm unless specified.)

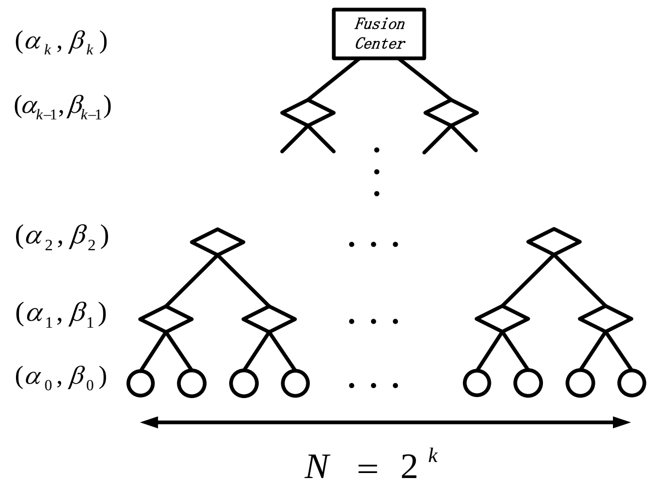


Figure 3.1: A balanced binary relay tree with height k. Circles represent sensors making measurements. Diamonds represent relay nodes which fuse binary messages. The rectangle at the root represents the fusion center making an overall decision.

We assume that all sensors are independent given each hypothesis, and that all sensors have identical Type I error probability α_0 and identical Type II error probability β_0. We apply the likelihood ratio test [69] with threshold 1 as the fusion rule at the relay nodes and at the fusion center. This fusion rule is locally (but not necessarily globally[5]) optimal in the case of equally likely hypotheses H_0 and H_1; i.e., it minimizes the total error probability locally at each fusion node. In the case where the hypotheses are not equally likely, the locally optimal fusion rule has a different threshold value, which is the ratio of the two hypothesis probabilities. However, this complicates the analysis without bringing any additional insights. Therefore, for simplicity, we henceforth assume a threshold value of 1 in our analysis. We are interested in following questions:

- What are these Type I and Type II error probabilities as functions of N?
- Will they converge to 0 at the fusion center?

---

[5]We will discuss the global optimality in Section 6.5 of Chapter 6.

- If yes, how fast will they converge with respect to N?

Fusion at a single node receiving information from the two immediate child nodes where these have identical Type I error probabilities α and identical Type II error probabilities β provides a detection with Type I and Type II error probabilities denoted by (α', β'), and given by [60]:

$$(\alpha', \beta') = f(\alpha, \beta) := \begin{cases} (1-(1-\alpha)^2, \beta^2), & \alpha \leq \beta, \\ (\alpha^2, 1-(1-\beta)^2), & \alpha > \beta. \end{cases} \tag{3.1}$$

Evidently, as all sensors have the same error probability pair (α_0, β_0), all relay nodes at level 1 will have the same error probability pair $(\alpha_1, \beta_1) = f(\alpha_0, \beta_0)$, and by recursion,

$$(\alpha_{k+1}, \beta_{k+1}) = f(\alpha_k, \beta_k), \qquad k = 0, 1, \ldots \log N - 1, \tag{3.2}$$

where (α_k, β_k) is the error probability pair of nodes at the kth level of the tree.

The recursive relation (3.2) allows us to consider the pair of the Type I and II error probabilities as a discrete dynamic system. In [60], which focuses on the convergence issues for the total error probability, convergence was proved using Lyapunov methods. The analysis of the precise evolution of the sequence $\{(\alpha_k, \beta_k)\}$ and the total error probability decay rate remains open. In this chapter, we will establish upper and lower bounds for the total error probability and deduce the precise decay rate of the total error probability.

To illustrate the ideas, consider first a single trajectory for the dynamic system given by (3.1), and starting at the initial state (α_0, β_0). This trajectory is shown in Fig. 3.2. It exhibits different behaviors depending on its distance from the $\beta = \alpha$ line. The trajectory approaches $\beta = \alpha$ very fast initially, but when (α_k, β_k) approaches within a certain neighborhood of the line $\beta = \alpha$, the next pair $(\alpha_{k+1}, \beta_{k+1})$ will appear on the other side of that line. In the next section, we will establish theorems that characterize the precise step-by-step behavior of the dynamic system (3.2). In Section 3.3, we derive upper and lower bounds for (twice) the total error probability P_N at the fusion center as functions of N. These bounds show that the convergence of the total error probability is sub-exponential. Specifically, the exponent of P_N is essentially $\sqrt{N}$ (cf., [52], [54], and [55], where the convergence of the total error probability is exponential in trees with bounded height; more precisely, under the Neyman-Pearson criterion, the optimal error exponent is the same as that of the parallel configuration if leaf nodes dominate; i.e., $\lim_{\tau_N \to \infty} \ell_N / \tau_N = 1$; but under the Bayesian criterion it is worse).

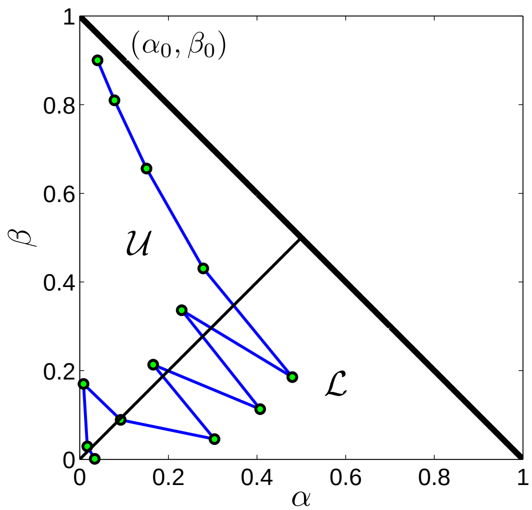


Figure 3.2: A trajectory of the sequence $\{(\alpha_k, \beta_k)\}$ in the (α, β) plane.

3.2 Evolution of Type I and II Error Probabilities

The relation (3.1) is symmetric about both of the lines $\alpha + \beta = 1$ and $\beta = \alpha$. Therefore, it suffices to study the evolution of the dynamic system $\{(\alpha_k, \beta_k)\}$ only in the region bounded by $\alpha + \beta < 1$ and $\beta \geq \alpha$. We denote

$$\mathcal{U} := \{(\alpha, \beta) \geq 0 | \alpha + \beta < 1 \text{ and } \beta \geq \alpha\}$$

to be this triangular region. Similarly, define the complementary triangular region

$$\mathcal{L} := \{(\alpha, \beta) \geq 0 | \alpha + \beta < 1 \text{ and } \beta < \alpha\}.$$

We denote the following region by B_1:

$$B_1 := \{(\alpha, \beta) \in \mathcal{U} | (1 - \alpha)^2 + \beta^2 \leq 1\}.$$

If $(\alpha_k, \beta_k) \in B_1$, then the next pair $(\alpha_{k+1}, \beta_{k+1}) = f(\alpha_k, \beta_k)$ crosses the line $\beta = \alpha$ to the opposite side from (α_k, β_k). More precisely, if $(\alpha_k, \beta_k) \in \mathcal{U}$, then $(\alpha_k, \beta_k) \in B_1$ if and only if $(\alpha_{k+1}, \beta_{k+1}) = f(\alpha_k, \beta_k) \in \mathcal{L}$. In other words, B_1 is the *inverse image* of $\mathcal{L}$ under mapping f in $\mathcal{U}$. The set B_1 is shown in Fig. 3.3(a). Fig. 3.3(b) illustrates this behavior of the trajectory for the example in Fig. 3.2. For instance, as shown in Fig. 3.3(b), if the state is at point 1 in B_1, then it jumps to the next state point 2, on the other side of $\beta = \alpha$.

Denote the following region by B_2:

$$B_2 := \{(\alpha, \beta) \in \mathcal{U} | (1-\alpha)^2 + \beta^2 \geq 1 \text{ and } (1-\alpha)^4 + \beta^4 \leq 1\}.$$

It is easy to show that if $(\alpha_k, \beta_k) \in \mathcal{U}$, then $(\alpha_k, \beta_k) \in B_2$ if and only if $(\alpha_{k+1}, \beta_{k+1}) = f(\alpha_k, \beta_k) \in B_1$. In other words, B_2 is the inverse image of B_1 in $\mathcal{U}$ under mapping f. The behavior of f is illustrated in the movement from point 0 to point 1 in Fig. 3.3(b). The set B_2 is identified in Fig. 3.3(a), lying directly above B_1.

Now for an integer $m > 1$, recursively define B_m to be the inverse image of B_{m-1} under mapping f, denoted by B_m. It is easy to see that

$$B_m := \{(\alpha, \beta) \in \mathcal{U} | (1-\alpha)^{2^{(m-1)}} + \beta^{2^{(m-1)}} \geq 1 \text{ and } (1-\alpha)^{2^m} + \beta^{2^m} \leq 1\}.$$

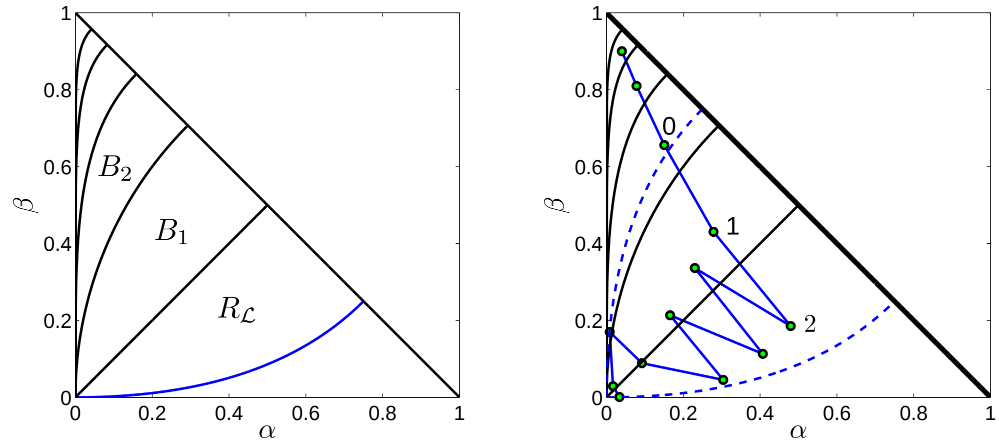
Notice that $\mathcal{U} = \bigcup_{m=1}^{\infty} B_m$. Hence, for any $(\alpha_0, \beta_0) \in \mathcal{U}$, there exists m such that $(\alpha_0, \beta_0) \in B_m$. This gives a complete description of how the dynamics of the system behaves in the upper triangular region $\mathcal{U}$. For instance, if the initial pair (α_0, β_0) lies in B_m, then the system evolves in the order

$$B_m \to B_{m-1} \to \ldots \to B_2 \to B_1.$$

Therefore, the system enters B_1 after $m-1$ levels of fusion; i.e., $(\alpha_{m-1}, \beta_{m-1}) \in B_1$.

As the next stage, we consider the behavior of the system after it enters B_1. The *image* of B_1 under mapping f, denoted by $R_{\mathcal{L}}$, is (see Fig. 3.3(a))

$$R_{\mathcal{L}} := \{(\alpha, \beta) \in \mathcal{L} | \sqrt{1-\alpha} + \sqrt{\beta} \geq 1\}.$$



(a) (b)

Figure 3.3: (a) Regions B_1, B_2, and $R_{\mathcal{L}}$ in the (α, β) plane. (b) The trajectory in Fig. 3.2 superimposed on (a), where solid lines represent boundaries of B_m and dashed lines represent boundaries of R.

We can define the reflection of B_m about the line $\beta = \alpha$ in the similar way for all m. Similarly, we denote by $R_{\mathcal{U}}$ the reflection of $R_{\mathcal{L}}$ about the line $\beta = \alpha$; i.e.,

$$R_{\mathcal{U}} := \{(\alpha, \beta) \in \mathcal{U} | \sqrt{1-\beta} + \sqrt{\alpha} \geq 1\}.$$

We denote the region $R_{\mathcal{U}} \cup R_{\mathcal{L}}$ by R. We will show that R is an *invariant region* in the sense that once the dynamic system enters R, it stays there. For example, as shown in Fig. 3.3(b), the system after point 1 stays inside R.

Proposition 3.2.1. *If* $(\alpha_{k_0}, \beta_{k_0}) \in R$ *for some* k_0*, then* $(\alpha_k, \beta_k) \in R$ *for all* $k \geq k_0$.

Proof. First we show that $B_1 \subset R_{\mathcal{U}} \subset B_1 \cup B_2$.

Notice that B_1, $R_{\mathcal{U}}$, and $B_1 \cup B_2$ share the same lower boundary $\beta = \alpha$. It suffices to show that the upper boundary of $R_{\mathcal{U}}$ lies between the upper boundary of B_2 and that of B_1 (see Fig. 3.4).

First, we show that the upper boundary of $R_{\mathcal{U}}$ lies above the upper boundary of B_1. We have

$$
\begin{aligned}
& 1-(1-\sqrt{\alpha})^2 \geq \sqrt{1-(1-\alpha)^2} \\
\iff \quad & 2\sqrt{\alpha}-\alpha \geq \sqrt{2\alpha-\alpha^2} \\
\iff \quad & \alpha^2+\alpha-2\alpha^{3/2} \geq 0,
\end{aligned}
$$

which holds for all α in $[0, 1)$. Thus, $B_1 \subset R_{\mathcal{U}}$.

Now we prove that the upper boundary of $R_{\mathcal{U}}$ lies below that of B_2. We have

$$
\begin{aligned}
& (1-(1-\alpha)^4)^{1/4} \geq 1-(1-\sqrt{\alpha})^2 \\
\iff \quad & 1-(1-\alpha)^4 \geq (2\sqrt{\alpha}-\alpha)^4 \\
\iff \quad & -2(\sqrt{\alpha}-1)^2\alpha(-\alpha^{3/2}+\alpha(\sqrt{\alpha}-1)+4\sqrt{\alpha}(\sqrt{\alpha}-1)+\alpha-2) \geq 0,
\end{aligned}
$$

which holds for all α in $[0, 1)$ as well. Hence, $R_{\mathcal{U}} \subset B_1 \cup B_2$.

Without loss of generality, we assume that $(\alpha_{k_0}, \beta_{k_0}) \in R_{\mathcal{U}}$. It means that either $(\alpha_{k_0}, \beta_{k_0}) \in B_1$ or $(\alpha_{k_0}, \beta_{k_0}) \in B_2 \cap R_{\mathcal{U}}$. If $(\alpha_{k_0}, \beta_{k_0}) \in B_1$, then the next pair $(\alpha_{k_0+1}, \beta_{k_0+1})$ lies in $R_{\mathcal{L}}$. If $(\alpha_{k_0}, \beta_{k_0}) \in B_2 \cap R_{\mathcal{U}}$, then $(\alpha_{k_0+1}, \beta_{k_0+1}) \in B_1 \subset R_{\mathcal{U}}$ and $(\alpha_{k_0+2}, \beta_{k_0+2}) \in R_{\mathcal{L}}$. By symmetry considerations, it follows that the system stays inside R for all $k \geq k_0$. $\square$

So far we have studied the precise evolution of the sequence $\{(\alpha_k, \beta_k)\}$ in the (α, β) plane. In the next section, we will consider the step-wise reduction in the total error probability and deduce upper and lower bounds for it.

3.3 Error Probability Bounds

In this section, we will first derive bounds for the total error probability in the case of equally likely hypotheses, where the fusion rule is the likelihood ratio test with unit threshold. Then we will deduce bounds for the total error probability in the case where the prior probabilities are unequal but the fusion rule remains the same.

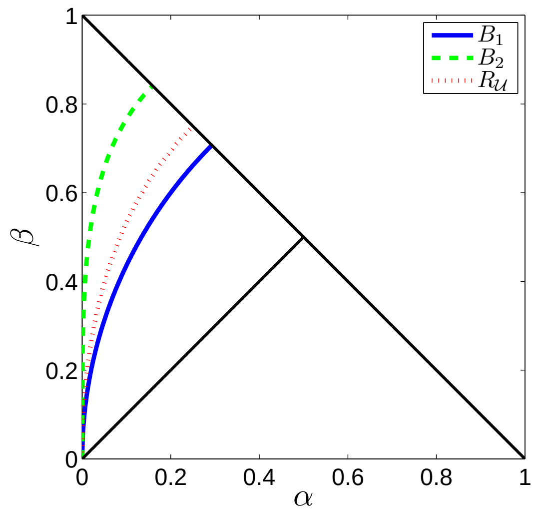


Figure 3.4: Upper boundaries of $B_1, B_2,$ and $R_{\mathcal{U}}$.

The total error probability for a node with (α_k, β_k) is $(\alpha_k + \beta_k)/2$ in the case of equal prior probabilities. Let $L_k = \alpha_k + \beta_k$, namely, twice the total error probability. Analysis of the total error probability results from consideration of the sequence $\{L_k\}$. In fact, we will derive bounds on $\log L_k^{-1}$, whose growth rate is related to the rate of convergence of L_k to 0. We divide our analysis into two parts:

I We study the shrinkage of the total error probability as the system propagates from B_m to B_1;

II We study the shrinkage of the total error probability after the system enters B_1.

3.3.1 Case I: Analysis as the System Propagates from B_m to B_1

Suppose that the initial state (α_0, β_0) lies in B_m, where m is a positive integer and $m \neq 1$. From the previous analysis, $(\alpha_{m-1}, \beta_{m-1}) \in B_1$. In this section, we study the rate of reduction of the total error probability as the system propagates from B_m to B_1.

Proposition 3.3.1. *Suppose that $(\alpha_k, \beta_k) \in B_m$, where m is a positive integer and $m \neq 1$. Then,*

$$1 \leq \frac{L_{k+1}}{L_k^2} \leq 2.$$

Proof. If $(\alpha_k, \beta_k) \in B_m$, where m is a positive integer and $m \neq 1$, then

$$\frac{L_{k+1}}{L_k^2} = \frac{1 - (1-\alpha_k)^2 + \beta_k^2}{(\alpha_k + \beta_k)^2}.$$

The following calculation establishes the lower bound of the ratio L_{k+1}/L_k^2:

$$\begin{aligned} L_{k+1} - L_k^2 &= 1 - (1-\alpha_k)^2 + \beta_k^2 - (\alpha_k + \beta_k)^2 \\ &= -2\alpha_k^2 - 2\alpha_k\beta_k + 2\alpha_k \\ &= 2\alpha_k(1 - (\alpha_k + \beta_k)) \geq 0, \end{aligned}$$

which holds in B_m.

To show the upper bound of the ratio L_{k+1}/L_k^2, it suffices to prove that

$$\begin{aligned} L_{k+1} - 2L_k^2 &= 1 - (1-\alpha_k)^2 + \beta_k^2 - 2(\alpha_k + \beta_k)^2 \\ &= -3\alpha_k^2 - 4\alpha_k\beta_k + 2\alpha_k - \beta_k^2 \leq 0. \end{aligned}$$

The partial derivative with respect to β_k is

$$\frac{\partial(L_{k+1} - 2L_k^2)}{\partial \beta_k} = -2\beta_k - 4\alpha_k \leq 0,$$

which is non-positive, and so it suffices to consider values on the upper boundary of B_1.

$$\begin{aligned} L_{k+1} - 2L_k^2 &= 1 - (1-\alpha_k)^2 + \beta_k^2 - 2(\alpha_k + \beta_k)^2 \\ &= 2\beta_k^2 - 2(\alpha_k + \beta_k)^2 \leq 0. \end{aligned}$$

In consequence, the claimed upper bound on the ratio L_{k+1}/L_k^2 holds. □

Fig. 3.5 shows a plot of values of L_{k+1}/L_k^2 in $\bigcup_{m=2}^{\infty} B_m$. With the recursive relation given in Proposition 3.3.1, we can derive the following bounds for $\log L_k^{-1}$.

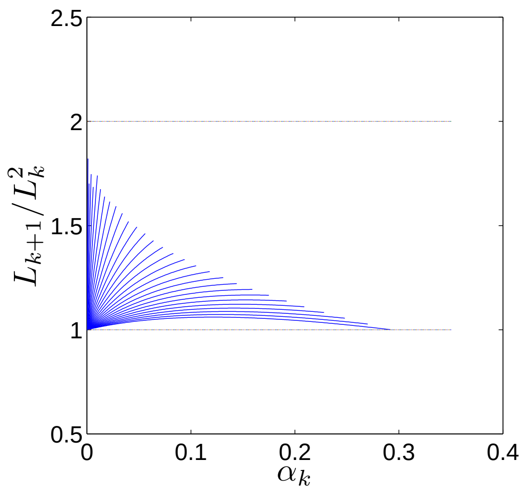


Figure 3.5: Ratio L_{k+1}/L_k^2 in $\bigcup_{m=2}^{\infty} B_m$. Each line depicts the ratio versus α_k for a fixed β_k.

Proposition 3.3.2. *Suppose that* $(\alpha_0, \beta_0) \in B_m$*, where* m *is a positive integer and* $m \neq 1$*. Then, for* $k = 1, 2, \ldots, m-1$*,*

$$2^k \left(\log L_0^{-1} - 1\right) \leq \log L_k^{-1} \leq 2^k \log L_0^{-1}.$$

Proof. From Proposition 3.3.1 we have, for $k = 0, 1, \ldots, m-2$,

$$L_{k+1} = a_k L_k^2$$

for some $a_k \in [1, 2]$. Then for $k = 1, 2, \ldots, m-1$,

$$L_k = a_{k-1} \cdot a_{k-2}^2 \ldots a_0^{2^{k-1}} L_0^{2^k},$$

where $a_i \in [1, 2]$ for each i. Hence,

$$\log L_k^{-1} = -\log a_{k-1} - 2\log a_{k-2} - \ldots - 2^{k-1}\log a_0 - \log L_0^{2^k}.$$

Since $\log L_0^{-1} > 0$ and $0 \leq \log a_i \leq 1$ for each i, we have

$$\log L_k^{-1} \leq 2^k \log L_0^{-1}.$$

Finally,

$$\begin{aligned}\log L_k^{-1} &\geq -1 - 2 - \ldots - 2^{k-1} + 2^k \log L_0^{-1} \\ &\geq -2^k + 2^k \log L_0^{-1} = 2^k \left(\log L_0^{-1} - 1\right).\end{aligned}$$

□

Suppose that the balanced binary relay tree has N leaf nodes. Then, the height of the fusion center is $\log N$. For convenience, let $P_N = L_{\log N}$ be (twice) the total error probability at the fusion center. Substituting $k = \log N$ into Proposition 3.3.2, we get the following result.

Corollary 3.3.1. *Suppose that $(\alpha_0, \beta_0) \in B_m$, where m is a positive integer and $m \neq 1$. If $\log N < m$, then*

$$N\left(\log L_0^{-1} - 1\right) \leq \log P_N^{-1} \leq N \log L_0^{-1}.$$

Notice that the lower bound of $\log P_N^{-1}$ is useful only if $L_0 < 1/2$. Next we derive a lower bound for $\log P_N^{-1}$ which is useful for all $L_0 \in (0, 1)$.

Proposition 3.3.3. *Suppose that $(\alpha_k, \beta_k) \in B_m$, where m is a positive integer and $m \neq 1$. Then,*

$$\frac{L_{k+1}}{L_k^{\sqrt{2}}} \leq 1.$$

Proof. If $(\alpha_k, \beta_k) \in B_m$, where m is a positive integer and $m \neq 1$, then

$$\frac{L_{k+1}}{L_k^{\sqrt{2}}} = \frac{1 - (1-\alpha_k)^2 + \beta_k^2}{(\alpha_k + \beta_k)^{\sqrt{2}}}.$$

To prove the upper bound of the ratio, it suffices to show that

$$\psi(\alpha_k, \beta_k) = 1 - (1-\alpha_k)^2 + \beta_k^2 - (\alpha_k + \beta_k)^{\sqrt{2}} \leq 0.$$

The second-order partial derivative of ψ with respect to α_k is non-positive:

$$\frac{\partial^2 \psi}{\partial \alpha_k^2} = -2 - \sqrt{2}(\sqrt{2}-1)(\alpha_k + \beta_k)^{\sqrt{2}-2} \le 0.$$

Therefore, the minimum of $\partial\psi/\partial\alpha_k$ is on the lines $\alpha_k + \beta_k = 1$ and $(1-\alpha_k)^2 + \beta_k^2 = 1$. It is easy to show that $\partial\psi/\partial\alpha_k \ge 0$. In consequence, the maximum of ψ is on the lines $\alpha_k + \beta_k = 1$ and $(1-\alpha_k)^2 + \beta_k^2 = 1$. If $\alpha_k + \beta_k = 1$, then it is easy to see that $\psi = 0$. If $(1-\alpha_k)^2 + \beta_k^2 = 1$, then $\psi = 2\beta_k^2 - (\alpha_k + \beta_k)^{\sqrt{2}}$. It is easy to show that the maximum value of ψ lies at the intersection of $\alpha_k + \beta_k = 1$ and $(1-\alpha_k)^2 + \beta_k^2 = 1$, where $\psi = 0$. Hence, the ratio $L_{k+1}/L_k^{\sqrt{2}}$ is upper bounded by 1. □

Fig. 3.6 shows a plot of values of $L_{k+1}/L_k^{\sqrt{2}}$ in $\bigcup_{m=2}^{\infty} B_m$. With the inequality given in Proposition 3.3.3, we can derive a new lower bound for $\log P_N^{-1}$, which is useful for all $L_0 \in (0,1)$.

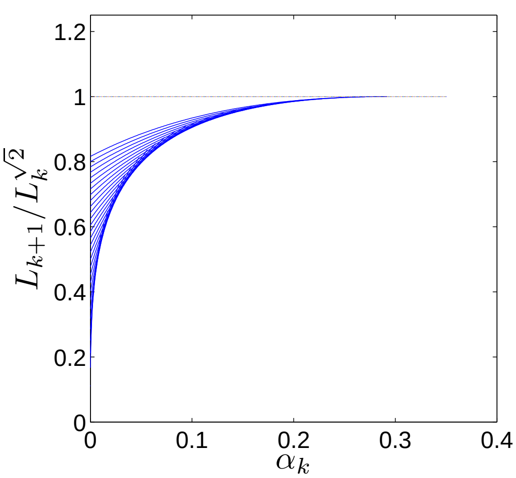


Figure 3.6: Ratio $L_{k+2}/L_k^{\sqrt{2}}$ in $\bigcup_{m=2}^{\infty} B_m$. Each line depicts the ratio versus α_k for a fixed β_k.

Proposition 3.3.4. *Suppose that $(\alpha_0, \beta_0) \in B_m$, where m is a positive integer and $m \neq 1$. If $\log N < m$, then*

$$\log P_N^{-1} \ge \sqrt{N} \log L_0^{-1}.$$

Proof. From Proposition 3.3.3 we have, for $k = 0, 1, \ldots, m-2$,

$$L_{k+1} = a_k L_k^{\sqrt{2}}$$

for some $a_k \in (0, 1]$. Then for $k = 1, 2, \ldots, m-1$,

$$L_k = a_{k-1} \cdot a_{k-2}^{\sqrt{2}} \ldots a_0^{\sqrt{2}^{k-1}} L_0^{\sqrt{2}^k},$$

where $a_i \in (0, 1]$ for each i. Hence,

$$\log L_k^{-1} = -\log a_{k-1} - \sqrt{2}\log a_{k-2} - \ldots - \sqrt{2}^{k-1}\log a_0 - \log L_0^{\sqrt{2}^k}.$$

Since $\log L_0^{-1} > 0$ and $\log a_i \leq 0$ for each i, we have

$$\log L_k^{-1} \geq \sqrt{2}^k \log L_0^{-1}.$$

Therefore, we have

$$\log P_N^{-1} \geq \sqrt{N} \log L_0^{-1}.$$

□

3.3.2 Case II: Analysis when the System Stays inside R

We have derived error probability bounds up until the point where the trajectory of the system enters B_1. In this section, we consider the total error probability reduction from that point on. First we will establish error probability bounds for even-height trees. Then we will deduce error probability bounds for odd-height trees.

3.3.2.1 Error probability bounds for even-height trees

If $(\alpha_0, \beta_0) \in B_m$ for some $m \neq 1$, then $(\alpha_{m-1}, \beta_{m-1}) \in B_1$. The system afterward stays inside the invariant region R (but not necessarily inside B_1). Hence, the decay rate of the total error probability in the invariant region R determines the asymptotic decay rate. Without loss of generality, we assume that (α_0, β_0) lies in the invariant region R. In contrast to Proposition 3.3.1, which bounds the ratio L_{k+1}/L_k^2, we will bound the ratio L_{k+2}/L_k^2 associated with taking two steps.

Proposition 3.3.5. *Suppose that* $(\alpha_k, \beta_k) \in R$. *Then,*

$$1 \leq \frac{L_{k+2}}{L_k^2} \leq 2.$$

Proof. Because of symmetry, we only have to prove the case where (α_k, β_k) lies in $R_{\mathcal{U}}$. We consider two cases: $(\alpha_k, \beta_k) \in B_1$ and $(\alpha_k, \beta_k) \in B_2 \cap R_{\mathcal{U}}$.

In the first case,

$$\frac{L_{k+2}}{L_k^2} = \frac{(1-(1-\alpha_k)^2)^2 + 1 - (1-\beta_k^2)^2}{(\alpha_k + \beta_k)^2}.$$

To prove the lower bound of the ratio, it suffices to show that

$$\begin{aligned} L_{k+2} - L_k^2 &= (1-(1-\alpha_k)^2)^2 + 1 - (1-\beta_k^2)^2 - (\alpha_k + \beta_k)^2 \\ &= (1-\alpha_k-\beta_k)((\beta_k-\alpha_k)^3 + 2\alpha_k\beta_k(\beta_k-\alpha_k) \\ &+ (\beta_k-\alpha_k)^2 + 2\alpha_k^2) \geq 0. \end{aligned}$$

We have $1-\alpha_k-\beta_k > 0$ and $\beta_k \geq \alpha_k$ for all $(\alpha_k, \beta_k) \in B_1$, resulting in the above inequality.

To prove the upper bound of the ratio, it suffices to show that

$$L_{k+2} - 2L_k^2 = \alpha_k^4 - 4\alpha_k^3 + 2\alpha_k^2 - 4\alpha_k\beta_k - \beta_k^4 \leq 0.$$

The partial derivative with respect to β_k is

$$\frac{\partial(L_{k+2} - 2L_k^2)}{\partial \beta_k} = -4\alpha_k - 4\beta_k^3 \leq 0,$$

which is non-positive. Therefore, it suffices to consider its values on the curve $\beta_k = \alpha_k$, on which $L_{k+2} - 2L_k^2$ is clearly non-positive.

Now we consider the second case, namely $(\alpha_k, \beta_k) \in B_2 \cap R_{\mathcal{U}}$, which gives

$$\frac{L_{k+2}}{L_k^2} = \frac{1-(1-\alpha_k)^4+\beta_k^4}{(\alpha_k+\beta_k)^2}.$$

To prove the lower bound of the ratio, it suffices to show that

$$\begin{aligned} L_{k+2} - L_k^2 &= (1-(1-\alpha_k)^4) + \beta_k^4 - (\alpha_k+\beta_k)^2 \\ &= (1-\alpha_k-\beta_k)(\alpha_k^3 - \alpha_k^2\beta_k - 3\alpha_k^2 + \alpha_k\beta_k^2 \\ &\quad + 2\alpha_k\beta_k - \beta_k^3 - \beta_k^2 + 4\alpha_k) \geq 0. \end{aligned}$$

Therefore, it suffices to show that

$$\phi(\alpha_k, \beta_k) = \alpha_k^3 - \alpha_k^2\beta_k - 3\alpha_k^2 + \alpha_k\beta_k^2 + 2\alpha_k\beta_k - \beta_k^3 - \beta_k^2 + 4\alpha_k \geq 0.$$

The partial derivative with respect to β_k is

$$\frac{\partial\phi}{\partial\beta_k} = -(\alpha_k-\beta_k)^2 - 2\beta_k^2 + 2(\alpha_k-\beta_k) \leq 0.$$

Thus, it is enough to consider the values on the upper boundaries $\sqrt{1-\beta_k} + \sqrt{\alpha_k} = 1$ and $\alpha_k + \beta_k = 1$.

If $\alpha_k + \beta_k = 1$, then the inequality is trivial, and if $\sqrt{1-\beta_k} + \sqrt{\alpha_k} = 1$, then

$$L_{k+2} - L_k^2 = 2\alpha_k^2(1-2\sqrt{\alpha_k})(2\alpha_k - 6\sqrt{\alpha_k} + 5)$$

and the inequality holds because $\alpha_k \leq \frac{1}{4}$ in region $B_2 \cap R_{\mathcal{U}}$.

The claimed upper bound for the ratio L_{k+2}/L_k^2 can be written as

$$\begin{aligned} L_{k+2} - 2L_k^2 &= (1-(1-\alpha_k)^4) + \beta_k^4 - 2(\alpha_k+\beta_k)^2 \\ &= -\alpha_k^4 + 4\alpha_k^3 - 8\alpha_k^2 + 4\alpha_k - 4\alpha_k\beta_k + \beta_k^4 - 2\beta_k^2 \leq 0. \end{aligned}$$

The partial derivative with respect to β_k is

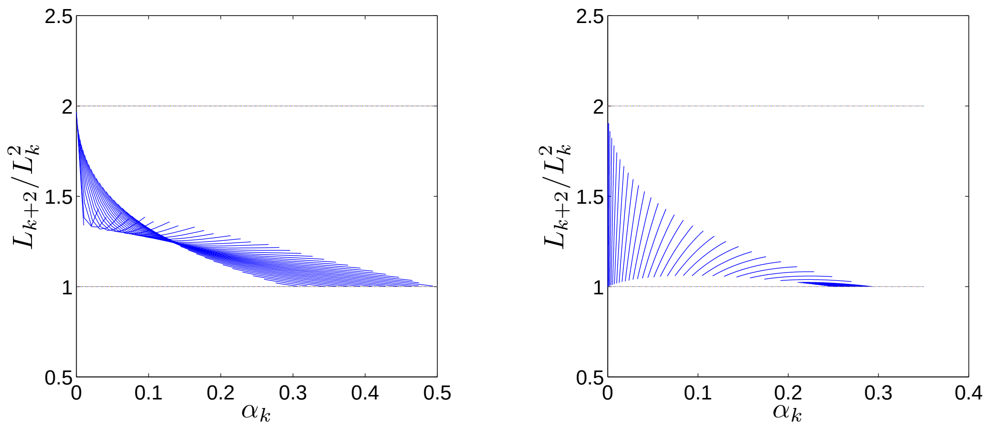
$$\frac{\partial(L_{k+2} - 2L_k^2)}{\partial\beta_k} = -4\alpha_k + 4\beta_k^3 - 4\beta_k \leq 0.$$

Again, it is sufficient to consider values on the upper boundary of B_1. Hence,

$$L_{k+2} - 2L_k^2 = 2\beta_k^2 - 2(\alpha_k + \beta_k)^2 \leq 0.$$

$\square$

Fig. 3.7(a) and Fig. 3.7(b) show plots of values of L_{k+2}/L_k^2 in B_1 and $B_2 \cap R_{\mathcal{U}}$, respectively.



(a) (b)

Figure 3.7: (a) Ratio L_{k+2}/L_k^2 in B_1. (b) Ratio L_{k+2}/L_k^2 in $B_2 \cap R_{\mathcal{U}}$. Each line depicts the ratio versus α_k for a fixed β_k.

Proposition 3.3.5 gives bounds on the relationship between L_k and L_{k+2} in the invariant region R. Hence, in the special case of trees with even height, that is, when $\log N$ is an even integer, it is easy to bound P_N in terms of L_0. In fact, we will bound $\log P_N^{-1}$ which in turn provides bounds for P_N.

Theorem 3.3.1. *Suppose that* $(\alpha_0, \beta_0) \in R$ *and* $\log N$ *is even. Then,*

$$\sqrt{N}\left(\log L_0^{-1} - 1\right) \leq \log P_N^{-1} \leq \sqrt{N} \log L_0^{-1}.$$

Proof. If $(\alpha_0, \beta_0) \in R$, then we have $(\alpha_k, \beta_k) \in R$ for $k = 0, 1, \ldots, \log N - 2$. From Proposition 3.3.5, we have

$$L_{k+2} = a_k L_k^2$$

for $k = 0, 2, \ldots, \log N - 2$ and some $a_k \in [1, 2]$. Therefore, for $k = 2, 4, \ldots, \log N$, we have

$$L_k = a_{(k-2)/2} \cdot a_{(k-4)/2}^2 \ldots a_0^{2^{(k-2)/2}} L_0^{2^{k/2}},$$

where $a_i \in [1, 2]$ for each i. Substituting $k = \log N$, we have

$$\begin{aligned} P_N &= a_{(k-2)/2} \cdot a_{(k-4)/2}^2 \ldots a_0^{2^{(k-2)/2}} L_0^{2^{\log \sqrt{N}}} \\ &= a_{(k-2)/2} \cdot a_{(k-4)/2}^2 \ldots a_0^{\sqrt{N}/2} L_0^{\sqrt{N}}. \end{aligned}$$

Hence,

$$\log P_N^{-1} = -\log a_{(k-2)/2} - 2 \log a_{(k-4)/2} - \ldots - \frac{\sqrt{N}}{2} \log a_0 + \sqrt{N} \log L_0^{-1}.$$

Notice that $\log L_0^{-1} > 0$ and $0 \leq \log a_i \leq 1$ for each i. Thus,

$$\log P_N^{-1} \leq \sqrt{N} \log L_0^{-1}.$$

Finally,

$$\begin{aligned} \log P_N^{-1} &\geq -1 - 2 - \ldots - \frac{\sqrt{N}}{2} + \sqrt{N} \log L_0^{-1} \\ &\geq -\sqrt{N} + \sqrt{N} \log L_0^{-1} = \sqrt{N} \left(\log L_0^{-1} - 1 \right). \end{aligned}$$

$\square$

Notice the lower bound for $\log P_N^{-1}$ in Theorem 3.3.1 is useful only if $L_0 < 1/2$. We further provide a lower bound for $\log P_N^{-1}$ which is useful for all $L_0 \in (0, 1)$.

Proposition 3.3.6. *Suppose that $(\alpha_k, \beta_k) \in R$. Then,*

$$\frac{L_{k+2}}{L_k^{\sqrt{2}}} \leq 1.$$

Proof. In the case where $(\alpha_k, \beta_k) \in B_2 \cap R_{\mathcal{U}}$, from Proposition 3.3.3, we have $L_{k+1} \leq L_k^{\sqrt{2}}$. Moreover, it is easy to show that $L_{k+2} \leq L_{k+1}$. Thus, we have $L_{k+2} \leq L_k^{\sqrt{2}}$.

In the case where $(\alpha_k, \beta_k) \in B_1$, it suffices to prove that

$$\vartheta(\alpha_k, \beta_k) = (1 - (1 - \alpha_k)^2)^2 + 1 - (1 - \beta_k^2)^2 - (\alpha_k + \beta_k)^{\sqrt{2}} \leq 0.$$

We take second-order partial derivative of ϑ with respect to α_k along the lines $\alpha_k + \beta_k = c$ in this region. It is easy to show that the derivative is non-negative:

$$\frac{\partial^2 \vartheta}{\partial \alpha_k^2} = 12((1 - \alpha_k)^2 - \beta_k^2) \geq 0.$$

Therefore, we conclude that the maximum of ϑ lies on the boundaries of this region. If $\alpha_k + \beta_k = 1$, then we have $\vartheta(\alpha_k, \beta_k) = 0$. If $(1 - \alpha_k)^2 + \beta_k^2 = 1$, then we have $\alpha_{k+1} = \beta_{k+1}$. Moreover, if $\alpha_{k+1} = \beta_{k+1}$, then we can show that $L_{k+2} = L_{k+1}$. Hence, it suffices to show that $L_{k+1}/L_k^{\sqrt{2}} \leq 1$ on the line $(1 - \alpha_k)^2 + \beta_k^2 = 1$, which has been proved in Proposition 3.3.3. If $\beta_k = \alpha_k$, then $L_{k+1} = L_k$ and $(\alpha_{k+1}, \beta_{k+1})$ lies on the lower boundary of $R_{\mathcal{L}}$, on which we have $L_{k+2}/L_{k+1}^{\sqrt{2}} \leq 1$. Thus, we have $L_{k+2}/L_k^{\sqrt{2}} \leq 1$. □

Fig. 3.8(a) and Fig. 3.8(b) show plots of the ratio inside B_1 and $B_2 \cap R_{\mathcal{U}}$, respectively. Next we derive a new lower bound for $\log P_N^{-1}$.

Proposition 3.3.7. *Suppose that* $(\alpha_0, \beta_0) \in R$ *and* $\log N$ *is even. Then,*

$$\log P_N^{-1} \geq \sqrt[4]{N} \log L_0^{-1}.$$

Proof. From Proposition 3.3.6 we have, for $k = 0, 2, \ldots, \log N - 2$,

$$L_{k+2} = a_k L_k^{\sqrt{2}}$$

for some $a_k \in (0, 1]$. Then for $k = 2, 4, \ldots, \log N$, we have

$$L_k = a_{(k-2)/2} \cdot a_{(k-4)/2}^{\sqrt{2}} \ldots a_0^{\sqrt{2}^{(k-2)/2}} L_0^{\sqrt{2}^{k/2}},$$

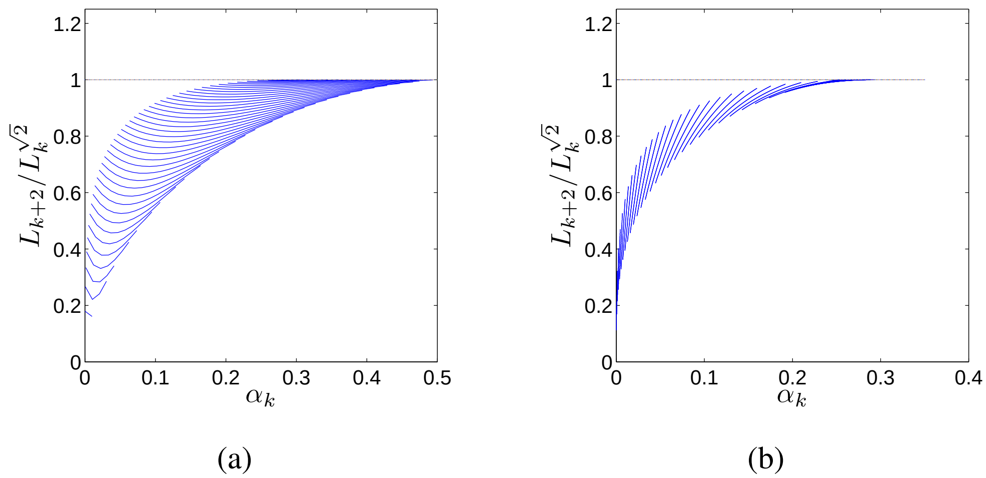


Figure 3.8: (a) Ratio $L_{k+2}/L_k^{\sqrt{2}}$ in B_1. (b) Ratio $L_{k+2}/L_k^{\sqrt{2}}$ in $B_2 \cap R_{\mathcal{U}}$. Each line depicts the ratio versus α_k for a fixed β_k.

where $a_i \in (0, 1]$ for each i. Therefore,

$$\log L_k^{-1} = -\log a_{(k-2)/2} - \sqrt{2}\log a_{(k-4)/2} - \ldots - \sqrt{2}^{(k-2)/2}\log a_0 - \log L_0^{\sqrt{2}^{k/2}}.$$

Since $\log L_0^{-1} > 0$ and $\log a_i \le 0$ for each i, we have

$$\log L_k^{-1} \ge \sqrt{2}^{k/2}\log L_0^{-1}.$$

Therefore, we have

$$\log P_N^{-1} \ge \sqrt[4]{N}\log L_0^{-1}.$$

□

3.3.2.2 Error probability bounds for odd-height trees

Next we explore the case of trees with odd height; i.e., $\log N$ is an odd integer. Assume that (α_0, β_0) lies in the invariant region R. First, we will establish general bounds for odd-height trees. Then we deduce bounds for the case where there exists $(\alpha_m, \beta_m) \in B_2 \cap R_{\mathcal{U}}$ for some $m \in \{0, 1, \ldots, \log N - 2\}$.

For odd-height trees, we need to know how much the total error probability is reduced by moving up one level in the tree.

Proposition 3.3.8. *Suppose that* $(\alpha_k, \beta_k) \in \mathcal{U}$. *Then,*

$$1 \leq \frac{L_{k+1}}{L_k^2} \text{ and } \frac{L_{k+1}}{L_k} \leq 1.$$

Proof. The first inequality is equivalent to

$$\begin{aligned} L_{k+1} - L_k^2 &= 1 - (1-\alpha_k)^2 + \beta_k^2 - (\alpha_k + \beta_k)^2 \\ &= 2\alpha_k(1 - (\alpha_k + \beta_k)) \geq 0, \end{aligned}$$

which holds for all $(\alpha_k, \beta_k) \in \mathcal{U}$.

The second inequality is equivalent to

$$\begin{aligned} L_{k+1} - L_k &= 1 - (1-\alpha_k)^2 + \beta_k^2 - (\alpha_k + \beta_k) \\ &= (\alpha_k - \beta_k)(1 - (\alpha_k + \beta_k)) \leq 0, \end{aligned}$$

which holds for all $(\alpha_k, \beta_k) \in \mathcal{U}$. □

Fig. 3.9(a) and Fig. 3.9(b) show plots of values of L_{k+1}/L_k^2 and L_{k+1}/L_k in $\mathcal{U}$.

Using Propositions 3.3.5 and 3.3.8, we are about to calculate error probability bounds for odd-height trees as follows.

Theorem 3.3.2. *Suppose that* $(\alpha_0, \beta_0) \in R$ *and* $\log N$ *is odd. Then*

$$\sqrt{\frac{N}{2}}\left(\log L_0^{-1} - 1\right) \leq \log P_N^{-1} \leq \sqrt{2N} \log L_0^{-1}.$$

Proof. From Proposition 3.3.8, we have

$$L_1 = \widetilde{a} L_0^2$$

for some $\widetilde{a} \geq 1$. And, by Proposition 3.3.5, the following identity holds.

$$L_{k+2} = a_k L_k^2$$

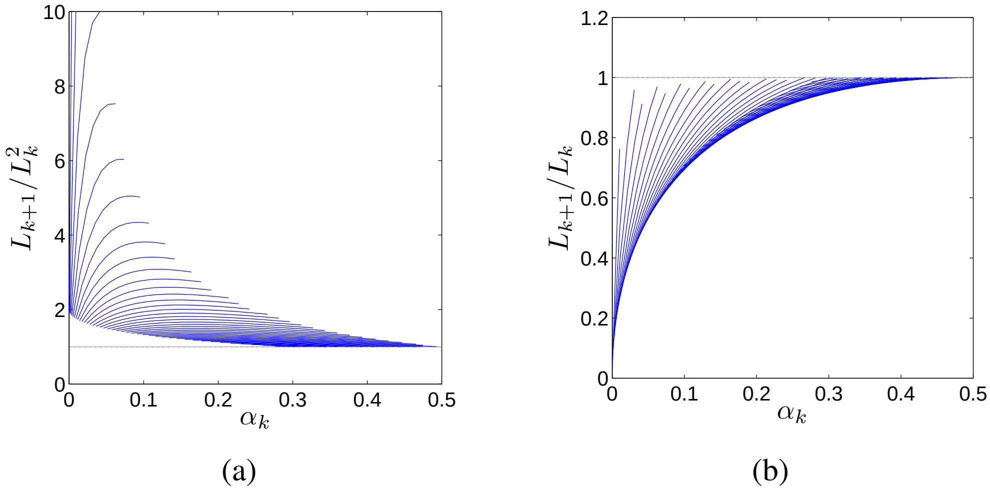


Figure 3.9: (a) Ratio L_{k+1}/L_k^2 in $\mathcal{U}$. (b) Ratio L_{k+1}/L_k in $\mathcal{U}$. Each line depicts the ratio versus α_k for a fixed β_k.

for $k = 1, 3, \ldots, \log N - 2$ and some $a_k \in [1, 2]$. Hence, we can write

$$L_k = \widetilde{a}^{2^{(k-1)/2}} \cdot a_{(k-1)/2} \cdot a_{(k-3)/2}^2 \ldots a_1^{2^{(k-3)/2}} L_0^{2^{(k+1)/2}},$$

where $a_i \in [1, 2]$ for each i and $\widetilde{a} \geq 1$. Let $k = \log N$, we have

$$\log P_N^{-1} = -2^{(k-1)/2} \log \widetilde{a} - \log a_{(k-1)/2} - \ldots - 2^{(k-3)/2} \log a_1 + \sqrt{2N} \log L_0^{-1}.$$

Notice that $\log L_0^{-1} > 0$ and for each i, $\log a_i \geq 0$. Moreover, $\log \widetilde{a} \geq 0$. Hence,

$$\log P_N^{-1} \leq \sqrt{2N} \log L_0^{-1}.$$

It follows by Proposition 3.3.8 that

$$L_k = \widetilde{a} L_{k-1}$$

for some $\widetilde{a} \in (0, 1]$. By Proposition 3.3.5, we have

$$L_{k+2} = a_k L_k^2$$

for $k = 0, 2, \ldots, \log N - 3$ and some $a_k \in [1, 2]$. Thus,

$$L_k = \widetilde{a} \cdot a_{(k-3)/2} \cdot a_{(k-3)/2}^2 \ldots a_0^{2^{(k-3)/2}} L_0^{2^{(k-1)/2}},$$

where $a_i \in [1,2]$ for each i and $\widetilde{a} \in (0,1]$. Hence,

$$\log P_N^{-1} = -\log \widetilde{a} - \log a_{(k-1)/2} - \ldots - 2^{(k-3)/2} \log a_1 + \sqrt{\frac{N}{2}} \log L_0^{-1}.$$

Notice that $\log L_0^{-1} > 0$ and for each i, $0 \leq \log a_i \leq 1$ and $\log \widetilde{a} \leq 0$. Thus,

$$\log P_N^{-1} \geq -\sqrt{\frac{N}{2}} + \sqrt{\frac{N}{2}} \log L_0^{-1} = \sqrt{\frac{N}{2}} \left(\log L_0^{-1} - 1\right).$$

$\square$

Next we consider the special case where there exists $m \in \{0, 1, \ldots, \log N - 2\}$ such that $(\alpha_m, \beta_m) \in B_2 \cap R_{\mathcal{U}}$.

Proposition 3.3.9. *Suppose that* $(\alpha_k, \beta_k) \in B_1$ *and* $(\alpha_{k-1}, \beta_{k-1}) \in B_2 \cap R_{\mathcal{U}}$. *Then,*

$$\frac{1}{2} \leq \frac{L_{k+1}}{L_k} \leq 1.$$

Proof. The upper bound for L_{k+1}/L_k is trivial. By Proposition 3.3.1, if $(\alpha_{k-1}, \beta_{k-1}) \in B_2 \cap R_{\mathcal{U}}$, then

$$1 \leq \frac{L_k}{L_{k-1}^2} \leq 2;$$

i.e.,

$$\frac{1}{2} \leq \frac{L_{k-1}^2}{L_k} \leq 1,$$

and in consequence of Proposition 3.3.5, if $(\alpha_{k-1}, \beta_{k-1}) \in B_2 \cap R_{\mathcal{U}}$, then

$$1 \leq \frac{L_{k+1}}{L_{k-1}^2} \leq 2.$$

Therefore, we have

$$\frac{1}{2} \leq \frac{L_{k+1}}{L_k}.$$

$\square$

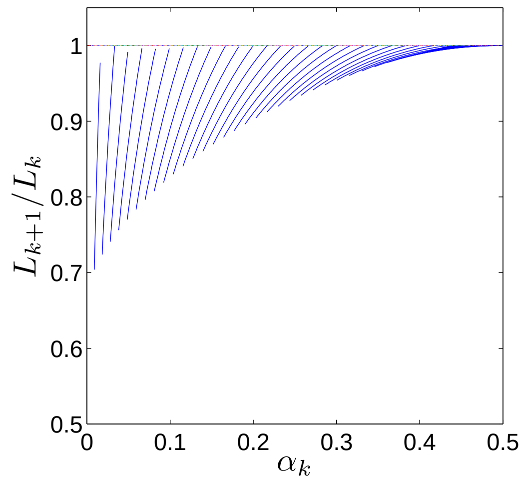


Figure 3.10: Ratio L_{k+1}/L_k in the region $f(B_2 \cup R_{\mathcal{U}})$. Each line depicts the ratio versus α for a fixed β.

Fig. 3.10 shows a plot of values of L_{k+1}/L_k in this case.

We have proved in Proposition 3.3.5 that if (α_k, β_k) is in $B_2 \cap R_{\mathcal{U}}$, then the ratio $L_{k+2}/L_k^2 \in [1, 2]$. However, if we analyze each level of fusion, it can be seen that the total error probability decreases exponentially fast from $B_2 \cap R_{\mathcal{U}}$ to B_1 (Proposition 3.3.1). Proposition 3.3.9 tells us that the fusion from B_1 to $R_{\mathcal{L}}$ is a bad step, which does not contribute significantly in decreasing the total error probability.

We can now provide bounds for the total error probability at the fusion center.

Theorem 3.3.3. *Suppose that* $(\alpha_0, \beta_0) \in R$*,* $\log N$ *is an odd integer, and there exists* $m \in \{0, 1, \ldots, \log N - 2\}$ *such that* $(\alpha_m, \beta_m) \in B_2 \cap R_{\mathcal{U}}$*.*

If m *is even, then*

$$\sqrt{2N}\left(\log L_0^{-1} - 1\right) \leq \log P_N^{-1} \leq \sqrt{2N} \log L_0^{-1}.$$

If m *is odd, then*

$$\sqrt{\frac{N}{2}}\left(\log L_0^{-1} - 1\right) \leq \log P_N^{-1} \leq \sqrt{\frac{N}{2}} \log L_0^{-1} + \sqrt{\frac{N}{2^{m+2}}}.$$

Proof. If $(\alpha_m, \beta_m) \in B_2 \cap R_{\mathcal{U}}$ and m is even, then by Proposition 3.3.1, we have

$$L_{m+1} = \widetilde{a} L_m^2$$

for some $\widetilde{a} \in [1, 2]$.

By Proposition 3.3.5, we have

$$L_{k+2} = a_k L_k^2$$

for $k = 0, 2, \ldots, m-2, m+1, \ldots, \log N - 2$, and some $a_k \in [1, 2]$. Hence,

$$L_k = a_{(k-1)/2} \cdot a_{(k-3)/2}^2 \ldots a_0^{2^{(k-1)/2}} L_0^{2^{(k+1)/2}},$$

where $a_i \in [1, 2]$ for each i.

Let $k = \log N$, we have

$$\log P_N^{-1} = -\log a_{(k-1)/2} - 2 \log a_{(k-3)/2} - \ldots - 2^{(k-1)/2} \log a_0 + \sqrt{2N} \log L_0^{-1}.$$

Notice that $\log L_0^{-1} > 0$ and for each i, $0 \le \log a_i \le 1$. Thus,

$$\log P_N^{-1} \le \sqrt{2N} \log L_0^{-1}.$$

Finally,

$$\log P_N^{-1} \ge -\sqrt{2N} + \sqrt{2N} \log L_0^{-1} = \sqrt{2N} \left(\log L_0^{-1} - 1 \right).$$

If $(\alpha_m, \beta_m) \in B_2 \cap R_{\mathcal{U}}$ and m is odd, then by Proposition 3.3.9 we have

$$L_{m+2} = \widetilde{a} L_{m+1}$$

for some $\widetilde{a} \in [1/2, 1]$.

It follows from Proposition 3.3.5 that

$$L_{k+2} = a_k L_k^2$$

for $k = 0, 2, \dots, m-1, m+2, \dots, \log N - 2$ and some $a_k \in [1, 2]$. Therefore,

$$L_k = a_{(k-3)/2} \cdot a_{(k-3)/2}^2 \dots a_0^{2^{(k-3)/2}} \cdot \widetilde{a}^{2^{(k-m-2)/2}} L_0^{2^{(k-1)/2}},$$

where $a_i \in [1, 2]$ for each i and $\widetilde{a} \in [1/2, 1]$. Hence,

$$\log P_N^{-1} = -\sqrt{\frac{N}{2^{m+2}}} \log \widetilde{a} - \log a_{(k-3)/2} - \dots - \frac{\sqrt{N/2}}{2} \log a_0 + \sqrt{\frac{N}{2}} \log L_0^{-1}.$$

Notice that $\log L_0^{-1} > 0$ and for each i, $0 \leq \log a_i \leq 1$ and $-1 \leq \log \widetilde{a} \leq 0$. Thus,

$$\log P_N^{-1} \leq \sqrt{\frac{N}{2}} \log L_0^{-1} + \sqrt{\frac{N}{2^{m+2}}}.$$

Finally,

$$\log P_N^{-1} \geq -\sqrt{\frac{N}{2}} + \sqrt{\frac{N}{2}} \log L_0^{-1} = \sqrt{\frac{N}{2}} \left(\log L_0^{-1} - 1\right).$$

□

Finally, by combining all of the analysis above for step-wise reduction of the total error probability, we can write general bounds when the initial error probability pair (α_0, β_0) lies inside B_m, where $m \neq 1$.

Theorem 3.3.4. *Suppose that $(\alpha_0, \beta_0) \in B_m$, where m is an integer and $m \neq 1$.*

*If $\log N < m$, then (*Corollary 3.3.1*)*

$$N\left(\log L_0^{-1} - 1\right) \leq \log P_N^{-1} \leq N \log L_0^{-1}.$$

If $\log N \geq m$, and $\log N - m$ is odd, then

$$\sqrt{2^{m-1} N}\left(\log L_0^{-1} - 1\right) \leq \log P_N^{-1} \leq \sqrt{2^{m-1} N} \log L_0^{-1}.$$

If $\log N \geq m$, and $\log N - m$ is even, then

$$\sqrt{2^{m-2} N}\left(\log L_0^{-1} - 1\right) \leq \log P_N^{-1} \leq \sqrt{2^{m} N} \log L_0^{-1}.$$

Proof. If $\log N < m$, then this scenario is the same as that of Corollary 3.3.1. Therefore,

$$N\left(\log L_0^{-1} - 1\right) \leq \log P_N^{-1} \leq N \log L_0^{-1}.$$

If $\log N \geq m$ and $\log N - m$ is odd, then it takes $(m-1)$ steps for the system to move into B_1. After it arrives in B_1, there is an even number of levels left because $\log N - m$ is odd.

By Proposition 3.3.1, we have

$$L_{k+1} = \widetilde{a}_k L_k^2$$

for $k = 0, 1, \ldots, m-2$ and some $\widetilde{a}_k \in [1, 2]$, and in consequence of Proposition 3.3.5,

$$L_{k+2} = a_k L_k^2$$

for $k = m-1, m-3, \ldots, \log N - 2$ and some $a_k \in [1, 2]$. Thus,

$$L_k = a_{(k+m-3)/2} \cdot a_{(k+m-5)/2}^2 \cdots a_0^{2^{(k+m-3)/2}} L_0^{2^{(k+m-1)/2}},$$

where $a_i \in [1, 2]$ for each i.

Let $k = \log N$. Then we obtain

$$\begin{aligned}\log P_N^{-1} = & -\log a_{(k+m-3)/2} - 2\log a_{(k+m-5)/2} - \ldots \\ & - \frac{\sqrt{2^{m-1}N}}{2} \log a_0 + \sqrt{2^{m-1}N} \log L_0^{-1}.\end{aligned}$$

Note that $\log L_0^{-1} > 0$, and for each i, $0 \leq \log a_i \leq 1$. Thus,

$$\log P_N^{-1} \leq \sqrt{2^{m-1}N} \log L_0^{-1}.$$

Finally,

$$\begin{aligned}\log P_N^{-1} & \geq -\sqrt{2^{m-1}N} + \sqrt{2^{m-1}N} \log L_0^{-1} \\ & = \sqrt{2^{m-1}N} \left(\log L_0^{-1} - 1\right).\end{aligned}$$

For the case where $\log N - m$ is even, the proof is similar and it is omitted. □

Remark 3.3.1. Notice again that the lower bounds for $\log P_N^{-1}$ above are useful only if $L_0 < 1/2$. However, similar to Proposition 3.3.7, we can derive a lower bound for $\log P_N^{-1}$, which is useful for all $L_0 \in (0, 1)$. It turns out that this lower bound differs from that in Proposition 3.3.7 by a constant term. Therefore, it is omitted.

3.3.3 Invariant Region in B_1

Consider the region $\{(\alpha,\beta)\in\mathcal{U}|\beta\le\sqrt{\alpha}$ and $\beta\ge 1-(1-\alpha)^2\}$, which is a subset of B_1 (see Fig. 3.11(a)). Denote the union of this region and its reflection with respect to $\beta=\alpha$ by S. It turns out that S is also invariant.

Proposition 3.3.10. *If $(\alpha_{k_0},\beta_{k_0})\in S$, then $(\alpha_k,\beta_k)\in S$ for all $k\ge k_0$.*

Proof. Without loss of generality, we consider the upper half of S, denoted by $S_{\mathcal{U}}$. As we shall see, the image of $S_{\mathcal{U}}$ is exactly the reflection of $S_{\mathcal{U}}$ with respect to the line $\beta=\alpha$ (denoted by $S_{\mathcal{L}}$). We know that $S_{\mathcal{U}}:=\{(\alpha,\beta)\in\mathcal{U}|\beta\le\sqrt{\alpha}$ and $\beta\ge 1-(1-\alpha)^2\}$.

The image of $S_{\mathcal{U}}$ under f can be calculated by

$$(\alpha',\beta')=f(\alpha,\beta)=(1-(1-\alpha)^2,\beta^2),$$

where $(\alpha,\beta)\in\mathcal{U}$. The above relation is equivalent to

$$(\alpha,\beta)=(1-\sqrt{1-\alpha'},\sqrt{\beta'}).$$

Therefore, we can calculate images of boundaries for $R_{\mathcal{U}}$ under f.

The image of the upper boundary $\beta\le\sqrt{\alpha}$ is

$$\sqrt{\beta'}\le\sqrt{1-\sqrt{1-\alpha'}};$$

i.e.,

$$\alpha'\ge 1-(1-\beta')^2,$$

and that of the lower boundary $\beta\ge 1-(1-\alpha)^2$ is

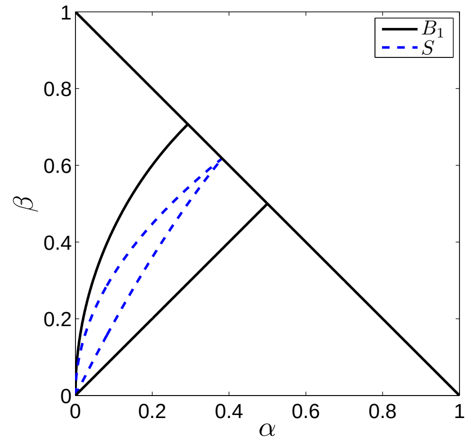
$$\sqrt{\beta'}\ge 1-(1-(1-\sqrt{1-\alpha'}))^2;$$

i.e.,

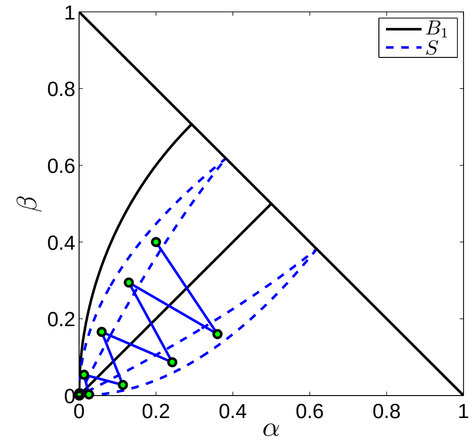
$$\alpha'\le\sqrt{\beta'}.$$

The function f is monotone. Hence, images of boundaries of $S_{\mathcal{U}}$ are boundaries of $S_{\mathcal{L}}$. Notice that boundaries of $R_{\mathcal{L}}$ are symmetric with those of $R_{\mathcal{U}}$ about $\beta=\alpha$. We conclude that S is an invariant region. $\square$

Fig. 3.11(b) shows a single trajectory of the dynamic system which stays inside S.



(a)



(b)

Figure 3.11: (a) Invariant region S (between dashed lines) lies inside B_1 (between solid lines). (b) A trajectory of the system which stays inside S.

We have given bounds for P_N, which is (twice) the total error probability. It turns out that for the case where $(\alpha_0, \beta_0) \in S$, we can bound the Type I and Type II errors individually.

Proposition 3.3.11. *If $(\alpha_k, \beta_k) \in S$, then*

$$1 \leq \frac{\alpha_{k+2}}{\alpha_k^2} \leq 4$$

and

$$1 \leq \frac{\beta_{k+2}}{\beta_k^2} \leq 4.$$

Remark 3.3.2. It is easy to see that as long as the system stays inside B_1, then in a similar vein, these ratios α_{k+2}/α_k^2 and β_{k+2}/β_k^2 are lower bounded by 1 and upper bounded by a constant. But recall that B_1 is not an invariant region. Thus, it is more interesting to consider S.

Proofs are omitted because they are along similar lines to those in the other proofs. As before, these inequalities give rise to bounds on sequences $\{\alpha_k\}$ and $\{\beta_k\}$. For example, for $\{\alpha_k\}$, we have the following.

Corollary 3.3.2. *If* $(\alpha_0, \beta_0) \in S$ *and* k *is even, then*

$$2^{k/2}\left(\log \alpha_0^{-1} - 2\right) \leq \log \alpha_k^{-1} \leq 2^{k/2} \log \alpha_0^{-1}.$$

3.3.4 Unequal Likely Hypotheses

In this section we consider the situation of unequally likely hypotheses; that is, $\pi_0 \neq \pi_1$. Suppose that the fusion rule is as before: The likelihood ratio test with unit threshold. The resulting total error probability for the nodes at level k is equal to $\hat{L}_k = \pi_0\alpha_k + \pi_1\beta_k$, and the total error probability at the fusion center is $\hat{P}_N = \hat{L}_{\log N}$. We are interested in bounds for $\hat{P}_N$.

Because the fusion rule is the same as before, the previous bounds for $\log L_k^{-1}$ hold. From these bounds, we now derive bounds for $\hat{P}_N$. Without loss of generality, we assume that $\pi_0 \leq \pi_1$. We obtain the following:

$$\pi_0 L_k \leq \pi_0\alpha_k + \pi_1\beta_k \leq \pi_1 L_k.$$

From these inequalities, we can derive upper and lower bounds for $\log \hat{P}_N^{-1}$. For example, in the case where $(\alpha_0, \beta_0) \in R$ and $\log N$ is even (even-height tree), from Theorem 3.3.1, we have

$$\sqrt{N}(\log L_0^{-1} - 1) \leq \log P_N^{-1} \leq \sqrt{N} \log L_0^{-1},$$

from which we obtain

$$\sqrt{N}(\log L_0^{-1} - 1) + \log \pi_1^{-1} \leq \log \hat{P}_N^{-1} \leq \sqrt{N} \log L_0^{-1} + \log \pi_0^{-1}.$$

We have derived error probability bounds for balanced binary relay trees under several scenarios. In the next section, we will use these bounds to study the asymptotic rate of convergence.

3.4 Asymptotic Rates

The asymptotic decay rate of the total error probability with respect to N is considered while the performance of the sensors is constant is the first problem to be tackled. Then we allow the sensors to be asymptotically crummy, in the sense that $\alpha_0 + \beta_0 \to 1$. We prove that the total error probability still converges to 0 under certain conditions. Last, we will compare the detection performance by applying different strategies in balanced binary relay trees.

3.4.1 Asymptotic Decay Rate

Notice that as N becomes large, the sequence $\{(\alpha_k, \beta_k)\}$ will eventually move into the invariant region R at some level and stays inside from that point. Therefore, it suffices to consider the decay rate in the invariant region R. Because error probability bounds for trees with odd height differ from those of the even-height tree by a constant term, without loss of generality, we will only consider trees with even height.

Proposition 3.4.1. *If $L_0 = \alpha_0 + \beta_0$ is fixed, then*

$$\log P_N^{-1} = \Theta(\sqrt{N}).$$

Proof. If $L_0 = \alpha_0 + \beta_0$ is fixed, then by Proposition 3.3.7 we immediately see that $P_N \to 0$ as $N \to \infty$ ($\log P_N^{-1} \to \infty$) and there exists a finite k such that $L_k < 1/2$. To analyze the asymptotic rate, we may assume that $L_0 < 1/2$. In this case, the bounds in Theorem 3.3.1 show that $\log P_N^{-1} = \Theta(\sqrt{N})$. $\square$

This implies that the convergence of the total error probability is sub-exponential; more precisely, the exponent is essentially $\sqrt{N}$.

In the special case where $(\alpha_0, \beta_0) \in S$, the Type I and Type II error probabilities decay to 0 with exponent $\sqrt{N}$ individually. Moreover, it is easy to show that the exponent is still $\sqrt{N}$ even if the prior probabilities are unequal.

Given $L_0 \in (0, 1)$ and $\epsilon \in (0, 1)$, suppose that we wish to determine how many sensors we need to have so that $P_N \leq \epsilon$. If $L_0 < 1/2$, then the solution is simply to find an N (e.g., the smallest) satisfying the inequality

$$\sqrt{N}\left(\log L_0^{-1} - 1\right) \geq -\log \epsilon.$$

In consequence, we have

$$N \geq ((\log L_0^{-1} - 1)\log \epsilon)^2.$$

The smallest N grows like $\Theta((\log \epsilon)^2)$ (cf., [60], in which the smallest N has a larger growth rate). If $L_0 \geq 1/2$, then by Proposition 3.3.7 we can deduce how many levels k are required so that $L_k < 1/2$:

$$\sqrt[4]{N}\log L_0^{-1} > -\log\frac{1}{2} = 1.$$

Therefore, N has to satisfy

$$N > (\log L_0^{-1})^{-4},$$

which implies that

$$k > 4\log(\log L_0{}^{-1})^{-1}.$$

Combining with the above analysis for the case where $L_0 < 1/2$, we can then determine the number of sensors required so that $P_N \leq \epsilon$.

3.4.2 Unequal Prior Probabilities

Using the Bayesian likelihood ratio test (the threshold given by the ratio of the prior probabilities), if the one of the two child nodes at level k sends '0' and the other node sends '1', then the test is given by

$$\frac{\alpha_k(1-\alpha_k)}{\beta_k(1-\beta_k)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0}{\pi_1} = c.$$

This rule reduces to 'AND' or 'OR' rule depends on the ratio. We wish to show that the system will not choose the same rule for consecutive 3 times. Suppose that α_k and β_k are sufficiently small and

$$\frac{\alpha_k(1-\alpha_k)}{\beta_k(1-\beta_k)} = c.$$

Then, we have $\alpha_k \approx c\beta_k$. Moreover, $\alpha_{k+1} \approx 2\alpha_k = 2c\beta_k$, $\beta_{k+1} = \beta_k^2$, and

$$\frac{\alpha_{k+1}(1-\alpha_{k+1})}{\beta_{k+1}(1-\beta_{k+1})} > c.$$

Then, we have $\alpha_{k+2} \approx \alpha_{k+1}^2 = 4c^2\beta_k^2$ and $\beta_{k+2} \approx 2\beta_{k+1} = 2\beta_k^2$. Also, $\alpha_{k+3} \approx \alpha_{k+2}^2 = 16c^4\beta_k^4$ and $\beta_{k+3} = 4\beta_k^2$. Then we have

$$\frac{\alpha_{k+3}(1-\alpha_{k+3})}{\beta_{k+3}(1-\beta_{k+3})} \approx 16c^4\beta_k^2 \leq c$$

for sufficiently small β_k. Hence, the system will not choose the same fusion rule for consecutive 3 times. In this case, we can show that

$$\log P_N^{-1} = \Theta(\sqrt{N}).$$

3.4.3 Crummy Sensors

In this part we allow the total error probability of each sensor, denoted by $L_0^{(N)}$, to depend on N but still to be constant across sensors.

If $L_0^{(N)}$ is bounded by some constant $L \in (0,1)$ for all N, then clearly $P_N \to 0$. It is more interesting to consider $L_0^{(N)} \to 1$, which means that sensors are asymptotically crummy.

Proposition 3.4.2. *Suppose that $L_0^{(N)} = 1 - \eta_N$ with $\eta_N \to 0$.*

(1) If $\eta_N \geq c_1/\sqrt[4]{N}$, then $P_N \leq e^{-c_1}$.

(2) If $\eta_N = \omega(1/\sqrt[4]{N})$, then $P_N \to 0$.

(3) If $\eta_N \leq c_2/\sqrt{N}$, then $P_N \geq e^{-c_2}$.

(4) If $\eta_N = o(1/\sqrt{N})$, then $P_N \to 1$.

Proof. First we consider part (1). We have

$$\sqrt[4]{N}\log(L_0^{(N)})^{-1} = -\sqrt[4]{N}\log(1-\eta_N).$$

But as $x \to 0$, $-\log(1-x) \sim x/\ln(2)$, from which we obtain

$$\sqrt[4]{N}\log(L_0^{(N)})^{-1} \sim \eta_N\sqrt[4]{N}/\ln(2).$$

From Proposition 3.3.7, it is easy to see that if we have $\eta_N \geq c_1/\sqrt[4]{N}$, then for sufficiently large N we obtain

$$\log P_N^{-1} \geq \sqrt[4]{N}\log(L_0^{(N)})^{-1} \geq c_1/\ln(2),$$

that is,

$$P_N \leq 2^{-c_1/\ln(2)} = e^{-c_1}.$$

Moreover, if $\eta_N\sqrt[4]{N} \to \infty$, that is, $\eta_N = \omega(1/\sqrt[4]{N})$, then $P_N \to 0$. This finishes the proof for part (2).

Next we consider parts (3) and (4). We have

$$\sqrt{N}\log(L_0^{(N)})^{-1} = -\sqrt{N}\log(1-\eta_N),$$

from which we obtain

$$\sqrt{N}\log(L_0^{(N)})^{-1} \sim \eta_N\sqrt{N}/\ln(2).$$

From Theorem 3.3.1, it is easy to see that if we have $\eta_N \leq c_2/\sqrt{N}$, then for sufficiently large N we obtain

$$\log P_N^{-1} \leq \sqrt{N}\log(L_0^{(N)})^{-1} \leq c_2/\ln(2),$$

that is,

$$P_N \geq 2^{-c_2/\ln(2)} = e^{-c_2}.$$

Moreover, if $\eta_N\sqrt{N} \to \infty$, that is, $\eta_N = o(1/\sqrt{N})$, then $P_N \to 1$.

□

Using part (3) of the above proposition, we derive a necessary condition for $P_N \to 0$.

Corollary 3.4.1. *Suppose that* $L_0^{(N)} = 1 - \eta_N$ *with* $\eta_N \to 0$. *Then,* $P_N \to 0$ *implies that* $\eta_N = \omega(1/\sqrt{N})$.

3.4.4 Comparison of Simulation Results

We end this section by comparing the quantitative behavior of the unit-threshold likelihood ratio rule with that of other fusion rules of interest. First, we define two particular fusion rules that can be applied at an individual node:

- OR rule: the parent node decides 0 if and only if both the child nodes send 0;
- AND rule: the parent node decides 1 if and only if both the child nodes send 1.

Notice that the unit-threshold likelihood ratio rule reduces to either the AND rule or the OR rule, depending on the values of the Type I and Type II error probabilities at the particular level of the tree. For our quantitative comparison, we consider three system-wide fusion strategies that we will compare with the case that uses the unit-threshold likelihood ratio rule at every node:

- OR strategy: Every fusion uses the OR rule;
- AND strategy: Every fusion uses the AND rule;
- RAND strategy: At each level of the tree, we randomly pick either the AND rule or the OR rule with equal probability, and independently over levels, and apply that rule to all the nodes at that level.

In Fig. 3.12, we show plots of the total error probability as a function of N for the tree that uses the unit-threshold likelihood ratio rule at every node (the one analyzed in this thesis). We also plot the total error probabilities for the AND and OR strategies, as well as the average total error probability over 100 independent trials of the RAND strategy. For comparison purposes, we also plot the error probability curve of the centralized parallel fusion strategy.

We can see from Fig. 3.12 that the total error probability for the centralized parallel strategy decays to 0 faster than that of the binary relay tree that uses the unit-threshold likelihood

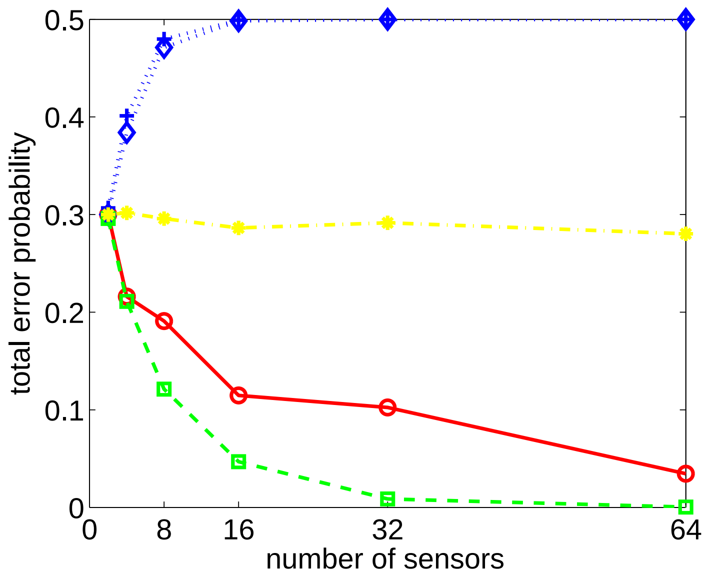


Figure 3.12: Total error probability plots. Dashed line: centralized parallel fusion strategy. Solid line: unit-threshold likelihood ratio rule for balanced binary relay tree. Dotted line with '$\diamond$' marker: OR strategy. Dotted line with '+' marker: AND strategy. Dash-dot line: RAND strategy.

ratio rule at every node. This is not surprising, because the former is known to be exponential, as discussed earlier, while the latter is sub-exponential with exponent $\sqrt{N}$, as shown in this thesis. The AND and OR strategies both result in total error probabilities converging monotonically to $1/2$, while the RAND strategy results in an average total error probability that does not decrease much with N.

CHAPTER 4

BALANCED BINARY RELAY TREES WITH NODE AND LINK FAILURES

We have studied the detection performance of balanced binary relay trees in the last chapter. However, we have not accounted for the possibility of node and link failures, and we have assumed that all messages are received reliably. In this chapter, we address these two issues.

4.1 Related Work

In practical scenarios, the nodes are failure-prone and the communication channels are not perfect in the decentralized network, wherein messages are subject to random erasures. The literature on hypothesis testing problem in tree networks with node and link failures is quite limited. Tay *et al.* [70] provide asymptotic analysis about the impact of imperfect nodes and links modeled as binary symmetric channels in trees with bounded height using branching process, Chernoff bounds, etc. However, the detection performance for unbounded-height trees with failure-prone nodes and links is still open.

In this thesis, we investigate the distributed detection problem in the context of balanced binary relay trees where nodes and links fail with certain probabilities. This is the first performance analysis for unbounded-height trees with imperfect nodes and links. We derive non-asymptotic bounds for the total error probability P_N as functions of N, which characterize the asymptotic decay rate of the total error probability. We show that the detection performance in this failure case cannot be better than that in the non-failure case. However, we derive an explicit necessary and sufficient condition on the decay rate of the local failure probabilities p_k (combination of node and link failure probabilities at each level) such that the decay rate of the total error probability in the failure case is the same as that of the non-failure case. More precisely, we show that $\log P_N^{-1} = \Theta(\sqrt{N})$ if and only if $\log p_k^{-1} = \Omega(2^{k/2})$.

4.2 Problem Formulation

We consider the problem of binary hypothesis testing between H_0 and H_1 in a balanced binary relay tree with failure-prone nodes and links, shown in Fig. 4.1 (the notation there will be defined below). Each sensor (circle) sends a binary message upward to its parent node. Each relay node (diamond) fuses two binary messages from its child nodes into a new binary message, which is then sent to the node at the next level. This process is repeated culminating at the fusion center, where an overall binary decision is made. We assume that all sensors are conditionally independent given each hypothesis, and that all sensor messages have identical Type I error probability α_0 (also known as probability of false alarm) and identical Type II error probability β_0 (also known as probability of missed detection). Moreover, we assume that each node at level k fails with identical node failure probability n_k (a failed node cannot transmit any message upward). We model each link as a binary erasure channel as shown in Fig. 2.2. With a certain probability, the input message X (either 0 or 1) gets erased and the receiver does not get any data. We assume that the links between nodes at height k and height $k+1$ have identical probability of erasure ℓ_k.

Consider a node $\mathcal{N}_k$ at level k connected to its parent node $\mathcal{N}_{k+1}$ at level $k+1$. We define several events as follows:

- $E_k^{(1)}$: the event that the node $\mathcal{N}_k$ does not have a message to transmit; i.e., $\mathcal{N}_k$ does not receive any messages from both its child nodes. We denote the probability of this event by $\mathcal{P}_k$ and we call it the *starvation probability*.
- $E_k^{(2)}$: the event that either the node $\mathcal{N}_k$ fails or the link from $\mathcal{N}_k$ to $\mathcal{N}_{k+1}$ fails. We call the occurrence of $E_k^{(2)}$ a *local failure* and we denote by p_k the local failure probability.
- $E_k^{(3)}$: the event that $\mathcal{N}_{k+1}$ does not receive a message from $\mathcal{N}_k$. We denote the probability of this event by q_k and we call it the *silence probability*.

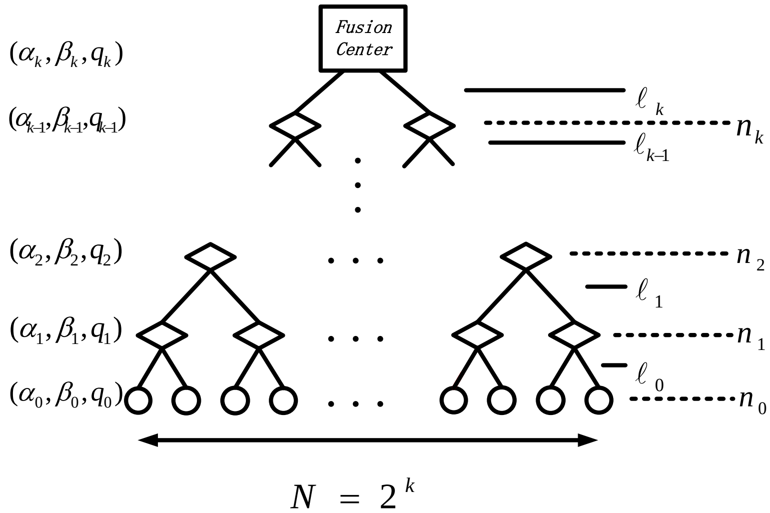


Figure 4.1: A balanced binary relay tree with node and link failures.

Note that $E_k^{(3)}$ occurs if and only if either (i) the node $\mathcal{N}_k$ does not have a message to transmit (event $E_k^{(1)}$), or (ii) the node $\mathcal{N}_k$ does have a message to transmit but a local failure occurs (event $E_k^{(2)}$). The probability of case (i) is simply $\mathcal{P}_k$. The probability of case (ii) is p_k, which equals the conditional probability of $E_k^{(3)}$ given $\bar{E}_k^{(1)}$ (the complement event $E_k^{(1)}$, which means that $\mathcal{N}_k$ has a message to transmit). Thus,

$$\begin{aligned} p_k &= \mathbb{P}(E_k^{(2)}) \\ &= \mathbb{P}(E_k^{(3)}|\bar{E}_k^{(1)}) \\ &= n_k + \ell_k - n_k\ell_k. \end{aligned}$$

By the law of total probability, we have

$$\begin{aligned} q_k &= \mathbb{P}(E_k^{(3)}) \\ &= \mathbb{P}(E_k^{(1)}) + \mathbb{P}(E_k^{(3)}|\bar{E}_k^{(1)})\mathbb{P}(\bar{E}_k^{(1)}) \\ &= \mathcal{P}_k + p_k(1 - \mathcal{P}_k). \end{aligned}$$

Consider the parent node $\mathcal{N}_{k+1}$. This node does not have a message to transmit (event $E_{k+1}^{(1)}$) if and only if it does not receive messages from both its two child nodes. The probability

$\mathcal{P}_{k+1}$ of this event is

$$\mathcal{P}_{k+1} = q_k^2 = (\mathcal{P}_k + p_k(1 - \mathcal{P}_k))^2.$$

Recursively, we can show that the probability of the event that the parent node of $\mathcal{N}_{k+1}$ does not receive messages from $\mathcal{N}_{k+1}$ is

$$\begin{aligned} q_{k+1} &= \mathcal{P}_{k+1} + p_{k+1}(1 - \mathcal{P}_{k+1}) \\ &= q_k^2 + p_{k+1}(1 - q_k^2), \end{aligned}$$

where $p_{k+1} = n_{k+1} + \ell_{k+1} - n_{k+1}\ell_{k+1}$ denotes the local failure probability for level $k+1$.

Again, we denote the Type I and Type II error probabilities for the nodes at level k by α_k and β_k, respectively. If $\mathcal{N}_{k+1}$ receives data from only one of its two child nodes, then the Type I and Type II error probabilities do not change since the parent node receives only one binary message and directly sends this message without fusion. The probability of this event is $2q_k(1 - q_k)$, in which case we have

$$(\alpha_{k+1}, \beta_{k+1}) = (\alpha_k, \beta_k).$$

If the parent node receives messages from both child nodes, then the scenario is the same as that in Chapter 3. The probability of this event is $(1 - q_k)^2$, in which case we have

$$(\alpha_{k+1}, \beta_{k+1}) = \begin{cases} (1 - (1 - \alpha_k)^2, \beta_k^2), & \text{if } \alpha_k \leq \beta_k, \\ (\alpha_k^2, 1 - (1 - \beta_k)^2), & \text{if } \alpha_k > \beta_k. \end{cases}$$

Consider the mean Type I and Type II error probabilities conditioned on the event that the parent node receives at least one message from its child nodes; i.e., the parent node has data. We have

$$(\alpha_{k+1}, \beta_{k+1}, q_{k+1}) = f(\alpha_k, \beta_k, q_k),$$

where

$$f(\alpha_k, \beta_k, q_k) := \begin{cases} \left(\frac{(1-q_k)(2\alpha_k - \alpha_k^2) + 2q_k\alpha_k}{1+q_k}, \frac{(1-q_k)\beta_k^2 + 2q_k\beta_k}{1+q_k}, q_k^2 + (1 - q_k^2)p_{k+1} \right), & \text{if } \alpha_k \leq \beta_k, \\ \left(\frac{(1-q_k)\alpha_k^2 + 2q_k\alpha_k}{1+q_k}, \frac{(1-q_k)(2\beta_k - \beta_k^2) + 2q_k\beta_k}{1+q_k}, q_k^2 + (1 - q_k^2)p_{k+1} \right), & \text{if } \alpha_k > \beta_k. \end{cases} \tag{4.1}$$

Recall that all sensors have the same *error probability triplet* (α_0, β_0, q_0), where $q_0 = p_0 = n_0 + \ell_0 - n_0\ell_0$. Therefore, by the above recursion (4.1), all relay nodes at level 1 will have the same error probability triplet $(\alpha_1, \beta_1, q_1) = f(\alpha_0, \beta_0, q_0)$ (where α_1 and β_1 are the conditional mean error probabilities). Similarly we can calculate error probability triplets for nodes at all other levels. We have

$$(\alpha_{k+1}, \beta_{k+1}, q_{k+1}) = f(\alpha_k, \beta_k, q_k), \quad k = 0, 1, \dots, \tag{4.2}$$

where (α_k, β_k, q_k) is the error probability triplet of nodes at the kth level of the tree.

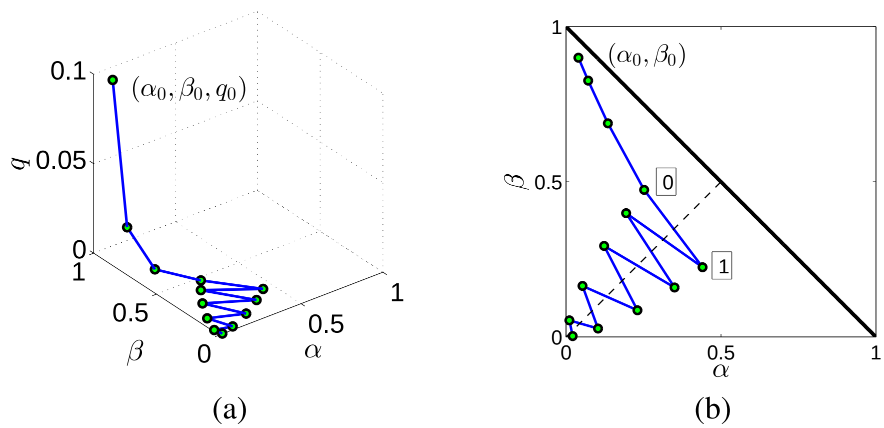


Figure 4.2: (a) An example trajectory of (α_k, β_k, q_k) in the (α, β, q) coordinates. (b) The trajectory in 4.2(a) projected onto the (α, β) plane.

Consider (α_k, β_k, q_k) as a discrete dynamic system governed by (4.2) with p_k as its input. Notice that the dynamic system depends on the exogenous parameters n_k and ℓ_k only through p_k. An example trajectory of this dynamic system is shown in Fig. 4.2(a), with the local failure probabilities given by $p_{k+1} = p_k^2$. We observe that q_k decreases very quickly to 0 in this case. In addition, as shown in Fig. 4.2(b), the trajectory approaches $\beta = \alpha$ at the beginning. After (α_k, β_k) gets too close to $\beta = \alpha$, the next pair $(\alpha_{k+1}, \beta_{k+1})$ will be repelled toward the other side of the line $\beta = \alpha$. This behavior is similar to the non-failure scenario (see Chapter 3), in which case there exists an invariant region in the sense that the system stays in the invariant region once the system enters it.

Is there an invariant region in the failure case where $p_k \neq 0$? We answer this question affirmatively by precisely describing this invariant region in $\mathbb{R}^3$ in Section 4.3. In doing so, we derive bounds on the total error probability as explicit functions of N in Section 4.4. This allows us to characterize the decay rate of the total error probability as N goes to infinity.

4.3 Evolution of Type I, Type II, and Silence Probabilities

Our analysis here builds on the method in Chapter 3. Notice that the recursion (4.1) is symmetric about the hyperplanes $\alpha + \beta = 1$ and $\beta = \alpha$. Thus, it suffices to study the evolution of the dynamic system only in the region bounded by $\alpha + \beta < 1$, $\beta \geq \alpha$, and $0 \leq q \leq 1$. Let

$$\mathcal{U} := \{(\alpha, \beta, q) \geq 0 | \alpha + \beta < 1,\ \beta \geq \alpha,\ \text{and}\ \ q \leq 1\}$$

be this triangular prism. Similarly, define the complementary triangular prism

$$\mathcal{L} := \{(\alpha, \beta, q) \geq 0 | \alpha + \beta < 1,\ \beta < \alpha,\ \text{and}\ \ q \leq 1\}.$$

First, we introduce the following region:

$$\begin{aligned} B :=& \{(\alpha, \beta, q) \in \mathcal{U} | \beta \leq -q/(1-q) + \\ & \sqrt{q^2 + (1-q)^2(2\alpha - \alpha^2) + 2q(1-q)\alpha}/(1-q)\}. \end{aligned}$$

It is easy to show that if $(\alpha_k, \beta_k, q_k) \in B$, then the next triplet $(\alpha_{k+1}, \beta_{k+1}, q_{k+1})$ jumps across the plane $\beta = \alpha$ away from (α_k, β_k, q_k). This process is shown in Fig. 4.2(b) from 0 to 1. More precisely, if $(\alpha_k, \beta_k, q_k) \in \mathcal{U}$, then $(\alpha_k, \beta_k, q_k) \in B$ if and only if $(\alpha_{k+1}, \beta_{k+1}, q_{k+1}) \in \mathcal{L}$. In other words, B is the *inverse image* of $\mathcal{L}$ in $\mathcal{U}$ under mapping f.

Note that if the initial error probability triplet is outside B; i.e., $(\alpha_0, \beta_0, q_0) \in \mathcal{U} \setminus B$, then before the system enters B, we have $\alpha_{k+1} > \alpha_k$ and $\beta_{k+1} < \beta_k$. Thus, the dynamic system moves toward the $\beta = \alpha$ plane, which means that if the number N of sensors is sufficiently large, then the dynamic system is guaranteed to enter B.

Next we consider the behavior of the system after it enters B. If $(\alpha_k, \beta_k, q_k) \in B$, we consider the position of the next pair $(\alpha_{k+1}, \beta_{k+1}, q_{k+1})$; i.e., we consider the *image* of B under f, which we denote by $R_{\mathcal{L}}$. Similarly we denote by $R_{\mathcal{U}}$ the reflection of $R_{\mathcal{L}}$ with respect to $\beta = \alpha$. This region is shown in Fig. 4.3 in the (α, β, q) coordinates. We find that

$$R_{\mathcal{U}} := \{(\alpha, \beta, q) \in \mathcal{U} | \beta \leq -\alpha + 2(\sqrt{q^2 + (1-q^2)\alpha} - q)/(1-q)\}.$$

The sets $R_{\mathcal{U}}$ and B have some interesting properties. We denote the projection of the upper boundary of $R_{\mathcal{U}}$ and B onto the (α, β) plane for a fixed q by $R_{\mathcal{U}}^q$ and B^q, respectively. It is easy to see that if $q_1 \leq q_2$, then $R_{\mathcal{U}}^{q_1}$ lies above $R_{\mathcal{U}}^{q_2}$ in the (α, β) plane. Similarly, if $q_1 \leq q_2$, then B^{q_1} lies above B^{q_2} in the (α, β) plane. Moreover, we have the following proposition.

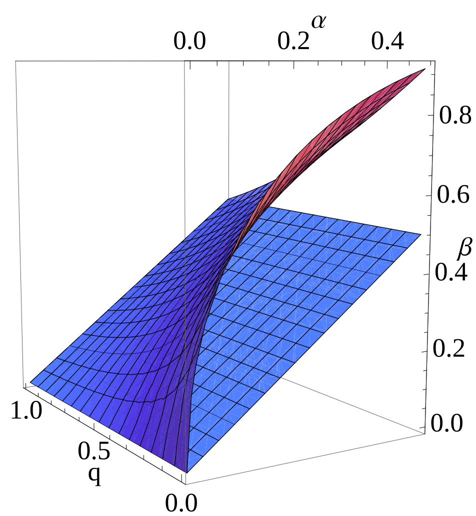


Figure 4.3: $R_{\mathcal{U}}$ in the (α, β, q) coordinates.

Proposition 4.3.1. $B \subset R_{\mathcal{U}}$.

Proof. B and $R_{\mathcal{U}}$ share the same lower boundary $\beta = \alpha$. Thus, it suffices to prove that the upper boundary of B lies below that of $R_{\mathcal{U}}$ for a fixed q; i.e., B^q lies above $R_{\mathcal{U}}^q$ in the (α, β) plane.

The upper boundary of B is given by

$$\beta = \frac{-q + \sqrt{q^2 + (1-q)^2(2\alpha - \alpha^2) + 2q(1-q)\alpha}}{1-q}.$$

The upper boundary of $R_{\mathcal{U}}$ is given by

$$\beta = -\alpha + 2\frac{\sqrt{q^2 + (1-q^2)\alpha} - q}{1-q}.$$

We need to prove the following:

$$\frac{-q + \sqrt{q^2 + (1-q)^2(2\alpha - \alpha^2) + 2q(1-q)\alpha}}{1-q} \leq -\alpha + 2\frac{\sqrt{q^2 + (1-q^2)\alpha} - q}{1-q}.$$

The above inequality can be simplified as follows:

$$\sqrt{q^2 + (1-q)^2(2\alpha - \alpha^2) + 2q(1-q)\alpha} \leq -\alpha(1-q) - q + 2\sqrt{q^2 + (1-q^2)\alpha}.$$

Squaring both sides and simplifying, we have

$$2\sqrt{q^2 + (1-q^2)\alpha}(\alpha(1-q) + q) \leq 2(q^2 + (1-q^2)\alpha) - (1-q)^2(\alpha - \alpha^2).$$

Again squaring both sides and simplifying, we have

$$\begin{aligned}
&4(q^2 + (1-q^2)\alpha)((1-q)\alpha + q)^2 \leq \\
&4(q^2 + (1-q^2)\alpha)^2 + (1-q)^4(\alpha - \alpha^2)^2 - 4(q^2 + (1-q^2)\alpha)(1-q)^2(\alpha - \alpha^2),
\end{aligned}$$

which can be simplified as follows:

$$\begin{aligned}
& 4(q^2 + (1-q^2)\alpha)(q^2 + 2q(1-q)\alpha + \\
& (1-q)^2\alpha^2 - q^2 - (1-q^2)\alpha + (1-q)^2(\alpha - \alpha^2)) \\
\leq\ & (1-q)^4(\alpha - \alpha^2)^2.
\end{aligned}$$

Fortuitously, the left-hand side turns out to be identically 0. Thus, the inequality holds. □

Note that B and $R_{\mathcal{U}}$ share the same lower boundary $\beta = \alpha$. Thus, it suffices to proof that the upper boundary of B lies below that of $R_{\mathcal{U}}$ for a fixed q; i.e., B^q lies above $R_{\mathcal{U}}^q$ in the (α, β) plane. The reader can refer to Figs. 4.4(a) and 4.4(b) for plots of the upper boundaries of $R_{\mathcal{U}}$ and B projected onto the (α, β) plane for two fixed values of q.

Let us denote by R the region $R_{\mathcal{U}} \cup R_{\mathcal{L}}$. Then, so far we have shown that if the tree height is sufficiently large the system enters R. Next we show below that R is an *invariant region* in the sense that once the system enters R, it stays there.

Proposition 4.3.2. *Suppose that* $(\alpha_{k_0}, \beta_{k_0}, q_{k_0}) \in R$ *for some* k_0 *and the sequence* $\{q_k\}$ *is non-increasing for* $k \geq k_0$. *Then,* $(\alpha_k, \beta_k, q_k) \in R$ *for all* $k \geq k_0$.

Proof. Without loss of generality, we assume that $(\alpha_k, \beta_k, q_k) \in R_{\mathcal{U}}$. We know that $R_{\mathcal{L}}$ is the image of $\mathcal{U}$ in $\mathcal{L}$. Thus if the next state $(\alpha_{k+1}, \beta_{k+1}, q_{k+1}) \in \mathcal{L}$, then it must be inside $R_{\mathcal{L}}$. We already have $q_{k+1} \leq q_k$, which indicates that $R_{\mathcal{U}}^{q_{k+1}}$ lies above $R_{\mathcal{U}}^{q_k}$ in the (α, β) plane. Moreover, for a fixed q, the upper boundary $R_{\mathcal{U}}^{q}$ is monotone increasing in the (α, β) plane. We already know that $\alpha_{k+1} > \alpha_k$ and $\beta_{k+1} < \beta_k$. As a result, if the next state $(\alpha_{k+1}, \beta_{k+1}, q_{k+1}) \in \mathcal{U}$, then the next state is in fact inside $R_{\mathcal{U}}$. Note that in Fig. 4.2(b), the dynamic system stays in a neighbor region of $\beta = \alpha$ after it gets close to $\beta = \alpha$.

□

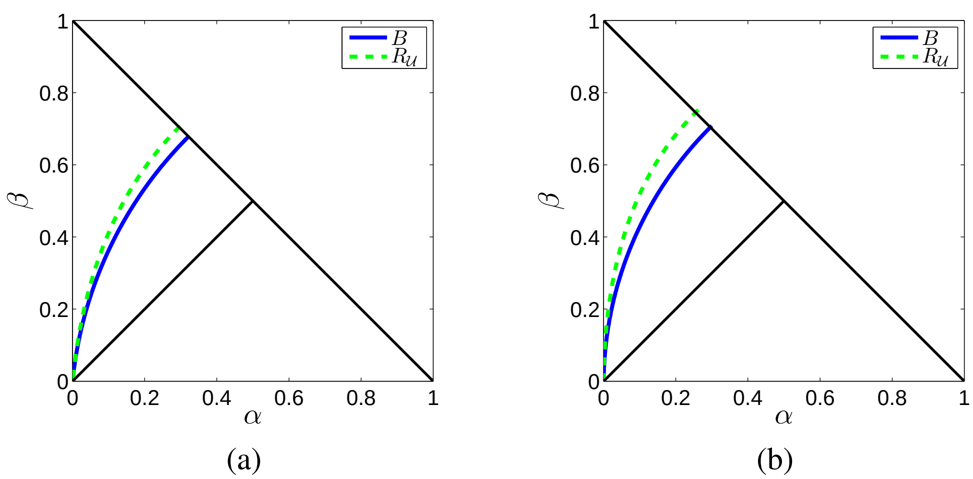


Figure 4.4: (a) Upper boundaries for $R_{\mathcal{U}}$ and B for $q = 0.1$. (b) Upper boundaries for $R_{\mathcal{U}}$ and B for $q = 0.01$.

To study the asymptotic detection performance, we can simply analyze the case where the system lies inside the invariant region and stays inside it. We assume that $\{q_k\}$ is a non-increasing sequence. We will show in the next section that without this assumption, the decay rate is strictly more slowly than that of the non-failure case. Note that $\{q_k\}$ is a sequence depending on the input p_k, which in turn depends on the exogenous parameters n_k and ℓ_k. Next we provide a sufficient condition for $\{q_k\}$ to be non-increasing.

Proposition 4.3.3. *Suppose that $p_{k+1} \leq p_k$ for all k and $q_1 \leq q_0$. Then, $\{q_k\}$ is a non-increasing sequence.*

Proof. The recursive relation for q_k is:

$$q_{k+1} = q_k^2 + (1 - q_k^2)p_{k+1}.$$

Since $\{p_k\}$ is non-increasing, we have

$$\begin{aligned} q_{k+2} &= q_{k+1}^2 + (1 - q_{k+1}^2)p_{k+2} \\ &\leq q_{k+1}^2 + (1 - q_{k+1}^2)p_{k+1}. \end{aligned}$$

Notice that this recursion is simply a weighted sum of 1 and p_{k+1}. From the initial condition that $q_1 \leq q_0$, it is easy to see that $q_{k+1} \leq q_k$ using mathematical induction.

□

Henceforth, we assume that p_k is non-increasing and therefore q_k is monotone non-increasing as well. Based on the above propositions, in the next section we study the reduction of the total error probability when the system lies in R to determine the asymptotic decay rate.

4.4 Error Probability Bounds and Asymptotic Decay Rates

In this section, we first compare the step-wise reduction of the total error probability between the failure case and non-failure case. Then, we show that the decay of the failure case cannot be faster than that of the non-failure case. However, we provide a sufficient condition such that the scaling law of the decay rate in the failure case remains the same as that of the non-failure case and we discuss how this sufficient condition is satisfied in terms of the input parameter p_k.

4.4.1 Step-wise Reduction and Asymptotic Decay Rate

We will first consider the case where the prior probabilities are equal; i.e., $\pi_0 = \pi_1 = 1/2$. We define $L_k = \alpha_k + \beta_k$ to be (twice) the total error probability for nodes at level k.

4.4.1.1 Step-wise reduction

In this part, we show that in the failure case, the decay of the total error probability for a single step cannot be faster than that of the non-failure case.

Proposition 4.4.1. *Let* $L_{k+1}^{(q)} = \alpha_{k+1}^{(q)} + \beta_{k+1}^{(q)}$ *be (twice) the total error probability at the next level from the current state* (α_k, β_k, q). *Suppose that* (α_k, β_k, q_1) *and* $(\alpha_k, \beta_k, q_2) \in \mathcal{U}$. *If* $q_1 < q_2$, *then*

$$L_{k+1}^{(q_1)} \leq L_{k+1}^{(q_2)}$$

with equality if and only if $\alpha_k = \beta_k$.

Proof. It is easy to show the following inequality

$$\begin{aligned} & 2\alpha_k - \alpha_k^2 + \beta_k^2 \leq \alpha_k + \beta_k \\ \Longleftrightarrow & \beta_k^2 - \alpha_k^2 \leq \beta_k - \alpha_k \end{aligned}$$

holds in the region $\alpha_k + \beta_k < 1$ and $\beta_k \geq \alpha_k$. The equality is satisfied if and only if $\beta_k = \alpha_k$.

From the recursion described in (4.1), we have

$$L_{k+1}^{(q)} = \frac{1-q}{1+q} L_{k+1}^{(0)} + \frac{2q}{1+q}(\alpha_k + \beta_k),$$

where $L_{k+1}^{(0)} = 2\alpha_k - \alpha_k^2 + \beta_k^2$. Notice that

$$\frac{1-q}{1+q} + \frac{2q}{1+q} = 1.$$

Therefore, we can write

$$L_{k+1}^{(q_1)} = \delta_1 L_{k+1}^{(0)} + (1-\delta_1)(\alpha_k + \beta_k),$$

where $\delta_1 = (1-q_1)/(1+q_1)$. Let $\delta_2 = (1-q_2)/(1+q_2)$. Then, it is easy to see that $\delta_1 \geq \delta_2$. Thus, we have

$$\begin{aligned} L_{k+1}^{(q_1)} &= \delta_1 L_{k+1}^{(0)} + (1-\delta_1)(\alpha_k + \beta_k) + (\delta_2 - \delta_1) L_{k+1}^{(0)} - (\delta_2 - \delta_1) L_{k+1}^{(0)} \\ &\leq \delta_1 L_{k+1}^{(0)} + (1-\delta_1)(\alpha_k + \beta_k) + (\delta_2 - \delta_1) L_{k+1}^{(0)} - (\delta_2 - \delta_1)(\alpha_k + \beta_k) = L_{k+1}^{(q_2)}. \end{aligned}$$

This completes the proof.

□

From Proposition 4.4.1, we immediately deduce that if $q > 0$, then $L^{(0)}_{k+1} \leq L^{(q)}_{k+1}$. This means that the decay of the total error probability for a single step is fastest if the silence probability is 0 (non-failure case). In other words, for the failure case, the step-wise shrinkage of the total error probability cannot be faster than that of the non-failure case, where the total error probability decays to 0 with exponent $\sqrt{N}$. In addition, we show in this section that the asymptotic decay rate for the failure case cannot be faster than that of the non-failure case.

4.4.1.2 Asymptotic decay rate

With the assumption of equally likely hypotheses, we denote (twice) the total error probability for nodes at the fusion center by $P_N := L_{\log N}$. Using Proposition 4.4.1, we provide an upper bound for $\log P_N^{-1}$, which in turn provides an upper bound for the decay rate.

Theorem 4.4.1. *Suppose that* $(\alpha_0, \beta_0, q_0) \in R$. *Then,*

$$\log P_N^{-1} \leq \sqrt{N}\left(\log L_0^{-1} + 1\right).$$

Proof. From the assumptions that q_k is monotone non-increasing and $(\alpha_0, \beta_0, q_0) \in R$, we shall see that the dynamic system stays inside R. First we show the following inequality for the system in R:

$$\frac{L_{k+2}}{L_k^2} \geq \frac{1}{2}. \tag{4.3}$$

The evolution of the system is

$$(\alpha_k, \beta_k, q_k) \to (\alpha_{k+1}, \beta_{k+1}, q_{k+1}) \to (\alpha_{k+2}, \beta_{k+2}, q_{k+2}).$$

From Proposition 4.4.1, we have

$$L_{k+2}^{(0)} \leq L_{k+2},$$

where $L_{k+2}^{(0)} = 2\alpha_{k+1} - \alpha_{k+1}^2 + \beta_{k+1}^2$ as defined before. To prove $L_{k+2}/L_k^2 \geq 1/2$, it suffices to show that $L_{k+2}^{(0)}/L_k^2 \geq 1/2$. We divide our proof into two cases: $(\alpha_k, \beta_k, q_k) \in R_u \setminus B$ and $(\alpha_k, \beta_k, q_k) \in B$.

Case I. If $(\alpha_k, \beta_k, q_k) \in R_u \setminus B$, then

$$\frac{L_{k+2}^{(0)}}{L_k^2} = \frac{2\alpha_{k+1} - \alpha_{k+1}^2 + \beta_{k+1}^2}{(\alpha_k + \beta_k)^2}.$$

From the recursion (4.1), we have

$$\alpha_{k+1} = \frac{1 - q_k}{1 + q_k}(2\alpha_k - \alpha_k^2) + \frac{2q_k}{1 + q_k}\alpha_k \geq \alpha_k$$

and

$$\beta_{k+1} = \frac{1 - q_k}{1 + q_k}\beta_k^2 + \frac{2q_k}{1 + q_k}\beta_k \geq \beta_k^2.$$

Thus, it suffices to show that

$$\frac{2\alpha_k - \alpha_k^2 + \beta_k^4}{(\alpha_k + \beta_k)^2} \geq \frac{1}{2}.$$

It is easy to see that

$$2(2\alpha_k - \alpha_k^2) \geq 1 - (1 - \alpha_k)^4.$$

Hence, it suffices to show that

$$(1 - (1 - \alpha_k)^4 + \beta_k^4) \geq (\alpha_k + \beta_k)^2,$$

which has been proved in Chapter 3.

Case II. If $(\alpha_k, \beta_k, q_k) \in B$, then it suffices to show that

$$\frac{\alpha_{k+1}^2 + 2\beta_{k+1} - \beta_{k+1}^2}{(\alpha_k + \beta_k)^2} \geq \frac{1}{2}.$$

Again from (4.1), we have

$$\alpha_{k+1} = \frac{1-q_k}{1+q_k}(2\alpha_k - \alpha_k^2) + \frac{2q_k}{1+q_k}\alpha_k \geq \alpha_k$$

and

$$\beta_{k+1} = \frac{1-q_k}{1+q_k}\beta_k^2 + \frac{2q_k}{1+q_k}\beta_k \geq \beta_k^2.$$

Thus, it suffices to prove that

$$\frac{\alpha_k^2 + \beta_k^2}{(\alpha_k + \beta_k)^2} \geq \frac{1}{2},$$

which is obvious. This proves (4.3). We now prove the claim of Theorem 4.4.1. From (4.3), we have

$$L_{k+2} = a_k L_k^2$$

for $k = 0, 1, \ldots, \log N - 2$ and some $a_k \geq 1/2$. Therefore, for $k = 2, 4, \ldots, \log N$, we have

$$L_k = a_{(k-2)/2} \cdot a_{(k-4)/2}^2 \cdots a_0^{2^{(k-2)/2}} L_0^{2^{k/2}},$$

where $a_i \geq 1/2$, $i = 0, 1, \ldots, (k-2)/2$. Taking logs and using $k = \log N$, we have

$$\begin{aligned}\log P_N^{-1} = &- \log a_{(k-2)/2} - 2\log a_{(k-4)/2} - \cdots \\ &- 2^{(k-2)/2}\log a_0 + \sqrt{N}\log L_0^{-1}.\end{aligned}$$

Notice that $\log L_0^{-1} > 0$ and $\log a_i \geq -1$ for all i. Thus,

$$\begin{aligned}\log P_N^{-1} &\leq \sqrt{N}\log L_0^{-1} + \sqrt{N} \\ &= \sqrt{N}\left(\log L_0^{-1} + 1\right).\end{aligned}$$

This completes the proof. $\square$

Theorem 4.4.1 provides an upper bound for $\log P_N^{-1}$. From this upper bound, it is easy to get an upper bound for the asymptotic decay rate.

Corollary 4.4.1. *Suppose that* $(\alpha_0, \beta_0, q_0) \in R$. *Then,*

$$\log P_N^{-1} = O(\sqrt{N}).$$

Compared with the decay rate for the non-failure case, the rate in Corollary 4.4.1 is not faster than $\sqrt{N}$ (note that the scaling law for decay rate for the non-failure case is exactly $\sqrt{N}$). This observation is unsurprising because the case where nodes and links are perfect has the best detection performance. But is it possible that the decay rate for the failure case remains $\sqrt{N}$? In the next section, we show that this is possible if the silence probabilities decay to 0 sufficiently fast.

4.4.2 Error Probability Bounds and Decay Rates

In this section, we first provide a sufficient condition for the ratio L_{k+2}/L_k^2 to be bounded. Then, we derive upper and lower bounds for the total error probability at the fusion center for trees with even and odd heights, in the equal prior scenario. Under the sufficient condition, we show that the decay rate of the total error probability remains the same as that of the non-failure case. We will also discuss the non-equal prior scenario.

Proposition 4.4.2. *Suppose that* $(\alpha_k, \beta_k, q_k) \in R$ *and* q_k *is monotone non-increasing. If* $q_k \leq CL_k$ *where* $C \geq 0$, *then the ratio* L_{k+2}/L_k^2 *is bounded as*

$$\frac{1}{2} \leq \frac{L_{k+2}}{L_k^2} \leq 6C + 2.$$

Proof. The lower bound of L_{k+2}/L_k^2 has been proved in Theorem 4.4.1. Here we derive the upper bound for L_{k+2}/L_k^2. Again we divide our proof into two cases: $(\alpha_k, \beta_k, q_k) \in R_u \setminus B$ and $(\alpha_k, \beta_k, q_k) \in B$.

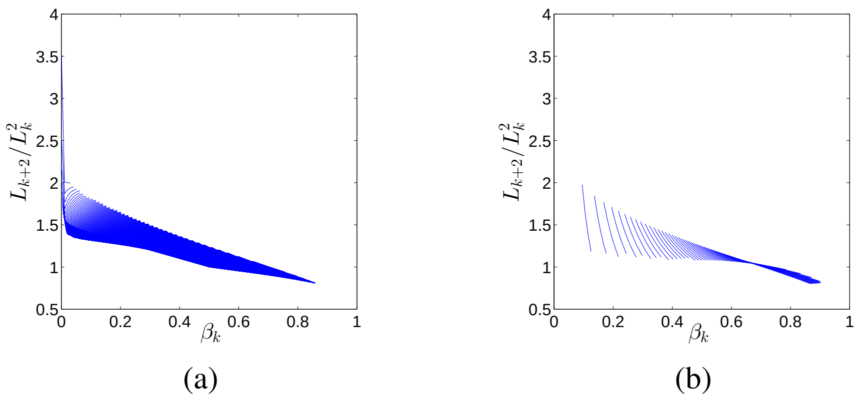


Figure 4.5: (a) The ratio L_{k+2}/L_k^2 in B for $C = 1$. (b) The ratio L_{k+2}/L_k^2 in $R_{\mathcal{U}} \setminus B$ for $C = 1$. Each line depicts the ratio versus β_k for a fixed α_k.

Case I. If $(\alpha_k, \beta_k, q_k) \in R_u \setminus B$, then

$$\frac{L_{k+2}}{L_k^2} \le \frac{L_{k+1}}{L_k^2} = \frac{1-q_k}{1+q_k}\frac{2\alpha_k - \alpha_k^2 + \beta_k^2}{(\alpha_k+\beta_k)^2} + \frac{2q_k}{(1+q_k)(\alpha_k+\beta_k)}. \tag{4.4}$$

Since $q_k \le CL_k$, the second term on the right-hand side of (4.4) is upper bounded as

$$\frac{2q_k}{(1+q_k)(\alpha_k+\beta_k)} \le 2C.$$

We now show the other term is bounded above, namely,

$$\frac{2\alpha_k - \alpha_k^2 + \beta_k^2}{(\alpha_k+\beta_k)^2} \le 4C + 2. \tag{4.5}$$

Let

$$\phi(\alpha_k, \beta_k) := 2\alpha_k - (4C+3)\alpha_k^2 - (4C+1)\beta_k^2 - 2(4C+2)\alpha_k\beta_k \le 0.$$

We have

$$\frac{\partial \phi}{\partial \beta_k} = -2(4C+1)\beta_k - 2(4C+2)\alpha_k \le 0.$$

Thus, the maximum of ϕ is on the line $\alpha_k + \beta_k = q_k/C$ and the upper boundary of B. If $\alpha_k + \beta_k = q_k/C$, then we have

$$\frac{2\alpha_k - \alpha_k^2 + \beta_k^2}{(\alpha_k+\beta_k)^2} = \frac{2(\frac{q_k}{C} - \beta_k) + \frac{q_k}{C}(2\beta_k - \frac{q_k}{C})}{(q_k/C)^2}.$$

The partial derivative of the above term with respect to β_k is non-positive. Therefore, the maximum lies on the intersection of $\alpha_k + \beta_k = q_k/C$ and the upper boundary of B. Hence, it suffices to show (4.5) on the upper boundary of B, which is given by

$$\beta = \frac{-q + \sqrt{q^2 + (1-q)^2(2\alpha - \alpha^2) + 2q(1-q)\alpha}}{1-q}.$$

Let $\varphi(\alpha, q) := \sqrt{q^2 + (1-q)^2(2\alpha - \alpha^2) + 2q(1-q)\alpha}$. We have

$$\begin{aligned}\phi(\alpha_k, \beta_k) =& -(2q_k^2 + (1-q_k)^2(2\alpha_k - \alpha_k^2) + 2q_k(1-q_k)\alpha_k - 2q_k\varphi(\alpha_k, q_k))/(1-q_k)^2 \\ & + 2\alpha_k - (4C+3)\alpha_k^2 - (4C)\beta_k^2 - 2(4C+2)\alpha_k\beta_k \\ =& \frac{2q_k\beta_k}{1-q_k} - \frac{2q_k\alpha_k}{1-q_k} - (4C+2)\alpha_k^2 - (4C)\beta_k^2 - 2(4C+2)\alpha_k\beta_k.\end{aligned}$$

Since $q_k \leq C(\alpha_k + \beta_k)$, $\phi(\alpha_k, \beta_k)$ is non-positive. This proves (4.5). Moreover, we have $(1-q_k)/(1+q_k) \leq 1$, which combined with (4.5), gives

$$\frac{1-q_k}{1+q_k}\frac{2\alpha_k - \alpha_k^2 + \beta_k^2}{(\alpha_k + \beta_k)^2} \leq 4C + 2.$$

Thus, we have

$$\frac{L_{k+2}}{L_k^2} \leq 6C + 2.$$

Case II. We now show that

$$\frac{L_{k+2}}{L_k^2} \leq 6C + 2$$

for the case where $(\alpha_k, \beta_k, q_k) \in B$, From Proposition 4.4.1 we have

$$L_{k+2}^{(q_k)} \geq L_{k+2},$$

where $L_{k+2}^{(q_k)}$ denotes the total error probability if we use q_k to calculate L_{k+2} from L_{k+1}. Therefore, it suffices to prove that

$$L_{k+2}^{(q_k)} - (6C+2)L_k^2 = \alpha_{k+2} + \beta_{k+2} - (6C+2)(\alpha_k + \beta_k)^2 \leq 0.$$

We have

$$\beta_{k+1} = \frac{1-q_k}{1+q_k}\beta_k^2 + \frac{2q_k}{1+q_k}\beta_k.$$

Since $q_k \leq CL_k$, we have $\beta_k \geq q_k/(2C)$ and

$$\frac{\partial \beta_{k+1}}{\partial \beta_k} = \frac{2(1-q_k)}{1+q_k}\beta_k + \frac{2q_k}{1+q_k} \leq (6C+2)\beta_k.$$

From recursion (4.1), we have

$$\begin{aligned}\beta_{k+2} &= \frac{1-q_k}{1+q_k}(2\beta_{k+1} - \beta_{k+1}^2) + \frac{2q_k}{1+q_k}\beta_{k+1} \\ &= -\frac{1-q_k}{1+q_k}\beta_{k+1}^2 + \frac{2}{1+q_k}\beta_{k+1}.\end{aligned}$$

Therefore,

$$\frac{\partial \beta_{k+2}}{\partial \beta_k} = -2\frac{1-q_k}{1+q_k}\beta_{k+1}\frac{\partial \beta_{k+1}}{\partial \beta_k} + \frac{2}{1+q_k}\frac{\partial \beta_{k+1}}{\partial \beta_k} \leq 2(6C+2)\beta_k.$$

Consequently,

$$\frac{\partial\left(L_{k+2}^{(q_k)} - (6C+2)L_k^2\right)}{\partial \beta_k} \leq 2(6C+2)\beta_k - 2(6C+2)\alpha_k - 2(6C+2)\beta_k \leq 0.$$

We can consider the line $\alpha_k + \beta_k = q_k/C$ and the lower boundary of B, which is given by $\beta_k = \alpha_k$. With a similar argument, the maximum can be shown to lies on the intersection of $\alpha_k + \beta_k = q_k/C$ and the lower boundary of B. Moreover, we know that if $\beta_k = \alpha_k$, then $L_{k+1} = L_k$ and $(\alpha_{k+1}, \beta_{k+1})$ lies on the lower boundary of $R_{\mathcal{L}}$. Following a similar argument to Case I, we arrive at

$$\frac{L_{k+2}}{L_k^2} = \frac{L_{k+2}}{L_{k+1}^2} \leq 6C+2.$$

□

Note that if $C = 0$, then $q_k = 0$ for all k and the problem reduces to the non-failure case, where the ratio L_{k+2}/L_k^2 is bounded above by 2 (see Proposition 3.3.5). Figs. 4.5(a) and (b) show the behavior of L_{k+2}/L_k^2 in the regions B and $R_{\mathcal{U}} \setminus B$ for the case where $C = 1$; i.e., $q_k \leq L_k$. This example provides a visualization of the two-step reduction of the system.

Proposition 4.4.2 establishes bounds on the reduction in the total error probability for every two steps. From these, we can derive bounds for $\log P_N^{-1}$ for even-height trees; i.e., $\log N$ is even.

Theorem 4.4.2. *Suppose that* $(\alpha_0, \beta_0, q_0) \in R$ *and* q_k *is monotone non-increasing. If* $q_k \leq CL_k$ *where* C *is a positive constant and* $k = 0, 1, \ldots, \log N - 1$*, then for the case where* $\log N$ *is even,*

$$\sqrt{N}\left(\log L_0^{-1} - \log(6C+2)\right) \leq \log P_N^{-1} \leq \sqrt{N}\left(\log L_0^{-1} + 1\right).$$

Proof. If $(\alpha_0, \beta_0, q_0) \in R$ and q_k is monotone non-increasing, then we have $(\alpha_k, \beta_k, q_k) \in R$ for $k = 0, 1, \ldots, \log N - 2$. From Proposition 4.4.2, we have

$$L_{k+2} = a_k L_k^2,$$

for $k = 0, 1, \ldots, \log N - 2$ and some $a_k \in [1/2, 6C+2]$. Hence, for $k = 2, 4, \ldots, \log N$, we have

$$L_k = a_{(k-2)/2} \cdot a_{(k-4)/2}^2 \cdots a_0^{2^{(k-2)/2}} L_0^{2^{k/2}},$$

where $a_i \in [1/2, 6C+2]$, $i = 0, 1, \ldots, (k-2)/2$. Taking logs and using $k = \log N$, we have

$$\log P_N^{-1} = -\log a_{(k-2)/2} - 2\log a_{(k-4)/2} - \ldots - 2^{(k-2)/2}\log a_0 + \sqrt{N}\log L_0^{-1}.$$

Notice that $\log L_0^{-1} > 0$ and $-1 \leq \log a_i \leq \log(6C+2)$ for all i. Thus,

$$\begin{aligned}\log P_N^{-1} &\leq \sqrt{N}\log L_0^{-1} + \sqrt{N} \\ &= \sqrt{N}\left(\log L_0^{-1} + 1\right).\end{aligned}$$

Finally,

$$\log P_N^{-1} \geq -\log(6C+2)\sqrt{N} + \sqrt{N}\log L_0^{-1}$$
$$= \sqrt{N}\left(\log L_0^{-1} - \log(6C+2)\right).$$

$\square$

For odd-height trees, we need to calculate the reduction in the total error probability associated with a single step. For this, we have the following proposition.

Proposition 4.4.3. *If $(\alpha_k, \beta_k, q_k) \in \mathcal{U}$, then we have*

$$\frac{L_{k+1}}{L_k^2} \geq 1 \text{ and } \frac{L_{k+1}}{L_k} \leq 1.$$

From Propositions 4.4.2 and 4.4.3, we give bounds for the total error probability at the fusion center for trees with odd height.

Theorem 4.4.3. *Suppose that $(\alpha_0, \beta_0, q_0) \in R$ and q_k is monotone non-increasing. If $q_k \leq CL_k$ where C is a positive constant and $k = 0, 1, \ldots, \log N - 1$, then for the case where $\log N$ is odd,*

$$\sqrt{\frac{N}{2}}\left(\log L_0^{-1} - \log(6C+2)\right) \leq \log P_N^{-1} \leq \sqrt{2N}\left(\log L_0^{-1} + 1\right).$$

Proof. From Proposition 4.4.3, we have

$$L_1 = \widetilde{a} L_0^2$$

for some $\widetilde{a} \geq 1$. And, by Proposition 4.4.2, the following identity holds.

$$L_{k+2} = a_k L_k^2$$

for $k = 1, 3, \ldots, \log N - 2$ and some $a_k \in [1/2, 6C+2]$. Hence, we can write

$$L_k = \widetilde{a}^{2^{(k-1)/2}} \cdot a_{(k-1)/2} \cdot a_{(k-3)/2}^2 \cdots a_1^{2^{(k-3)/2}} L_0^{2^{(k+1)/2}},$$

where $a_i \in [1/2, 6C+2]$ for each $i = 1, 3, \ldots, (k-1)/2$, and $\widetilde{a} \geq 1$. Taking logs and using $k = \log N$, we have

$$\begin{aligned} \log P_N^{-1} = &- 2^{(k-1)/2} \log \widetilde{a} - \log a_{(k-1)/2} - \cdots \\ &- 2^{(k-3)/2} \log a_1 + \sqrt{2N} \log L_0^{-1}. \end{aligned}$$

Notice that $\log L_0^{-1} > 0$ and $\log a_i \geq -1$ for all i. Moreover, $\log \widetilde{a} \geq 0$. Hence,

$$\log P_N^{-1} \leq \sqrt{2N}(\log L_0^{-1} + 1).$$

We now establish the lower bound. It follows from Proposition 4.4.3 that

$$L_k = \widetilde{a} L_{k-1}$$

for some $\widetilde{a} \in (0, 1]$. By Proposition 4.4.2, we have

$$L_{k+2} = a_k L_k^2$$

for $k = 0, 2, \ldots, \log N - 3$ and some $a_k \in [1/2, 6C+2]$. Thus,

$$L_k = \widetilde{a} \cdot a_{(k-3)/2} \cdot a_{(k-3)/2}^2 \cdots a_0^{2^{(k-3)/2}} L_0^{2^{(k-1)/2}},$$

where $a_i \in [1/2, 6C+2]$ for each $i = 0, 2, \ldots, (k-3)/2$, and $\widetilde{a} \in (0, 1]$. Hence,

$$\begin{aligned} \log P_N^{-1} = &- \log \widetilde{a} - \log a_{(k-1)/2} - \cdots \\ &- 2^{(k-3)/2} \log a_1 + \sqrt{\frac{N}{2}} \log L_0^{-1}. \end{aligned}$$

Notice that $\log L_0^{-1} > 0$ and $-1 \leq \log a_i \leq \log(6C+2)$ for all i, and $\log \widetilde{a} \leq 0$. Thus,

$$\begin{aligned} \log P_N^{-1} &\geq - \log(6C+2)\sqrt{\frac{N}{2}} + \sqrt{\frac{N}{2}} \log L_0^{-1} \\ &= \sqrt{\frac{N}{2}} \left(\log L_0^{-1} - \log(6C+2) \right). \end{aligned}$$

This completes the proof. $\square$

Theorems 4.4.2 and 4.4.3, respectively, establish upper and lower bounds for $\log P_N^{-1}$ for trees with even and odd heights, for the case where hypotheses H_0 and H_1 are equally likely. For the case where the prior probabilities are not equal; i.e., $\pi_0 \neq \pi_1$, we can derive bounds for the total error probability in a similar fashion. Suppose that the fusion rule is as before; i.e., the likelihood ratio test with unit-threshold. The total error probability at the fusion center is $\hat{P}_N = \pi_0 \alpha_{\log N} \pi_1 \beta_{\log N}$. Without loss of generality, we assume that $\pi_0 \leq \pi_1$. We are interested in bounds for $\log \hat{P}_N^{-1}$.

Theorem 4.4.4. *Suppose that $(\alpha_0, \beta_0, q_0) \in R$ and q_k is monotone non-increasing. If $q_k \leq CL_k$ where C is a positive constant and $k = 0, 1, \ldots, \log N - 1$, then for the case where $\log N$ is even, we have*

$$\begin{aligned}\sqrt{N}(\log L_0^{-1} - \log(6C+2)) + \log \pi_1^{-1} &\leq \log \hat{P}_N^{-1} \\ &\leq \sqrt{N}(\log L_0^{-1} + 1) + \log \pi_0^{-1}.\end{aligned}$$

For the case where $\log N$ is odd, we have

$$\begin{aligned}\sqrt{\frac{N}{2}}(\log L_0^{-1} - \log(6C+2)) + \log \pi_1^{-1} &\leq \log \hat{P}_N^{-1} \\ &\leq \sqrt{2N}(\log L_0^{-1} + 1) + \log \pi_0^{-1}.\end{aligned}$$

Proof. First we consider the even-height tree case. Recall that $P_N = L_{\log N} = \alpha_{\log N} + \beta_{\log N}$. We have

$$\pi_0 P_N \leq \hat{P}_N = \pi_0 \alpha_k + \pi_1 \beta_k \leq \pi_1 P_N.$$

From the upper and lower bounds for $\log P_N^{-1}$ derived in Theorem 4.4.2, we can get the upper and lower bounds for $\log \hat{P}_N^{-1}$:

$$\begin{aligned}\log \hat{P}_N^{-1} &\geq \log \pi_1^{-1} + \log P_N^{-1} \\ &\geq \log \pi_1^{-1} + \sqrt{N}(\log L_0^{-1} - \log(6C+2))\end{aligned}$$

and

$$\begin{aligned}\log \hat{P}_N^{-1} &\leq \log \pi_0^{-1} + \log P_N^{-1} \\ &\leq \log \pi_0^{-1} + \sqrt{N}(\log L_0^{-1} + 1).\end{aligned}$$

For the odd-height tree case, we can mimic the proof using the bounds in Theorem 4.4.4. The details are omitted.

□

We now discuss the asymptotic decay rates. The system enters the invariant region R eventually if the height of the tree is sufficiently large. Therefore to consider the asymptotic decay rate, it suffices just to consider the decay rate when the system lies in R. In addition, the bounds in Theorems 4.4.2–4.4.4 only differ by constant terms, and so it suffices to consider only the asymptotic decay rate for trees with even height in the equal prior probability case. Moreover, when we consider the asymptotic regime; i.e., $N \to \infty$, the sufficient condition in Theorems 4.4.2–4.4.4; i.e., $q_k \leq CL_k$, can be written as $q_k = O(L_k)$. We have the following result.

Corollary 4.4.2. *Suppose that* $(\alpha_0, \beta_0, q_0) \in R$ *and* q_k *is monotone non-increasing. If* $q_k = O(L_k)$*, then the asymptotic decay rate is*

$$\log P_N^{-1} = \Theta(\sqrt{N}).$$

This implies that the decay of the total error probability is sub-exponential with exponent $\sqrt{N}$. Thus, compared to the non-failure case, the scaling law of the asymptotic decay rate does not change when we have node and link failures in the tree, provided that the probabilities of silence q_k decay to 0 sufficiently fast.

4.4.3 Discussion on the Sufficient Condition

We have shown that if $q_k = O(L_k)$, then the scaling law for the asymptotic decay rate remains the same as that of the non-failure case discussed in Chapter 3. Notice that the silence probability sequence $\{q_k\}$ depends on the local failure probabilities $\{p_k\}$, which we regard as an exogenous input. Next we consider how the decay rate of p_k determines the decay rate of q_k. Recall that the recursion of q_k is

$$q_{k+1} = q_k^2 + (1 - q_k^2)p_{k+1}.$$

Since q_k is non-increasing, the first term q_k^2 decays at least quadratically fast to 0 and $(1 - q_k^2) \nearrow 1$ in the second term. Therefore, if p_k decays more slowly than quadratically, then the value of q_k linearly depends on p_k.

Proposition 4.4.4. *Suppose that $\{p_k\}$ is monotone non-increasing. Then, the decay rate of the total error probability remains $\sqrt{N}$, i.e., $\log P_N^{-1} = \Theta(\sqrt{N})$, if and only if the decay rate of p_k is not smaller than $2^{k/2}$, i.e., $\log p_k^{-1} = \Omega(2^{k/2})$.*

Proof. By Corollary 4.4.1, we have $\log P_N^{-1} = O(\sqrt{N})$. This together with monotonicity of P_N imply that $\log P_N^{-1}$ is either $\Theta(\sqrt{N})$ or $o(\sqrt{N})$.

First we show that if $\log p_k^{-1} = \Omega(2^{k/2})$, then $\log P_N^{-1} = \Theta(\sqrt{N})$. From Corollary 4.4.1, we know that the decay rate of the total error probability is not better than $\sqrt{N}$, that is, $\log P_N^{-1} = O(\sqrt{N})$. We divide our proof into three cases based on the decay rate of p_k. If $\log p_k^{-1} = \Omega(2^k)$, that is, if p_k decays at least exponentially fast with respect to 2^k, then we can easily show that $q_k = O(L_k)$. If p_k decays more slowly than the above rate and $\log p_k^{-1} = \omega(2^{k/2})$, then for sufficiently large k we have

$$q_{k+1} = q_k^2 + (1 - q_k^2)p_{k+1} \leq 2p_{k+1}.$$

In consequence, q_k decays faster than the sequence $2p_k$ and therefore it decays faster than L_k, that is., $q_k = O(L_k)$, in which case by Corollary 4.4.2, the decay rate of the total error probability at the fusion center remains $\sqrt{N}$. In the case where $\log p_k^{-1} = \Theta(2^{k/2})$, we prove the claim by contradiction. We assume that $\log P_N^{-1} = o(\sqrt{N})$. Therefore, we can write $L_k = P_{\log N} > 2^{-c2^{k/2}}$ for all $c > 0$. Moreover, there exists c_1 such that $q_k \leq 2p_k \leq 2^{-c_1 2^{k/2}}$. In this case the ratio L_{k+2}/L_k^2 is upper bounded:

$$\begin{aligned}
\frac{L_{k+2}}{L_k^2} &\leq 1 + \frac{q_k}{(1+q_k)L_k^2} + \frac{q_{k+1}}{(1+q_{k+1})L_k^2} + \frac{q_k}{(1+q_k)}\frac{q_{k+1}}{(1+q_{k+1})}L_k^{-2} \\
&< 1 + \frac{2^{-c_1 2^{k/2}} + 2^{-c_1 2^{(k+1)/2}} + 2^{-c_1 2^{k/2}} 2^{-c_1 2^{(k+1)/2}}}{L_k^2} \\
&< 1 + 3\frac{2^{-c_1 2^{k/2}}}{L_k^2}.
\end{aligned}$$

Because $L_k > 2^{-c2^{k/2}}$ for all $c > 0$, we have $L_{k+2}/L_k^2 < 4$. Using the same analysis as that of Theorem 4.4.2, we can show that $\log P_N^{-1} = \Theta(\sqrt{N})$, which contradicts with the assumption. Hence, we conclude that if $\log p_k^{-1} = \Omega(2^{k/2})$, then the decay rate of the total error probability remains $\sqrt{N}$, i.e., $\log P_N^{-1} = \Theta(\sqrt{N})$.

Next we show that if $\log p_k^{-1} = o(2^{k/2})$, then $\log P_N^{-1} = o(\sqrt{N})$. This claim is also proved by contradiction. Suppose that the local failure probability does not decay sufficiently fast, more precisely, $\log p_k^{-1} = o(2^{k/2})$ and the decay rate of the total error probability remains $\sqrt{N}$. For sufficiently large k we have

$$q_{k+1} = q_k^2 + (1 - q_k^2)p_{k+1} \geq p_{k+1}/2.$$

Therefore we can write $q_k > 2^{-c2^{k/2}}$ for all $c > 0$, in which case the ratio L_{k+2}/L_k^2 is lower bounded:

$$\begin{aligned}
\frac{L_{k+2}}{L_k^2} &\geq \frac{q_k}{(1+q_k)}\frac{q_{k+1}}{(1+q_{k+1})}(\alpha_k + \beta_k)^{-1} \\
&> \frac{2^{-c_1 2^{k/2}} 2^{-c_2 2^{(k+1)/2}}}{4(\alpha_k + \beta_k)} \\
&= \frac{2^{-2^{k/2}(c_1 + \sqrt{2}c_2)}}{4(\alpha_k + \beta_k)},
\end{aligned} \tag{4.6}$$

for all positive c_1 and c_2. However, from the assumption that $\log P_N^{-1} = \Theta(\sqrt{N})$, we have $L_k \leq 2^{-c_3 2^{k/2}}$ for sufficiently large k, where c_3 is a positive constant. In consequence, we have shown the ratio (4.6) is not bounded above and $L_{k+2}/L_k^2 \to \infty$. Therefore, the decay rate of the total error probability cannot remain $\sqrt{N}$ and this rate is dominated by that of the non-failure case, i.e., $\log P_N^{-1} = o(\sqrt{N})$.

$\square$

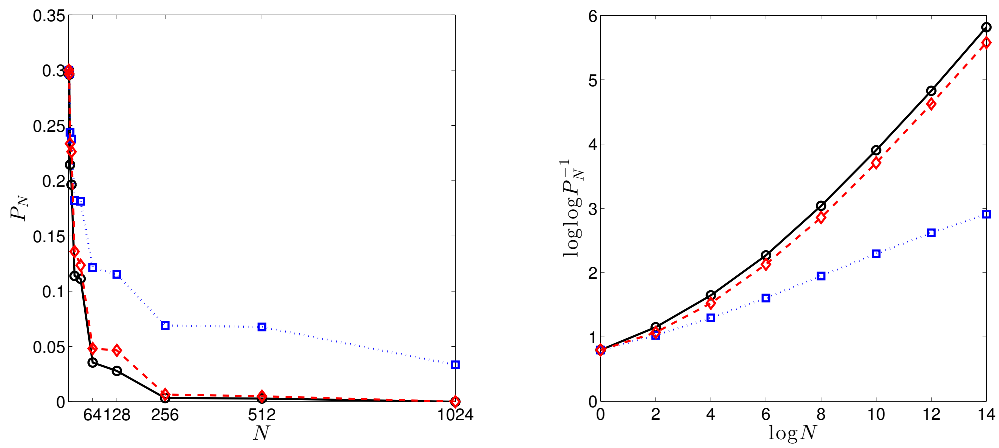
The above proposition tells us that the decay exponent of the total error probability remains $\sqrt{N}$ if and only if the local failure probability decays to 0 sufficiently fast. For illustration purposes, in Figs. 4.6(a) and (b) we plot the total error probability P_N versus the number N of sensors and $\log \log P_N^{-1}$ versus $\log N$, respectively. We set the prior probability $P(H_0) = 0.4$ and the local failure probability $p_0 = 0.1$. As shown in Figs. 4.6(a) and (b), the solid (black) lines represent the total error probability curves in the non-failure case. The dashed (red) lines represent the total error probability curves in the failure case where the local failure probabilities decay quadratically, i.e., $p_{k+1} = p_k^2$. This corresponds to a special case where $q_k < L_k$ for sufficiently large k, for which the decay rate remains $\sqrt{N}$. The dotted (blue) lines represent the total error probability curves in the failure case where the local failure probabilities are identical, i.e., $p_{k+1} = p_k$. This corresponds to a case where $q_k \geq 0.05$ for all k, for which the decay rate is strictly smaller than $\sqrt{N}$. The plots are illustrative of the differences in decay rates as reflected by our analytical results.

In the non-failure case and the quadratically decaying case described above, we have $\log P_N^{-1} = \Theta(\sqrt{N})$, which means that there exist positive constants c_1 and c_2 such that $c_1\sqrt{N} \leq \log P_N^{-1} \leq c_2\sqrt{N}$. Therefore, we have

$$\log c_1 + \frac{1}{2}\log N \leq \log\log P_N^{-1} \leq \log c_2 + \frac{1}{2}\log N.$$

Notice that in Fig. 4.6(b) for sufficiently large $\log N$ (> 8), the slopes for the non-failure case and the quadratically decaying case are approximately 1/2, consistent with the bounds above.

We have studied the detection performance of balanced binary relay trees with node and link failures. We have shown that the decay rate of the total error probability is $O(\sqrt{N})$,



(a) (b)

Figure 4.6: (a) Total error probability P_N versus the number N of sensors. (b) Plot of $\log\log P_N^{-1}$ versus $\log N$. Solid (black) lines represent the non-failure case. Dashed (red) lines represent the case where the local failure probabilities decay quadratically, i.e., $p_{k+1} = p_k^2$. Dotted (blue) lines represent the case where the local failure probabilities are identical, i.e., $p_{k+1} = p_k$.

which cannot be faster than that of the non-failure case. We have also derived upper and lower bounds for the total error probability at the fusion center as functions of N in the case where the silence probabilities decay to 0 sufficiently fast. These bounds imply that the total error probability converges to 0 sub-exponentially with exponent $\sqrt{N}$. Compared to balanced binary relay trees with no failures, the step-wise shrinkage of the total error probability in the failure case is slower, but the scaling law of the asymptotic decay rate remains the same. By contrast, if the silence probabilities do not decay to 0 sufficiently fast, then the decay rate in the failure case is strictly smaller than that in the non-failure case.

CHAPTER 5

M-ARY RELAY TREES AND NON-BINARY MESSAGE ALPHABETS

5.1 Problem Formulation

We consider the problem of binary hypothesis testing in the context of M-ary relay tree shown in Fig. 5.1, in which leaf agents (circles) are agents making independent measurements of the underlying true hypothesis. Only these leaves have direct access to the measurements in the tree structure. These leaf agents then make binary decisions based on their measurements and forward their decisions (messages) to their parent agents at the next level. Each nonleaf agent, with the exception of the root, is a *relay* agent (diamond), which aggregates M binary messages received from its child agents into one new binary message and forwards it to its parent agent again. This process takes place at each agent, culminating at the root (rectangle) where the final decision is made between the two hypotheses based on the messages received. We denote the number of leaf agents by N, which also represents the number of measurements. The height of the tree is $\log_M N$, which grows unboundedly as the number of leaf agents goes to infinity.

We assume that the decisions at all the leaf agents are independent given each hypothesis, and that they have identical Type I error probability (also known as false alarm probability, denoted by α_0) and identical Type II error probability (also known as missed detection probability, denoted by β_0). In this chapter, we answer the following questions about the Type I and Type II error probabilities:

- How do they change as we move upward in the tree?
- What are their explicit forms as functions of N?
- Do they converge to 0 at the root?

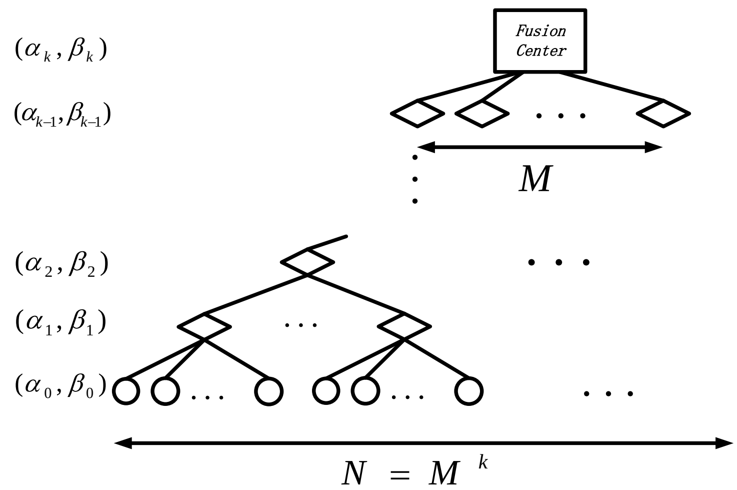


Figure 5.1: An M-ary relay tree with height k. Circles represent leaf agents making direct measurements. Diamonds represent relay agents which fuse M binary messages. The rectangle at the root makes an overall decision.

- If yes, how fast will they converge with respect to N?

For each nonleaf agent, we consider two ways of aggregating M binary messages:

- In the first case, each nonleaf agent simply aggregates M binary messages into a new binary decision using the majority dominance rule (with random tie-breaking), which is a typical non-Bayesian fusion rule. This way of aggregating information is common in daily life (e.g., voting). For this fusion rule, we provide explicit recursions for the Type I and Type II error probabilities as we move towards the root. We derive bounds for the Type I, Type II, and total error probabilities at the root as explicit functions of N, which in turn characterize the convergence rates.

- In the second case, each nonleaf agent knows the error probabilities associated with the binary messages received and it aggregates M binary messages into a new binary decision using the Bayesian likelihood ratio test, which is locally optimal in the sense that the total error probability after fusion is minimized. We derive an upper bound for the total error probability, which shows that the convergence speed of the total error probability using this fusion rule is at least as fast as that using the majority dominance rule.

5.2 Majority Dominance

In this section, we consider the case where each nonleaf agent uses the majority dominance rule. We derive explicit upper and lower bounds for the Type I, Type II, and total error probabilities with respect to N. Then, we use these bounds to characterize the asymptotic convergence rates.

5.2.1 Error Probability Bounds

We divide our analysis into two cases: oddary tree (M odd) and evenary tree (M even). In each case, we first derive the recursions for the Type I and Type II error probabilities and show that all agents at level k have the same error probability pair (α_k, β_k). Then, we study the step-wise reduction of each kind of error probability. From these we derive upper and lower bounds for the Type I, Type II, and the total error probability at the root.

5.2.1.1 Oddary tree

We first study the case where the degree of branching M is an odd integer. Consider an agent at level k, which aggregates M binary messages $\mathbf{u}_i^{k-1} = \{u_1^{k-1}, u_2^{k-1}, \ldots, u_M^{k-1}\}$ from its child agents at level $k-1$, where $u_t^{k-1} \in \{0,1\}$ for all t. Suppose that u_o^k is the output binary message after fusion, which is again sent to the parent agent at the next level. The majority dominance rule, when M is odd, is simply

$$u_o^k := \begin{cases} 1, & \text{if } \sum_{t=1}^M u_t^{k-1} \geq M/2, \\ 0, & \text{if } \sum_{t=1}^M u_t^{k-1} \leq M/2. \end{cases}$$

Suppose that the binary messages $\{u_t^{k-1}\}_{t=1}^M$ have identical Type I error probability α and identical Type II error probability β. Then, the Type I and Type II error probability pair (α', β')

associated with the output binary message u_o^k is given by:

$$\begin{aligned}
\alpha' &= \mathbb{P}_0(u_o^k = 1) \\
&= \prod_{t=1}^{M} \mathbb{P}_0(u_t^{k-1} = 1) + \binom{M}{1} \mathbb{P}_0(u_s^{k-1} = 0) \prod_{t=1}^{M-1} \mathbb{P}_0(u_t^{k-1} = 1) + \dots \\
&\quad + \binom{M}{(M-1)/2} \prod_{s=1}^{(M-1)/2} \mathbb{P}_0(u_s^{k-1} = 0) \prod_{t=1}^{(M+1)/2} \mathbb{P}_0(u_t^{k-1} = 1) \\
&= f(\alpha),
\end{aligned}$$

where $f(\alpha) := \alpha^M + \binom{M}{1}\alpha^{M-1}(1-\alpha) + \dots + \binom{M}{(M-1)/2}\alpha^{(M+1)/2}(1-\alpha)^{(M-1)/2}$ and

$$\begin{aligned}
\beta' &= \mathbb{P}_1(u_o^k = 0) \\
&= \prod_{t=1}^{M} \mathbb{P}_1(u_t^{k-1} = 0) + \binom{M}{1} \mathbb{P}_1(u_s^{k-1} = 1) \prod_{t=1}^{M-1} \mathbb{P}_1(u_t^{k-1} = 0) + \dots \\
&\quad + \binom{M}{(M-1)/2} \prod_{s=1}^{(M+1)/2} \mathbb{P}_1(u_s^{k-1} = 1) \prod_{t=1}^{(M-1)/2} \mathbb{P}_1(u_t^{k-1} = 0) \\
&= f(\beta).
\end{aligned}$$

We assume that all the binary messages from leaf agents have the same error probability pair (α_0, β_0). Hence, all agent decisions at level 1 will have the same error probability pair after fusion: $(\alpha_1, \beta_1) = (f(\alpha_0), f(\beta_0))$. By induction, we have

$$(\alpha_{k+1}, \beta_{k+1}) = (f(\alpha_k), f(\beta_k)), \qquad k = 0, 1, \dots, \log_M N - 1,$$

where (α_k, β_k) represents the error probability pair for agents at the kth level of the tree. Note that the recursions for α_k and β_k are identical. Hence, it suffices to consider only the Type I error probability α_k in deriving the error probability bounds. Before proceeding, we provide the following lemma.

Lemma 5.2.1. *Let* $h_k^M(x) = x^k + \binom{M}{1}x^{k-1}(1-x) + \dots + \binom{M}{k}(1-x)^k$, *where* k *and* M *are integers. Suppose that* $0 < k < M$. *Then,* h_k^M *is a monotone decreasing function of* $x \in (0,1)$.

Proof. We use induction in M to prove the claim. First we note that $h_0^M(x) = 1$ for all M. Suppose that $M = 2$. Then, we have $h_1^2(x) = 2 - x$. Suppose that $M = 3$. Then, we have $h_1^3(x) = 3 - 2x$ and $h_2^3(x) = x^2 - 3x + 3$. Clearly, in these cases h_k^M are monotone decreasing functions of $x \in (0, 1)$.

Now suppose that h_k^j are monotone decreasing functions of $x \in (0, 1)$ for all $j = 2, \dots, m-1$ and $k = 1, \dots, j-1$. We wish to show that h_k^m are monotone decreasing functions of $x \in (0, 1)$ for all $k = 1, \dots, m-1$. We know that the binomial coefficients satisfy

$$\begin{aligned}\binom{m}{i} &= \binom{m-1}{i-1} + \binom{m-1}{i} \\ &= \binom{m-1}{i-1} + \binom{m-2}{i-1} + \binom{m-2}{i} = \dots \\ &= \binom{m-1}{i-1} + \binom{m-2}{i-1} + \dots + \binom{k}{i-1} + \binom{k}{i}.\end{aligned}$$

We apply the above expansion for all the coefficients in $h_k^m(x)$:

$$\begin{aligned}h_k^m(x) &= x^k + \binom{m}{1} x^{k-1}(1-x) + \dots + \binom{m}{k}(1-x)^k \\ &= x^k + \binom{k}{1} x^{k-1}(1-x) + \dots + \binom{k}{k}(1-x)^k \\ &\quad + \binom{k}{0} x^{k-1}(1-x) + \dots + \binom{k}{k-1}(1-x)^k + \dots \\ &\quad + \binom{m-1}{0} x^{k-1}(1-x) + \dots + \binom{m-1}{k-1}(1-x)^k \\ &= 1 + (1-x)h_{k-1}^k(x) + \dots + (1-x)h_{k-1}^{m-1}(x) \\ &= 1 + (1-x)\sum_{j=k}^{m-1} h_{k-1}^j(x).\end{aligned}$$

By the induction hypothesis, h_{k-1}^j are monotone decreasing for all $j = k, \dots, m-1$. Moreover, it is easy to see that h_{k-1}^j are positive for all $j = k, \dots, m-1$. Therefore, because the product of two positive monotone decreasing functions is also monotone decreasing, h_k^m is a monotone decreasing function of $x \in (0, 1)$. This completes the proof. $\square$

Next we will analyze the step-wise shrinkage of the Type I error probability after each fusion step. This analysis will in turn provide upper and lower bounds for the Type I error probability at the root.

Proposition 5.2.1. *Consider an M-ary relay tree, where M is an odd integer. Suppose that we apply the majority dominance rule as the fusion rule. Then, for all k we have*

$$1 \leq \frac{\alpha_{k+1}}{\alpha_k^{(M+1)/2}} \leq \binom{M}{(M-1)/2}.$$

Proof. Consider the ratio of α_{k+1} and $\alpha_k^{(M+1)/2}$:

$$\frac{\alpha_{k+1}}{\alpha_k^{(M+1)/2}} = \alpha_k^{(M-1)/2} + \binom{M}{1}\alpha_k^{(M-3)/2}(1-\alpha_k) + \dots \\ + \binom{M}{(M-1)/2}(1-\alpha_k)^{(M-1)/2}.$$

First, we derive the lower bound of the ratio. We know that

$$1 = (\alpha_k + 1 - \alpha_k)^{(M-1)/2} = \alpha_k^{(M-1)/2} + \binom{(M-1)/2}{1}\alpha_k^{(M-3)/2}(1-\alpha_k) \\ + \dots + \binom{(M-1)/2}{(M-1)/2}(1-\alpha_k)^{(M-1)/2}.$$

Moreover, it is easy to see that $\binom{M}{k} \geq \binom{(M-1)/2}{k}$ for all $k = 1, 2, \dots, (M-1)/2$. Consequently, we have $\alpha_{k+1}/\alpha_k^{(M+1)/2} \geq 1$. Next, we derive the upper bound of the ratio. By Lemma 5.2.1, we know that the ratio $\alpha_{k+1}/\alpha_k^{(M+1)/2}$ is monotone increasing as $\alpha_k \to 0$. Hence, we have

$$\frac{\alpha_{k+1}}{\alpha_k^{(M+1)/2}} \leq \binom{M}{(M-1)/2}.$$

□

The bounds in Proposition 5.2.1 hold for all $\alpha_k \in (0,1)$. Furthermore, the upper bound is achieved at the limit as $\alpha_k \to 0$: $\lim_{\alpha_k \to 0} \alpha_{k+1}/\alpha_k^{(M+1)/2} = \binom{M}{(M-1)/2}$. Using the above proposition, we now derive upper and lower bounds for $\log_2 \alpha_k^{-1}$.

Theorem 5.2.1. *Consider an M-ary relay tree, where M is an odd integer. Let $\lambda_M = (M+1)/2$. Suppose that we apply the majority dominance rule as the fusion rule. Then, for all k we have*

$$\lambda_M^k \left(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{\lambda_M} \right) \leq \log_2 \alpha_k^{-1} \leq \lambda_M^k \log_2 \alpha_0^{-1}.$$

Proof. From the inequalities in Proposition 5.2.1, we have $\alpha_{k+1} = c_k \alpha_k^{(M+1)/2} = c_k \alpha_k^{\lambda_M}$, where $c_k \in \left[1, \binom{M}{(M-1)/2}\right]$. From these we obtain

$$\alpha_k = c_{k-1} c_{k-2}^{\lambda_M} \dots c_0^{{\lambda_M}^{k-1}} \alpha_0^{{\lambda_M}^k},$$

where $c_i \in \left[1, \binom{M}{(M-1)/2}\right]$ for all i, and

$$\log_2 \alpha_k^{-1} = -\log_2 c_{k-1} - \lambda_M \log_2 c_{k-2} - \dots - \lambda_M^{k-1} \log_2 c_0 + \lambda_M^k \log_2 \alpha_0^{-1}.$$

Since $\log_2 c_i \in \left[0, \log_2 \binom{M}{(M-1)/2}\right]$, we have $\log_2 \alpha_k^{-1} \leq \lambda_M^k \log_2 \alpha_0^{-1}$. Moreover, we obtain

$$\begin{aligned}
\log_2 \alpha_k^{-1} \geq & -\log_2 \binom{M}{(M-1)/2} - \lambda_M \log_2 \binom{M}{(M-1)/2} - \dots \\
& - \lambda_M^{k-1} \log_2 \binom{M}{(M-1)/2} + \lambda_M^k \log_2 \alpha_0^{-1} \\
= & -\frac{\lambda_M^k - 1}{\lambda_M - 1} \log_2 \binom{M}{(M-1)/2} + \lambda_M^k \log_2 \alpha_0^{-1} \\
\geq & \lambda_M^k \left(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{(M-1)/2} \right) \\
= & \lambda_M^k \left(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{\lambda_M} \right).
\end{aligned}$$

□

The bounds for $\log_2 \beta_k^{-1}$ are similar and they are omitted for brevity. Note that our result holds for all finite integer k. In addition, our approach provides explicit bounds for both Type I and Type II error probabilities respectively. From the above results, we immediately obtain bounds at the root simply by substituting $k = \log_M N$ into the bounds in Theorem 5.2.1.

Corollary 5.2.1. *Let $P_{F,N}$ be the Type I error probability at the root of an M-ary relay tree, where M is an odd integer. Suppose that we apply the majority dominance rule as the fusion rule. Then, we have*

$$N^{\log_M \lambda_M}\left(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{\lambda_M}\right) \leq \log_2 P_{F,N}^{-1} \leq N^{\log_M \lambda_M} \log_2 \alpha_0^{-1}.$$

5.2.1.2 Evenary tree

We now study the case where M is an even integer and derive upper and lower bounds for the Type I error probabilities. The majority dominance rule in this case is

$$u_o^k := \begin{cases} 1, & \text{if } \sum_{t=1}^M u_t^{k-1} > M/2, \\ 1 \text{ w.p. } P_b, & \text{if } \sum_{t=1}^M u_t^{k-1} = M/2, \\ 0 \text{ w.p. } 1-P_b, & \text{if } \sum_{t=1}^M u_t^{k-1} = M/2, \\ 0, & \text{if } \sum_{t=1}^M u_t^{k-1} < M/2, \end{cases}$$

where $P_b \in (0,1)$ denotes the Bernoulli parameter for tie-breaking. We first assume that the tie-breaking is fifty-fifty; i.e., $P_b = 1/2$. We will show later that this assumption can be relaxed. The recursions for the Type I and Type II error probabilities are as follows:

$$\begin{aligned}
\alpha_k &= \mathbb{P}_0(u_o^k = 1) \\
&= \prod_{t=1}^{M} \mathbb{P}_0(u_t^{k-1} = 1) + \binom{M}{1}\mathbb{P}_0(u_s^{k-1} = 0) \prod_{t=1}^{M-1} \mathbb{P}_0(u_t^{k-1} = 1) + \dots \\
&+ \frac{1}{2}\binom{M}{M/2} \prod_{s=1}^{M/2} \mathbb{P}_0(u_s^{k-1} = 0) \prod_{t=1}^{M/2} \mathbb{P}_0(u_t^{k-1} = 1) \\
&= g(\alpha_{k-1}),
\end{aligned}$$

where $g(\alpha_{k-1}) := \alpha_{k-1}^M + \binom{M}{1}\alpha_{k-1}^{M-1}(1-\alpha_{k-1}) + \dots + \frac{1}{2}\binom{M}{M/2}\alpha_{k-1}^{M/2}(1-\alpha_{k-1})^{M/2}$ and

$$\begin{aligned}
\beta_k &= \mathbb{P}_1(u_o^k = 0) \\
&= \prod_{t=1}^{M} \mathbb{P}_1(u_t^{k-1} = 0) + \binom{M}{1}\mathbb{P}_1(u_s^{k-1} = 1) \prod_{t=1}^{M-1} \mathbb{P}_1(u_t^{k-1} = 0) + \dots \\
&+ \frac{1}{2}\binom{M}{M/2} \prod_{s=1}^{M/2} \mathbb{P}_1(u_s^{k-1} = 1) \prod_{t=1}^{M/2} \mathbb{P}_1(u_t^{k-1} = 0) \\
&= g(\beta_{k-1}).
\end{aligned}$$

Next we study the step-wise reduction of each type of error probability when each nonleaf agent uses the majority dominance rule. Again it suffices to consider α_k since the recursions are the same.

Proposition 5.2.2. *Consider an M-ary relay tree, where M is an even integer. Suppose that we apply the majority dominance rule as the fusion rule. Then, for all k we have*

$$1 \leq \frac{\alpha_{k+1}}{\alpha_k^{M/2}} \leq \frac{1}{2}\binom{M}{M/2}.$$

Proof. We consider the ratio of α_{k+1} and $\alpha_k^{M/2}$:

$$\frac{\alpha_{k+1}}{\alpha_k^{M/2}} = \alpha_k^{M/2} + \binom{M}{1}\alpha_k^{(M-2)/2}(1-\alpha_k) + \ldots + \frac{1}{2}\binom{M}{M/2}(1-\alpha_k)^{M/2}.$$

First, we show the lower bound of the ratio. We know that

$$\begin{aligned} 1 &= (\alpha_k + 1 - \alpha_k)^{M/2} \\ &= \alpha_k^{M/2} + \binom{M/2}{1}\alpha_k^{(M-2)/2}(1-\alpha_k) + \ldots + \binom{M/2}{M/2}(1-\alpha_k)^{M/2} \end{aligned}$$

and $\binom{M}{k} \geq \binom{M/2}{k}$ for all $k = 1, 2, \ldots, M/2$. Moreover, we have $\binom{M}{M/2}/2 \geq \binom{M/2}{M/2} = 1$. In consequence, we have $\alpha_{k+1}/\alpha_k^{M/2} \geq 1$. Notice that $\alpha_{k+1}/\alpha_k^{M/2} = h_{M/2}^M(\alpha_k)/2 + h_{M/2-1}^M(\alpha_k)/2$. By Lemma 5.2.1, the ratio is monotone increasing as $\alpha_k \to 0$. Hence, we have $\alpha_{k+1}/\alpha_k^{M/2} \leq \frac{1}{2}\binom{M}{M/2}$. $\square$

The upper bound is achieved at the limit as $\alpha_k \to 0$; i.e., $\lim_{\alpha_k \to 0} \alpha_{k+1}/\alpha_k^{M/2} = \binom{M}{M/2}/2$.

In deriving the above results, we assumed that the tie-breaking rule uses $P_b = 1/2$. Suppose now that the tie is broken with Bernoulli distribution with some arbitrary probability $P_b \in (0, 1)$. Then, it is easy to show that

$$P_b \leq \frac{\alpha_{k+1}}{\alpha_k^{M/2}} \leq 2^M.$$

The bounds above are not as tight as those in Proposition 5.2.2. However, the asymptotic convergence rates remain the same as we shall see later.

Next we derive upper and lower bounds for the Type I error probability at each level k.

Theorem 5.2.2. *Consider an M-ary relay tree, where M is an even integer. Let $\lambda_M = M/2$. Suppose that we apply the majority dominance rule as the fusion rule. Then, for all k we have*

$$\lambda_M^k \left(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{\lambda_M}\right) \leq \log_2 \alpha_k^{-1} \leq \lambda_M^k \log_2 \alpha_0^{-1}.$$

Proof. From the inequalities in Proposition 5.2.2 been derived, we have

$$\alpha_{k+1} = c_k \alpha_k^{M/2} = c_k \alpha_k^{\lambda_M},$$

where $c_k \in \left[1, \binom{M}{M/2}/2\right]$. From these we obtain

$$\alpha_k = c_{k-1} c_{k-2}^{\lambda_M} \ldots c_0^{\lambda_M{}^{k-1}} \alpha_0^{\lambda_M{}^k},$$

where $c_i \in \left[1, \binom{M}{M/2}/2\right]$ for all i, and

$$\log_2 \alpha_k^{-1} = -\log_2 c_{k-1} - \lambda_M \log_2 c_{k-2} - \ldots - \lambda_M^{k-1} \log_2 c_0 + \lambda_M^k \log_2 \alpha_0^{-1}.$$

Since $\log_2 c_i \in \left[0, \log_2 \binom{M}{M/2} - 1\right]$, we have $\log_2 \alpha_k^{-1} \leq \lambda_M^k \log_2 \alpha_0^{-1}$. Moreover, we obtain

$$\begin{aligned}
\log_2 \alpha_k^{-1} \geq &-\log_2 \binom{M}{M/2} - \lambda_M \log_2 \binom{M}{M/2} - \ldots \\
&- \lambda_M^{k-1} \log_2 \binom{M}{M/2} + \lambda_M^k \log_2 \alpha_0^{-1} \\
\geq &\ \lambda_M^k \left(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{M/2}\right).
\end{aligned}$$

$\square$

Similar to the oddary tree case, we can provide upper and lower bounds for the Type I error probability at the root.

Corollary 5.2.2. *Let $P_{F,N}$ be the Type I error probability at the root of an M-ary relay tree, where M is an even integer. Suppose that we apply the majority dominance rule as the fusion rule. Then, we have*

$$N^{\log_M \lambda_M} \left(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{\lambda_M}\right) \leq \log_2 P_{F,N}^{-1} \leq N^{\log_M \lambda_M} \log_2 \alpha_0^{-1}.$$

Remark 5.2.1. Notice that the above result is only useful when $M \geq 4$. For the case where $M = 2$ (balanced binary relay trees), we have $\alpha_{k+1} = \alpha_k^2 + \alpha_k(1-\alpha_k) = \alpha_k$ and $\beta_{k+1} = \beta_k^2 + \beta_k(1-\beta_k) = \beta_k$; that is, the Type I and Type II error probabilities remain the same after fusing with the majority dominance rule.

Remark 5.2.2. We have provided a detail analysis in Chapter 3 of the convergence rate of the total error probability in balanced binary relay trees ($M = 2$) using the unit-threshold likelihood ratio test at every nonleaf agent. We show explicit upper and lower bounds for the total error probability at the root as function of the number N of leaf agents, which in turn characterizes the convergence rate $\sqrt{N}$. Moreover, we show that the unit-threshold likelihood ratio test, which is locally optimal, is close-to globally optimal in terms of the reduction in the total error probability.

Remark 5.2.3. Notice that the bounds in Corollaries 5.2.1 and 5.2.2 have the same form. Therefore, the odd and even cases can be unified if we simply let $\lambda_M = \lfloor (M+1)/2 \rfloor$.

In the next section, we use the bounds above to derive upper and lower bounds for the total error probability at the root in the majority dominance rule case.

5.2.1.3 Total error probability bounds

In this section, we provide upper and lower bounds for the total error probability P_N at the root. Let π_0 and π_1 be the prior probabilities for the two underlying hypotheses. It is easy to see that $P_N = \pi_0 P_{F,N} + \pi_1 P_{M,N}$, where $P_{F,N}$ and $P_{M,N}$ correspond to the Type I and Type II error probabilities at the root. With the bounds for each type of error probability in the case where the majority dominance rule is used, we provide bounds for the total error probability as follows.

Theorem 5.2.3. *Consider an M-ary relay tree, let $\lambda_M = \lfloor (M+1)/2 \rfloor$. Suppose that we apply the majority dominance rule as the fusion rule. Then, we have*

$$
\begin{aligned}
N^{\log_M \lambda_M} \left(\log_2 \max\{\alpha_0, \beta_0\}^{-1} - \log_2 \binom{M}{\lambda_M} \right) &\leq \log_2 P_N^{-1} \\
&\leq N^{\log_M \lambda_M} (\pi_0 \log_2 \alpha_0^{-1} + \pi_1 \log_2 \beta_0^{-1}).
\end{aligned}
$$

Proof. From the definition of P_N; that is, $P_N = \pi_0 P_{F,N} + \pi_1 P_{M,N}$, we have $P_N \leq \max\{P_{F,N}, P_{M,N}\}$. In addition, we know that α_k and β_k have the same recursion. Therefore, the maximum between the Type I and Type II error probabilities at the root corresponds to the maximum at the leaf agents. Hence, we have

$$N^{\log_M \lambda_M} \left(\log_2 \max\{\alpha_0, \beta_0\}^{-1} - \log_2 \binom{M}{\lambda_M} \right) \leq \log_2 P_N^{-1}.$$

By the fact that $\log_2 x^{-1}$ is a convex function, we have $\log_2 P_N^{-1} \leq (\pi_0 \log_2 P_{F,N}^{-1} + \pi_1 \log_2 P_{M,N}^{-1})$. Therefore, we have $\log_2 P_N^{-1} \leq N^{\log_M \lambda_M}(\pi_0 \log_2 \alpha_0^{-1} + \pi_1 \log_2 \beta_0^{-1})$. $\square$

These non-asymptotic results are useful. For example, if we want to know how many measurements are required such that $P_N \leq \epsilon$, the answer is simply to find the smallest N that satisfies the inequality in Theorem 5.2.3; i.e.,

$$N^{\log_M \lambda_M} \left(\log_2 \max\{\alpha_0, \beta_0\}^{-1} - \log_2 \binom{M}{\lambda_M} \right) \geq \log_2 \epsilon^{-1}.$$

Hence we have

$$N \geq \left(\frac{\log_2 \epsilon^{-1}}{\log_2 \max\{\alpha_0, \beta_0\}^{-1} - \log_2 \binom{M}{\lambda_M}} \right)^{\log_{\lambda_M} M}.$$

The growth rate for the number of measurements is $\Theta((\log_2 \epsilon^{-1})^{\log_{\lambda_M} M})$.

5.2.2 Asymptotic Convergence Rates

From Corollaries 5.2.1 and 5.2.2, we can easily derive the decay rates of the Type I and Type II error probabilities. For example, for the Type I error probability, we have the following.

Proposition 5.2.3. *Consider an M-ary relay tree, let $\lambda_M = \lfloor (M+1)/2 \rfloor$. Suppose that we apply the majority dominance rule as the fusion rule. Then, we have $\log_2 P_{F,N}^{-1} = \Theta(N^{\log_M \lambda_M})$.*

Proof. To analyze the asymptotic rate, we may assume that α_0 is sufficiently small. More specifically, we assume that $\alpha_0 < 1/\binom{M}{\lambda_M}$. In this case, the bounds in Corollaries 5.2.1 and 5.2.2 show that $\log_2 P_{F,N}^{-1} = \Theta(N^{\log_M \lambda_M})$. $\square$

Remark 5.2.4. Note that $\log_M \lambda_M$ is monotone increasing with respect to M. Moreover, as M goes to infinity, the limit of $\log_M \lambda_M$ is 1. That is to say, when M is very large, the decay is close to exponential, which is the rate for star configuration and bounded-height trees. In terms of tree structures, when M is very large, the tree becomes short, and therefore achieves similar performance to that of bounded-height trees.

Remark 5.2.5. From the fact that the Type I and Type II error probabilities follow the same recursion, it is easy to see that the Type II error probability at the root also decays to 0 with exponent $N^{\log_M \lambda_M}$.

Next, we compute the decay rate of the total error probability.

Corollary 5.2.3. *Consider an M-ary relay tree, let $\lambda_M = \lfloor (M+1)/2 \rfloor$. Suppose that we apply the majority dominance rule as the fusion rule. Then, we have $\log_2 P_N^{-1} = \Theta(N^{\log_M \lambda_M})$.*

For the total error probability at the root, we have similar arguments with that for individual error probabilities. For large M, the decay of the total error probability is close to exponential.

5.3 Bayesian Likelihood Ratio Test

In this section, we consider the case where the Bayesian likelihood ratio test is used as the fusion rule. We derive an upper bound for the total error probability, which in turn characterizes the convergence rate. We show that the convergence rate in this case is at least as fast or faster than that with the majority dominance rule.

Theorem 5.3.1. *Let $\mathbb{P}_N$ be the total error probability at the root in the case where the Bayesian likelihood ratio test is used as the fusion rule in M-ary relay trees. We have*

$$\log_2 \mathbb{P}_N^{-1} \geq N^{\log_M \lambda_M} \left(\log_2 L_0^{-1} - \log_2 \left(\frac{2\binom{M}{\lambda_M} \max(\pi_0, \pi_1)}{\min(\pi_0, \pi_1)^{\lambda_M}} \right) \right).$$

Proof. In the case where the majority dominance rule is used, from Propositions 5.2.1 and 5.2.2, it is easy to show that

$$\frac{1}{2} \leq \frac{\alpha_{k+1} + \beta_{k+1}}{\alpha_k^{\lambda_M} + \beta_k^{\lambda_M}} \leq 2\binom{M}{\lambda_M}.$$

Since x^{λ_M} is a convex function for all $M \geq 2$, we have

$$\frac{\alpha_k^{\lambda_M} + \beta_k^{\lambda_M}}{2} \geq \left(\frac{\alpha_k + \beta_k}{2}\right)^{\lambda_M},$$

which implies the following:

$$2^{-\lambda_M+1} \leq \frac{\alpha_k^{\lambda_M} + \beta_k^{\lambda_M}}{(\alpha_k + \beta_k)^{\lambda_M}} \leq 1.$$

Hence, we obtain

$$2^{-\lambda_M} \leq \frac{\alpha_{k+1} + \beta_{k+1}}{(\alpha_k + \beta_k)^{\lambda_M}} \leq 2\binom{M}{\lambda_M}.$$

From these bounds and the fact that

$$\min(\pi_0, \pi_1)(\alpha_k + \beta_k) \leq \pi_0\alpha_k + \pi_1\beta_k \leq \max(\pi_0, \pi_1)(\alpha_k + \beta_k),$$

we have

$$\frac{2^{-\lambda_M}\min(\pi_0, \pi_1)}{\max(\pi_0, \pi_1)^{\lambda_M}} \leq \frac{\pi_0\alpha_{k+1} + \pi_1\beta_{k+1}}{(\pi_0\alpha_k + \pi_1\beta_k)^{\lambda_M}} \leq \frac{2\binom{M}{\lambda_M}\max(\pi_0, \pi_1)}{\min(\pi_0, \pi_1)^{\lambda_M}}.$$

Note that $\pi_0\alpha_k + \pi_1\beta_k$ is the total error probability for agents at level k and we denote it by L_k.

The Bayesian likelihood ratio test is the optimal rule in the sense that the total error probability is minimized after fusion. Let L_k^{LRT} be the total error probability after fusing with the Bayesian likelihood ratio test. We have

$$\frac{L_{k+1}^{LRT}}{L_k^{\lambda_M}} \leq \frac{L_{k+1}}{L_k^{\lambda_M}} \leq \frac{2\binom{M}{\lambda_M}\max(\pi_0, \pi_1)}{\min(\pi_0, \pi_1)^{\lambda_M}}.$$

Using a similar approach as that used in proving Theorem 5.2.1, we can derive the following lower bound for $\log_2 \mathbb{P}_N^{-1}$:

$$\log_2 \mathbb{P}_N^{-1} \geq N^{\log_M \lambda_M}\left(\log_2 L_0^{-1} - \log_2\left(\frac{2\binom{M}{\lambda_M}\max(\pi_0, \pi_1)}{\min(\pi_0, \pi_1)^{\lambda_M}}\right)\right).$$

□

From the above bound, we immediately obtain the following.

Corollary 5.3.1. *Consider an M-ary relay tree, and let $\lambda_M = \lfloor (M+1)/2 \rfloor$. Suppose that we apply the Bayesian likelihood ratio test as the fusion rule. Then, we have $\log_2 \mathbb{P}_N^{-1} = \Omega(N^{\log_M \lambda_M})$.*

Note that in the case where the majority dominance rule is used, the convergence rate is exactly $\Theta(N^{\log_M \lambda_M})$. Therefore, the convergence rate for the Bayesian likelihood ratio test is at least as good as that for the majority dominance rule.

5.4 Asymptotic Optimality of Fusion Rules

In this section, we discuss the asymptotic optimality of the two fusion rules considered in this thesis by comparing our asymptotic convergence rates with those in [61], in which it is shown that with any combination of fusion rules, the convergence rate is upper bounded as

$$\log_2 P_N^{-1} = O(N^{\log_M \frac{M+1}{2}}). \tag{5.1}$$

5.4.1 Oddary Case

In the oddary tree case, if each nonleaf agent uses the majority dominance rule, then the upper bound in (5.1) is achieved; i.e.,

$$\log_2 P_N^{-1} = \Theta(N^{\log_M \lfloor \frac{M+1}{2} \rfloor}) = \Theta(N^{\log_M \frac{M+1}{2}}).$$

This result is also mentioned in [61]. Tay *et al.* [55] find a similar result in bounded-height trees; that is, if the degree of branching for all the agents except those at level 1 is an odd constant, then the majority dominance rule achieves the optimal exponent.

Now we consider the case where each nonleaf agent uses the Bayesian likelihood ratio test. Since the convergence rate for this fusion rule is at least as good as that for the majority dominance rule, it is evident that the Bayesian likelihood ratio test, which is only locally optimal (the total error probability after each fusion is minimized), achieves the globally optimal convergence rate. This result is also of interest in decentralized detection problems, in which the objective is usually to find the globally optimal strategy. In oddary trees, the myopically optimal Bayesian likelihood ratio test, which is relevant to social learning problems because of the selfishness of agents, is essentially globally optimal in terms of achieving the optimal exponent.

Remark 5.4.1. Suppose that each nonleaf agent uses the Bayesian likelihood ratio test and we assume that the two hypotheses are equally likely. In this case, the output message is given by the unit-threshold likelihood ratio test:

$$\frac{\prod_{t=1}^{M} \mathbb{P}_1(u_t^{k-1})}{\prod_{t=1}^{M} \mathbb{P}_0(u_t^{k-1})} \underset{H_0}{\overset{H_1}{\gtrless}} 1.$$

If the Type I and Type II error probabilities at level 0 are equal; i.e., $\alpha_0 = \beta_0$, then the unit-threshold likelihood ratio test reduces to the majority dominance rule. The bounds for the error probabilities in this case and those in the majority dominance rule case are identical.

5.4.2 Evenary Case

In the evenary tree case, our results show that with the majority dominance rule, we have

$$\log_2 P_N^{-1} = \Theta(N^{\log_M \lfloor \frac{M+1}{2} \rfloor}) = \Theta(N^{\log_M \frac{M}{2}}). \tag{5.2}$$

This characterizes the explicit convergence rate of the total error probability (c.f. [61], in which there is a gap between the upper and lower bounds for $\log_2 P_N^{-1}$). It is evident that the majority dominance rule in this evenary tree case does not achieve the upper bound in (5.1). However, the gap between the rates described in (5.1) and (5.2) becomes smaller and more negligible as the degree M of branching grows.

In the case of binary relay trees ($M = 2$), the gap is most significant because the total error probability does not change after fusion with the majority dominance rule. In contrast, we have shown in Chapter 3 that the likelihood-rate test achieves convergence rate $\sqrt{N}$. For

$M \geq 4$, we have shown that the convergence rate using the Bayesian likelihood ratio test is at least as good as that using the majority dominance rule.

Now we consider the case where the *alternative majority dominance strategy* (tie is broken alternatively for agents at consecutive levels) is used throughout the tree. In this case we have

$$\alpha_k = \alpha_{k-1}^M + \binom{M}{1}\alpha_{k-1}^{M-1}(1-\alpha_{k-1}) + \ldots + \binom{M}{M/2}\alpha_{k-1}^{M/2}(1-\alpha_{k-1})^{M/2}$$

and

$$\alpha_{k+1} = \alpha_k^M + \binom{M}{1}\alpha_k^{M-1}(1-\alpha_k) + \ldots + \binom{M}{M/2-1}\alpha_k^{M/2+1}(1-\alpha_k)^{M/2-1}$$

Using Lemma 5.2.1, it is easy to show that

$$1 \leq \frac{\alpha_k}{\alpha_{k-1}^{M/2}} \leq \binom{M}{M/2} \text{ and } 1 \leq \frac{\alpha_{k+1}}{\alpha_k^{M/2+1}} \leq \binom{M}{M/2-1}. \tag{5.3}$$

Theorem 5.4.1. *Consider an M-ary relay tree, where M is an even integer, and let $\lambda_M = M/2$. Suppose that we apply the alternative majority dominance strategy. Then, for even k we have*

$$\lambda_M^{k/2}(\lambda_M+1)^{k/2}\left(\log_2\alpha_0^{-1} - \log_2\binom{M}{\lambda_M}\right) \leq \log_2\alpha_k^{-1} \leq \lambda_M^{k/2}(\lambda_M+1)^{k/2}\log_2\alpha_0^{-1}.$$

Proof. The case where $M = 2$ is easy to show using the recursion for α_k and the proof is omitted. Now let us consider the case where $M \geq 4$. From the inequalities in (5.3), we have

$$\alpha_{k+1} = c_k\alpha_k^{\lambda_M+1} = c_k c_{k-1}^{\lambda_M}\alpha_k^{\lambda_M(\lambda_M+1)},$$

where c_{k-1} and $c_k \in \left[1, \binom{M}{M/2}\right]$. From these we obtain

$$\alpha_k = c_{k-1}c_{k-2}^{\lambda_M}c_{k-3}^{\lambda_M(\lambda_M+1)}\ldots c_0^{{\lambda_M}^{k/2}(\lambda_M+1)^{k/2-1}}\alpha_0^{{\lambda_M}^{k/2}(\lambda_M+1)^{k/2}},$$

where $c_i \in \left[1, \binom{M}{M/2}\right]$ for all i. Therefore,

$$\begin{aligned}\log_2\alpha_k^{-1} = &-\log_2 c_{k-1} - \ldots - {\lambda_M}^{k/2}(\lambda_M+1)^{k/2-1}\log_2 c_0\\ &+ {\lambda_M}^{k/2}(\lambda_M+1)^{k/2}\log_2\alpha_0^{-1}.\end{aligned}$$

Since $\log_2 c_i \in \left[0, \log_2 \binom{M}{M/2}\right]$, we have $\log_2 \alpha_k^{-1} \leq {\lambda_M}^{k/2}(\lambda_M+1)^{k/2}\log_2 \alpha_0^{-1}$. Moreover, we have $\log_2 c_i \leq \log_2 \binom{M}{M/2}$. Hence,

$$\begin{aligned}\log_2 \alpha_k^{-1} \geq &-\log_2 \binom{M}{\lambda_M}(1+\lambda_M+\lambda_M(\lambda_M+1)+\ldots+\lambda_M^{k/2}(\lambda_M+1)^{k/2-1}) \\ &+{\lambda_M}^{k/2}(\lambda_M+1)^{k/2}\log_2 \alpha_0^{-1}.\end{aligned} \tag{5.4}$$

Next we use induction to show that

$$1+\lambda_M+\lambda_M(\lambda_M+1)+\ldots+\lambda_M^{k/2}(\lambda_M+1)^{k/2-1} \leq \lambda_M^{k/2}(\lambda_M+1)^{k/2}. \tag{5.5}$$

Suppose that $k=2$. Then, we have $1+\lambda_M \leq \lambda_M(\lambda_M+1)$, which holds because $\lambda_M \geq 2$. Suppose that (5.5) holds when $k=k_0$. We wish to show that it also holds when $k=k_0+1$, in which case we have

$$\begin{aligned}&1+\lambda_M+\ldots+\lambda_M^{k_0/2}(\lambda_M+1)^{k_0/2-1}+\lambda_M^{k_0/2}(\lambda_M+1)^{k_0/2}+\lambda_M^{k_0/2+1}(\lambda_M+1)^{k_0/2} \\ &\leq 2\lambda_M^{k_0/2}(\lambda_M+1)^{k_0/2}+\lambda_M^{k_0/2+1}(\lambda_M+1)^{k_0/2} \\ &\leq 2\lambda_M^{k_0/2+1}(\lambda_M+1)^{k_0/2} \leq \lambda_M^{k_0/2+1}(\lambda_M+1)^{k_0/2+1}.\end{aligned}$$

Therefore, we have proved (5.5). Substituting this result in (5.4), we obtain the desired lower bound. □

The bounds for $\log_2 \beta_k^{-1}$ are similar and they are omitted for brevity.

Corollary 5.4.1. *Let $P_{F,N}$ be the Type I error probability at the root of an M-ary relay tree, where M is an even integer. Suppose that we apply the alternative majority dominance strategy. Then, we have*

$$N^{\log_M \sqrt{M(M+2)}/2}\left(\log_2 \alpha_0^{-1}-\log_2 \binom{M}{\lambda_M}\right) \leq \log_2 P_{F,N}^{-1} \leq N^{\log_M \sqrt{M(M+2)}/2}\log_2 \alpha_0^{-1}.$$

Corollary 5.4.2. *Let P_N be the total error probability at the root of an M-ary relay tree, where M is an even integer. Suppose that we apply the alternative majority dominance strategy. Then, we have $\log_2 P_{F,N}^{-1} = \Theta(N^{\log_M \sqrt{M(M+2)}/2})$ and $\log_2 P_N^{-1} = \Theta(N^{\log_M \sqrt{M(M+2)}/2})$.*

Note that when $M = 2$, $\log_2 P_N^{-1} = \Theta(\sqrt{N})$. Therefore, the decay rate with this strategy is identical with that using the Bayesian likelihood ratio test. This is not surprising because we show in Chapter 3 that the Bayesian likelihood ratio test is essentially either 'AND' rule or 'OR' rule depending on the values of the Type I and II error probabilities. We also show that the same rule will repeat no more than two consecutive times. Therefore, the decay rate in this case is the same as that using the alternative majority dominance strategy.

For the case where $M \geq 4$, suppose that α_0 and β_0 are sufficiently small and sufficiently close to each other. Then, it is easy to show that the Bayesian likelihood ratio test is majority dominance rule with tie-breaking given by the values of the Type I and II error probabilities. Moreover, we can show that the same tie-breaking will repeat no more than two consecutive times. In this case, the error probability decays as $\Theta(N^{\log_M \sqrt{M(M+2)}/2})$.

Recall that the upper bound for the decay rate of the total error probability with all combinations of fusion rules is $O(N^{\log_M \frac{M+1}{2}})$, which involves an arithmetic mean of $M + 2$ and M. In contrast, the decay rate using the alternative majority dominance strategy and Bayesian likelihood ratio test involves the geometric mean of $M + 2$ and M, which means that these two strategies are almost asymptotic optimal, especially when M is large.

The convergence rate of the total error probability using the alternative majority dominance strategy is better that that the random tie-breaking case. For illustration purposes, in Fig. 5.2 we plot the exponent for the decay rate of the total error probability versus the spanning factor M in these two cases. For comparison purposes, we also plot the exponent in the upper bound (5.1). We can see from Fig. 5.2 that alternative majority dominance strategy achieves a larger exponent than that of the majority dominance rule with random tie-breaking. Moreover, the gap between the exponents in the alternative majority dominance strategy case and the upper bound (5.1) is small and almost negligible.

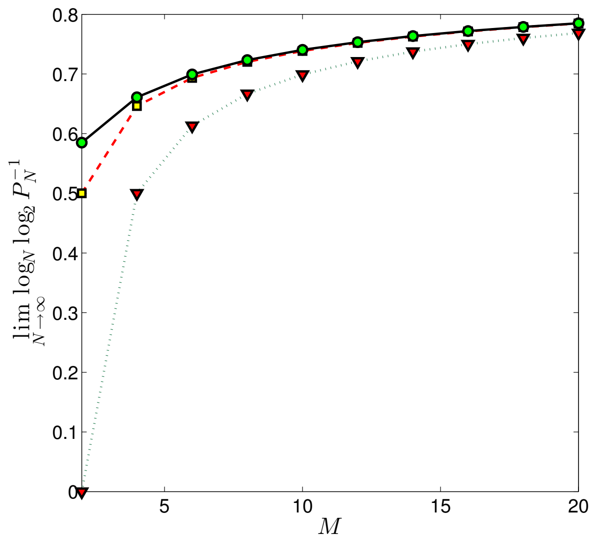


Figure 5.2: Plot of error exponents versus the spanning factor M. Dashed (red) line represents the alternative majority dominance strategy. Dotted (blue) line represents the majority dominance rule with random tie-breaking. Solid (black) line represents the upper bound on exponent in (5.1).

5.5 Non-binary Message Alphabets

In the previous sections, each agent in the tree is only allowed to pass a binary message to its supervising agent at the next level. A natural question is, what if each agent can transmit a 'richer' message? In this section, we provide a message-passing scheme that allows general message alphabet of size $\mathcal{D}$ (non-binary). We call this M-ary relay tree with message alphabet size $\mathcal{D}$ an $(M, \mathcal{D})$-tree. We have studied the convergence rates of $(M, 2)$-trees by investigating how fast the total error probability decays to 0. What about the convergence rate when $\mathcal{D}$ is an arbitrary finite integer?

We denote by u_o^k the output message for each agent at the kth level after fusing M input messages $\mathbf{u}_i^{k-1} = \{u_1^{k-1}, u_2^{k-1}, \ldots, u_M^{k-1}\}$ from its child agents at the $(k-1)$th level, where $u_t^{k-1} \in \{0, 1, \ldots, \mathcal{D}\}$ for all $t \in \{1, 2, \ldots, M\}$.

Case I: First, we consider an $(M, \mathcal{D})$-tree with height k_0, in which there are M^{k_0} leaf agents, and the message alphabet size is sufficiently large; more precisely,

$$\mathcal{D} \geq M^{k_0-1} + 1. \tag{5.6}$$

For our analysis, we need the following terminology:

Definition: Given a nonleaf agent in the tree, a *subtree leaf* of this agent is any leaf agent of the subtree rooted at the agent. An *affirmative subtree leaf* is any subtree leaf that sends a message of '1' upward.

Suppose that each leaf agent still generates a binary message $u_o^0 \in \{0, 1\}$ and sends it upward to its parent agent. Moreover, each intermediate agent simply sums up the messages it receives from its immediate child agents and sends the summation to its parent agent; that is, $u_o^k = \sum_{t=1}^{M} u_t^{k-1}$. Then we can show that the output message for each agent at the kth level is an integer from $\{0, 1, \ldots, M^k\}$ for all $k \in \{0, 1, \ldots, k_0 - 1\}$. Moreover, this message essentially represents the number of its affirmative subtree leaf.

Because of inequality (5.6), at each level k in the tree, the message alphabet size $\mathcal{D}$ is large enough to represent all possible values of u_o^k ($k \in \{0, \ldots, k_0 - 1\}$). In particular, the root (at level k_0) knows the number of its affirmative subtree leaves. In this case, the convergence rate is the same as that of the star configuration, where each leaf agent sends a binary message to the root directly. Recall that in the star configurations, the total error probability decays exponentially fast to 0.

Case II: We now consider the case where the tree height is very large; i.e., (5.6) does not hold. As shown in Fig. 5.3, we apply the scheme described in Case I; that is, the leaf agents send binary compressions of their measurements upward to their parent agents. Moreover, each intermediate agent simply sends the sum of the messages received to its parent agent; i.e.,

$$u_o^k = \sum_{t=1}^{M} u_t^{k-1}. \tag{5.7}$$

From the assumption of large tree height, it is easy to see that the message alphabet size is not large enough for all the relay agents to use the fusion rule described in (5.7). With some abuse of notation, we let k_0 to be the integer $k_0 = \lfloor \log_M(\mathcal{D}-1) \rfloor + 1$ (here, k_0 is not the height of the tree; it is strictly less than the height). Note that $M^{k_0-1}+1 \leq \mathcal{D} < M^{k_0}+1$.

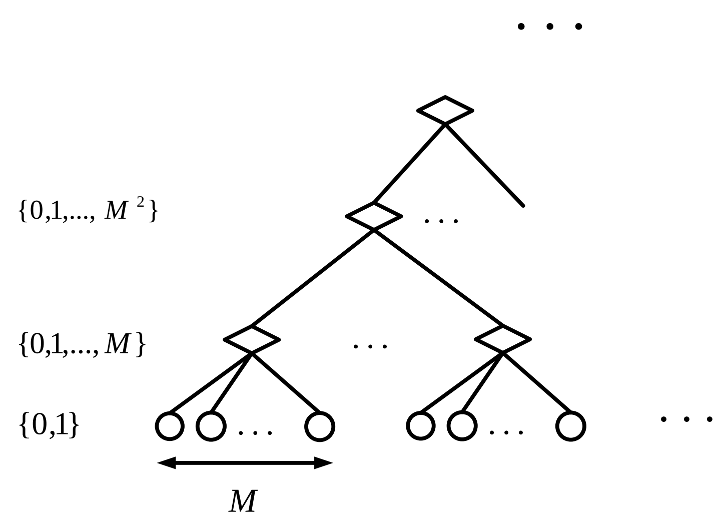


Figure 5.3: A message-passing scheme for non-binary message alphabets in an M-ary relay tree.

From the previous analysis, we can see that with this scheme, each agent at the k_0th level knows the number of its affirmative subtree leaves. Therefore, it is equivalent to consider the case where each agent at level k_0 connects to its M^{k_0} subtree leaves directly (all the intermediate agents in the subtree can be ignored). However, we cannot use the fusion rule described in (5.7) for the agents at k_0th level to generate the output messages because the message alphabet size is not large enough. Hence, we let each agent at level k_0 aggregate the M^{k_0} binary messages from its subtree leaves into a new binary message (using some fusion rule). By doing so, the output message from each agent at the k_0th level is binary again. Henceforth, we can simply apply the fusion rule (5.7) and repeat this process throughout the tree, culminating at the root. We now provide an upper bound for the asymptotic decay rate in this case.

Theorem 5.5.1. *The convergence rate of the total error probability for an* $(M, \mathcal{D})$*-tree is equal to that for an* $(M^{k_0}, 2)$*-tree, where* $k_0 = \lfloor \log_M(\mathcal{D}-1) \rfloor + 1$*. In particular, let* P_N *be the total* *error probability at the root for an* $(M, \mathcal{D})$*-tree. With any combination of fusion rules at level* ℓk_0*,* $\ell = 1, 2, \dots,$ *we have* $\log_2 P_N^{-1} = O\left(N^{\rho}\right),$ *where*

$$\rho := \frac{\ln(M^{k_0} + 1)}{\ln M^{k_0}} - \frac{\log_M 2}{k_0}.$$

Proof. Consider an $(M, \mathcal{D})$-tree with the scheme described above. It is easy to see that equivalently we can consider a tree where the leaf agents connect to the agents at the k_0th level directly. In addition, because of the recursive strategy applied throughout the tree, it suffices to consider the tree where the agents at the ℓk_0th level connect to the agents at the $(\ell + 1)k_0$th level directly for all non-negative integers ℓ. Therefore, the convergence rate of an $(M, \mathcal{D})$-tree is equal to that of the corresponding $(M^{k_0}, 2)$-tree.

In the asymptotic regime, the decay rate in $(M, 2)$-trees is bounded above as follows [61]:

$$\log_2 P_N^{-1} = O(N^{\log_M \frac{(M+1)}{2}}).$$

Therefore, the decay rate for $(M^{k_0}, 2)$-trees is also bounded above as

$$\log_2 P_N^{-1} = O(N^{\log_{M^{k_0}} \frac{(M^{k_0}+1)}{2}}),$$

which upon simplification gives the desired result. $\square$

Suppose that each agent at level ℓk_0 for all ℓ uses the majority dominance rule. Then, we can derive the convergence rate for the total error probability as follows.

Theorem 5.5.2. *Consider* $(M, \mathcal{D})$*-trees where the majority dominance rule is used. Let* $k_0 = \lfloor \log_M(\mathcal{D} - 1) \rfloor + 1$*. We have* $\log_2 P_N^{-1} = \Theta\left(N^{\varrho}\right),$ *where*

$$\varrho := \begin{cases} \frac{\ln(M^{k_0}+1)}{\ln M^{k_0}} - \frac{\log_M 2}{k_0}, & \text{if } M \text{ is odd}, \\ 1 - \frac{\log_M 2}{k_0}, & \text{if } M \text{ is even}. \end{cases}$$

Proof. By Theorem 5.5.1, the performance of $(M, \mathcal{D})$-trees is equal to that of $(M^{k_0}, 2)$-trees, where $k_0 = \lfloor \log_M(\mathcal{D} - 1) \rfloor + 1$. For the asymptotic rate, we have

$$\log_2 P_N^{-1} = \Theta(N^{\log_{M^{k_0}} \left\lfloor \frac{M^{k_0}+1}{2} \right\rfloor}),$$

which upon simplification gives the desired result. $\square$

Remark 5.5.1. Notice that $\lim_{M\to\infty} \ln(M^{k_0} + 1)/\ln M^{k_0} = 1$, which means that the even and odd cases in the expression for ϱ are similar when M is large.

Remark 5.5.2. From Theorem 5.5.1, we can see that with larger message alphabet size, the total error probability decays more quickly. However, the change in the decay exponent is not significant because k_0 depends on $\mathcal{D}$ logarithmically. Furthermore, if M is large, then the change in the performance is less sensitive to the increase in $\mathcal{D}$.

Remark 5.5.3. Comparing the results in Theorems 5.5.1 and 5.5.2, we can see that the majority dominance rule achieves the optimal exponent in the oddary case and it almost achieves the optimal exponent in the evenary case.

For the Bayesian likelihood ratio test, we have the following result.

Theorem 5.5.3. *The convergence rate using the likelihood ratio test is at least as good as that using the majority dominance rule; i.e.,* $\log_2 \mathbb{P}_N^{-1} = \Omega\left(N^{\varrho}\right)$.

In the case where M is even, we can derive the decay rate using the alternative majority dominance strategy.

Theorem 5.5.4. *The convergence rate using the alternative majority dominance strategy is* $\log_2 P_N^{-1} = \Omega\left(N^{\sigma}\right)$, *where*

$$\sigma = \frac{1}{2}\left(1 + \frac{\ln(M^{k_0} + 2)}{\ln M^{k_0}}\right) - \frac{\log_M 2}{k_0}.$$

Theorem 5.5.3 and 5.5.4 follow by applying the same arguments as those made in proofs of Corollary 5.3.1 and Theorem 5.5.1 and the proofs are omitted for brevity.

The message-passing scheme provided here requires message alphabets with maximum size $\mathcal{D}$. However, most of the agents use much ‘smaller’ messages. For example, the leaf agents generate binary messages. It is interesting to characterize the *average* message size used in our scheme. Because of the recursive strategy, it suffices to calculate the average message size in a subtree with height $k_0 - 1$ since the message sizes in our scheme repeat every k_0 levels.

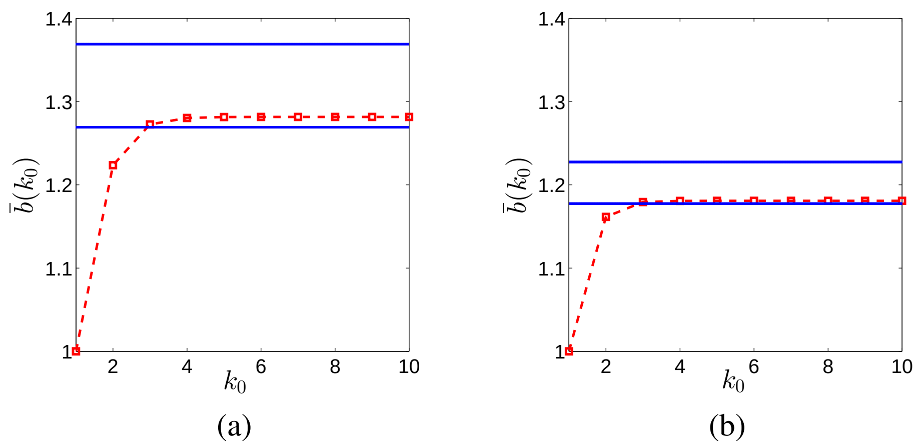


Figure 5.4: (a) Average message size (dashed red line) in $M = 10$ case. (b) Average message size (dashed red line) in $M = 20$ case. The blue lines represent the bounds in (5.8).

The message size (in bits) for agents at level $t \in \{0, 1, \dots, k_0 - 1\}$ is $\log_2(M^t + 1)$ and the number of agents at level t is $M^{k_0 - t}$. Therefore, the average size $\bar{b}(k_0)$ in bits used in our scheme is

$$\bar{b}(k_0) = \frac{M^{k_0} + \ldots + M \log_2(M^{k_0-1} + 1)}{M^{k_0} + M^{k_0-1} + \ldots + M} = \frac{\sum_{t=0}^{k_0-1} M^{k_0-t} \log_2(M^t + 1)}{\sum_{t=0}^{k_0-1} M^{t+1}}.$$

We have

$$\log_2(M^t + 1) > \log_2 M^t = t \log_2 M$$

and

$$\log_2(M^t + 1) < \log_2(2M^t) = 1 + t \log_2 M$$

for all $t \geq 1$. Therefore, the average size in bits is lower bounded as

$$\begin{aligned}
\bar{b}(k_0) >& \frac{M^{k_0} + M^{k_0-1}\log_2 M + \ldots + M(k_0-1)\log_2 M}{M^{k_0} + M^{k_0-1} + \ldots + M} \\
=& \frac{M^{k_0}}{M^{k_0} + M^{k_0-1} + \ldots + M} + \frac{\log_2 M(M^2(M^{k_0-1}-1) - M(M-1)(k_0-1))}{(M^{k_0} + M^{k_0-1} + \ldots + M)(M-1)^2} \\
=& \frac{M^{k_0} - M^{k_0-1}}{M^{k_0}-1} + \frac{M\log_2 M}{M-1}\frac{M^{k_0-1} - 1 - M(M-1)(k_0-1)}{M^{k_0}-1}.
\end{aligned}$$

In addition, it is upper bounded as

$$\bar{b}(k_0) < 1 + \frac{M\log_2 M}{M-1}\frac{M^{k_0-1} - 1 - M(M-1)(k_0-1)}{M^{k_0}-1} \leq 1 + \frac{\log_2 M}{M-1}.$$

Recall that, with sufficiently large k_0, the error probability convergence rates are close to exponential. However, from the above bounds the average message size in terms of bits in our scheme is still very small, specifically for sufficiently large k_0 we have

$$1 + \frac{\log_2 M}{M-1} - \frac{1}{M} \leq \bar{b}(k_0) \leq 1 + \frac{\log_2 M}{M-1}. \tag{5.8}$$

Fig. 5.4 shows plots of the average message sizes $\bar{b}(k_0)$ versus k_0 in the $M = 10$ and 20 cases. Note that as M increases, the average message size becomes smaller and the bounds in (5.8) become tighter.

CHAPTER 6

STRING SUBMODULARITY

We introduce the notion of string submodularity to optimal control problems. We show that the greedy strategy, consisting of a string of actions that only locally maximizes the step-wise gain in the objective function, achieves at least a $(1-e^{-1})$-approximation to the optimal strategy. Moreover, we can improve this approximation by introducing additional constraints on curvature, namely, *total backward curvature, total forward curvature*, and *elemental forward curvature*. We show that if the objective function has total backward curvature σ, then the greedy strategy achieves at least a $\frac{1}{\sigma}(1-e^{-\sigma})$-approximation of the optimal strategy. If the objective function has total forward curvature ϵ, then the greedy strategy achieves at least a $(1-\epsilon)$-approximation of the optimal strategy. Moreover, we consider a generalization of the diminishing-return property by defining the elemental forward curvature. We also introduce the notion of *string-matroid* and consider the problem of maximizing the objective function subject to a string-matroid constraint.

6.1 Introduction

6.1.1 Background

We consider the problem of optimally choosing a string of actions over a finite horizon to maximize an objective function. Let $\mathbb{A}$ be a set of all possible actions. At each stage i, we choose an action a_i from $\mathbb{A}$. We use $A = (a_1, a_2, \ldots, a_k)$ to denote a string of actions taken over k consecutive stages, where $a_i \in \mathbb{A}$ for $i = 1, 2, \ldots, k$. We use $\mathbb{A}^*$ to denote the set of all possible strings of actions (of arbitrary length, including the empty string). Let $f : \mathbb{A}^* \to \mathbb{R}$ be an objective function, where $\mathbb{R}$ denotes the real numbers. Our goal is to find a string $M \in \mathbb{A}^*$, with a length $|M|$ not larger than K, to maximize the objective function:

$$\begin{array}{l} \text{maximize } f(M) \\ \text{subject to } M \in \mathbb{A}^*, |M| \leq K. \end{array} \tag{6.1}$$

The solution to (6.1), which we call the *optimal strategy*, can be found using dynamic programming (see, e.g., [72]). More specifically, this solution can be expressed with *Bellman's equations*. However, the computational complexity of finding an optimal strategy grows exponentially with respect to the size of $\mathbb{A}$ and the length constraint K. On the other hand, the greedy strategy, though suboptimal in general, is easy to compute because at each stage, we only have to find an action to maximize the step-wise gain in the objective function. The question we are interested in is: How good is the greedy strategy compared to the optimal strategy in terms of the objective function? This question has attracted widespread interest, which we will review in the next section.

In this chapter, we extend the concept of set submodularity in combinatorial optimization to bound the performance of the greedy strategy with respect to that of the optimal strategy. Moreover, we will introduce additional constraints on curvatures, namely, total backward curvature, total forward curvature, and elemental forward curvature, to provide more refined lower bounds on the effectiveness of the greedy strategy relative to the optimal strategy. Therefore, the greedy strategy serves as a good approximation to the optimal strategy. We will investigate the relationship between the approximation bounds for the greedy strategy and the values of the curvature constraints. These results have many potential applications in closed-loop control problems such as portfolio management (see, e.g., [73]), sensor management (see, e.g., [74]), and influence in social networks (see, e.g., [75]).

6.1.2 Related Work

Submodular set functions play an important role in combinatorial optimization. Let X be a ground set and $g : 2^X \to \mathbb{R}$ be an objective function defined on the power set 2^X of X. Let $\mathcal{I}$ be a non-empty collection of subsets of X. Suppose that $\mathcal{I}$ has the *hereditary* and *augmentation* properties: 1. For any $S \subset T \subset X$, $T \in I$ implies that $S \in I$; 2. For any $S, T \in I$, if T has a larger cardinality than S, then there exists $j \in T \setminus S$ such that $S \cup \{j\} \in I$. Then, we call $(X, \mathcal{I})$ a matroid [76]. The goal is to find a set in $\mathcal{I}$ to maximize the objective function:

$$\begin{aligned}&\text{maximize } g(N)\\&\text{subject to } N \in \mathcal{I}.\end{aligned} \tag{6.2}$$

Suppose that $\mathcal{I} = \{S \subset X : \text{card}(S) \leq k\}$ for a given k, where $\text{card}(S)$ denotes the cardinality of S. Then, we call $(X, \mathcal{I})$ a *uniform* matroid.

The main difference between (6.1) and (6.2) is that the objective function in (6.1) depends on the order of elements in the string M, while the objective function in (6.2) is independent of the order of elements in the set N. To further explain the difference, we use $\mathcal{P}(M)$ to denote a permutation of a string M. Note that for M with length k, there exist $k!$ permutations. In (6.1), suppose that for any $M \in \mathbb{A}^*$ we have $f(M) = f(\mathcal{P}(M))$ for any $\mathcal{P}$. Then, under these special circumstances, problem (6.1) is equivalent to problem (6.2). In other words, we can view the second problem as a special case of the first problem. Moreover, there can be repeated identical elements in a string, while a set does not contain identical elements (but we note that this difference can be bridged by allowing the notion of multisets in the formulation of submodular set functions).

Finding the solution to (6.2) is NP-hard—a tractable alternative is to use a greedy algorithm. The greedy algorithm starts with the empty set, and incrementally adds an element to the current solution giving the largest gain in the objective function. Theories for maximizing submodular set functions and their applications have been intensively studied in recent years [77]–[101]. The main idea is to compare the performance of the greedy algorithm with that of the optimal solution. Suppose that the set objective function g is non-decreasing: $g(A) \leq g(B)$ for all $A \subset B$; and $g(\emptyset) = 0$ where $\emptyset$ denotes the empty set. Moreover, suppose that the function has the *diminishing-return property*: For all $A \subset B \subset X$ and $j \in X \setminus B$, we have $g(A \cup \{j\}) - g(A) \geq g(B \cup \{j\}) - g(B)$. Then, we say that g is a *submodular set* function. Nemhauser *et al.* [77] showed that the greedy algorithm achieves at least a $(1 - e^{-1})$-approximation for the optimal solution given that $(X, \mathcal{I})$ is a uniform matroid and the objective function is submodular. (By this we mean that the ratio of the objective function value of the greedy solution to that of the optimal solution is at least $(1 - e^{-1})$.) Fisher *et al.* [78] proved that the greedy algorithm provides at least a $1/2$-approximation of the optimal solution for a general matroid. Conforti and Cornuéjols [79] showed that if the function g has a total curvature c, where

$$c = \max_{j \in X} \left\{ 1 - \frac{g(X) - g(X \setminus \{j\})}{g(\{j\}) - g(\emptyset)} \right\},$$

then the greedy algorithm achieves at least $\frac{1}{c}(1 - e^{-c})$ and $\frac{1}{1+c}$-approximations of the optimal solution given that $(X, \mathcal{I})$ is a uniform matroid and a general matroid, respectively. Note that $c \in [0, 1]$ for a submodular set function, and if $c = 0$, then the greedy algorithm is optimal; if $c = 1$, then the result is the same as that in [77]. Vondrák [80] showed that the *continuous greedy algorithm* achieves at least a $\frac{1}{c}(1 - e^{-c})$-approximation for any matroid. Wang *et al.* [81] provided approximation bounds in the case where the function has an elemental curvature α, defined as

$$\alpha = \max_{S \subset X, i,j \in X, i \neq j} \left\{ \frac{g(S \cup \{i, j\}) - g(S \cup \{i\})}{g(S \cup \{j\}) - g(S)} \right\}.$$

The notion of elemental curvature generalizes the notion of diminishing return.

Some recent papers [71], [82]–[84] have extended the notion of set submodularity to problem (6.1). Streeter and Golovin [82] showed that if the function f is *forward* and *backward* monotone: $f(M \oplus N) \geq f(M)$ and $f(M \oplus N) \geq f(N)$ for all $M, N \in \mathbb{A}^*$, where $\oplus$ means string concatenation, and f has the diminishing-return property:

$$f(M \oplus (a)) - f(M) \geq f(N \oplus (a)) - f(N)$$

for all $a \in \mathbb{A}$, $M, N \in \mathbb{A}^*$ such that M is a prefix of N, then the greedy strategy achieves at least a $(1 - e^{-1})$-approximation of the optimal strategy. The notion of *string submodularity* and weaker sufficient conditions are established in [71] under which the greedy strategy still achieves at least a $(1 - e^{-1})$-approximation of the optimal strategy. Golovin and Krause [84] introduced adaptive submodularity for solving stochastic optimization problems under partial observability.

6.1.3 Contributions

In this chapter, we study the problem of maximizing submodular functions defined on strings. We impose additional constraints on curvatures, namely, total backward curvature, total forward curvatures, and elemental forward curvature, which will be rigorously defined in Section 6.2. The notion of total forward and backward curvatures is inspired by the work of Conforti and Cornuéjols [79]. However, the forward and backward algebraic structures are not exposed in the setting of set functions because the objective function defined on sets is independent of the order of elements in a set. The notion of elemental forward curvature is inspired by the work of Wang *et al.* [81]. We have exposed the forward algebraic structure of this elemental curvature in the setting of string functions. Moreover, the result and technical approach in [81] are different from those in this chapter. More specifically, the work in [81] requires the objective function to be a "set function"; that is, independent of order of elements in the set. In our case, order is a crucial component.

In Section 6.3, we consider the maximization problem in the case where the strings are chosen from a uniform structure. For this case, our results are summarized as follows. Suppose that the string submodular function f has total backward curvature $\sigma(O)$ with respect to the optimal strategy. Then, the greedy strategy achieves at least a $\frac{1}{\sigma(O)}(1 - e^{-\sigma(O)})$-approximation of the optimal strategy. Suppose that the string submodular function f has total forward curvature ϵ. Then, the greedy strategy achieves at least a $(1-\epsilon)$-approximation of the optimal strategy. We also generalize the notion of diminishing return by defining the elemental forward curvature η. The greedy strategy achieves at least a $1-(1-\frac{1}{K_\eta})^K$-approximation, where $K_\eta = (1-\eta^K)/(1-\eta)$ if $\eta \neq 1$ and $K_\eta = K$ if $\eta = 1$.

In Section 6.4, we consider the maximization problem in the case where the strings are chosen from a non-uniform structure by introducing the notion of string-matroid. Our results for this case are as follows. Suppose that the string submodular function f has total backward curvature $\sigma(O)$ with respect to the optimal strategy. Then, the greedy strategy achieves at least a $1/(1+\sigma(O))$-approximation. We also provide approximation bounds for the greedy strategy when the function has total forward curvature and elemental forward curvature.

In Section 6.5, we consider three applications of string submodular functions with curvature constraints: 1) Designing a string of learning/fusion rules in balanced binary relay trees; 2) choosing a string of actions to maximize the expected fraction of accomplished tasks; and 3) designing a string of measurement matrices such that the information gain is maximized.

6.2 String Submodularity, Curvature, and Strategies

6.2.1 String Submodularity

We now introduce notation (same to those in [71]) to define string submodularity. Consider a set $\mathbb{A}$ of all possible actions. At each stage i, we choose an action a_i from $\mathbb{A}$. Let $A = (a_1, a_2, \ldots, a_k)$ be a *string* of actions taken over k stages, where $a_i \in \mathbb{A}$, $i = 1, 2, \ldots, k$. Let the set of all possible strings of actions be

$$\mathbb{A}^* = \{(a_1, a_2, \ldots, a_k) | k = 0, 1, \ldots \text{ and } a_i \in \mathbb{A},\ i = 1, 2\ldots, k\}.$$

Note that $k = 0$ corresponds to the empty string (no action taken), denoted by $\varnothing$. For a given string $A = (a_1, a_2, \ldots, a_k)$, we define its *string length* as k, denoted $|A| = k$. If $M = (a_1^m, a_2^m, \ldots, a_{k_1}^m)$ and $N = (a_1^n, a_2^n, \ldots, a_{k_2}^n)$ are two strings in $\mathbb{A}^*$, we say $M = N$ if $|M| = |N|$ and $a_i^m = a_i^n$ for each $i = 1, 2, \ldots, |M|$. Moreover, we define string *concatenation* as follows:

$$M \oplus N = (a_1^m, a_2^m, \ldots, a_{k_1}^m, a_1^n, a_2^n, \ldots, a_{k_2}^n).$$

If M and N are two strings in $\mathbb{A}^*$, we write $M \preceq N$ if we have $N = M \oplus L$, for some $L \in \mathbb{A}^*$. In other words, M is a *prefix* of N.

A function from strings to real numbers, $f : \mathbb{A}^* \to \mathbb{R}$, is *string submodular* if

i. f has the *forward-monotone* property, i.e.,

$$\forall M, N \in \mathbb{A}^*, \qquad f(M \oplus N) \geq f(M).$$

ii. f has the *diminishing-return* property, i.e.,

$$\forall M \preceq N \in \mathbb{A}^*, \forall a \in \mathbb{A},$$
$$f(M \oplus (a)) - f(M) \geq f(N \oplus (a)) - f(N).$$

In the rest of the chapter, we assume that $f(\varnothing) = 0$. Otherwise, we can replace f with the marginalized function $f - f(\varnothing)$. From the forward-monotone property, we know that $f(M) \geq 0$ for all $M \in \mathbb{A}^*$.

We first state an immediate result from the definition of string submodularity.

Lemma 6.2.1. *Suppose that the objective function f is string submodular. Then, for any string $N = (n_1, n_2, \ldots, n_{|N|})$, we have*

$$f(N) \leq \sum_{i=1}^{|N|} f((n_i)).$$

Proof. We use mathematical induction to prove this lemma. If $|N| = 1$, then the result is trivial. Suppose the claim in the lemma holds for any string with length k, we wish to prove the claim for any string with length $k+1$. Let $N = (n_1, n_2, \ldots, n_k, n_{k+1})$. By the diminishing return property, we have

$$f((n_{k+1})) - f(\varnothing) \geq f(N) - f((n_1, \ldots, n_k)).$$

Therefore, by the assumption of the induction, we obtain

$$f(N) \leq f((n_{k+1})) + f((n_1, \ldots, n_k)) \leq \sum_{i=1}^{k+1} f((n_i)).$$

This completes the induction proof. $\square$

6.2.2 Curvature

We define the *total backward curvature* of f by

$$\sigma = \max_{a \in \mathbb{A}, M \in \mathbb{A}^*} \left\{ 1 - \frac{f((a) \oplus M) - f(M)}{f((a)) - f(\varnothing)} \right\}. \tag{6.3}$$

We define the total backward curvature of f with respect to string $M \in \mathbb{A}^*$ by

$$\sigma(M) = \max_{N \in \mathbb{A}^*, 0 < |N| \leq K} \left\{ 1 - \frac{f(N \oplus M) - f(M)}{f(N) - f(\varnothing)} \right\}, \tag{6.4}$$

where K is the length constraint in (6.1). Suppose that f is backward-monotone; i.e., $\forall M, N \in \mathbb{A}^*$, $f(M \oplus N) \geq f(N)$. Then, we have $\sigma \leq 1$ and f has total curvature at most σ with respect to any $M \in \mathbb{A}^*$; i.e, $\sigma(M) \leq \sigma \ \forall M \in \mathbb{A}^*$. This fact can be shown using a simple derivation: For any $N \in \mathbb{A}^*$, we have

$$\begin{aligned} &f(N \oplus M) - f(M) = \\ &\sum_{i=1}^{|N|} f((n_i, \ldots, n_{|N|}) \oplus M) - f((n_{i+1}, \ldots, n_{|N|}) \oplus M), \end{aligned}$$

where n_i represents the ith element of N. From the definition of total backward curvature and Lemma 6.2.1, we obtain

$$\begin{aligned} f(N \oplus M) - f(M) &\geq \sum_{i=1}^{|N|} (1 - \sigma) f((n_i)) \\ &\geq (1 - \sigma) f(N), \end{aligned}$$

which implies that $\sigma(M) \leq \sigma \leq 1$. We will give a lower bound for $\sigma(M)$ in the next section.

Symmetrically, we define the *total forward curvature* of f by

$$\epsilon = \max_{a \in \mathbb{A}, M \in \mathbb{A}^*} \left\{ 1 - \frac{f(M \oplus (a)) - f(M)}{f((a)) - f(\varnothing)} \right\}. \tag{6.5}$$

Moreover, we define the total forward curvature with respect to M by

$$\epsilon(M) = \max_{N \in \mathbb{A}^*, 0 < |N| \leq K} \left\{ 1 - \frac{f(M \oplus N) - f(M)}{f(N) - f(\varnothing)} \right\}. \tag{6.6}$$

If f is string submodular and has total forward curvature ϵ, then it has total forward curvature at most ϵ with respect to any $M \in \mathbb{A}^*$; i.e., $\epsilon(M) \le \epsilon\ \forall M \in \mathbb{A}^*$. Moreover, for a string submodular function f, it is easy to see that for any M, we have $\epsilon(M) \le \epsilon \le 1$ because of the forward-monotone property and $\epsilon(M) \ge 0$ because of the diminishing-return property.

We define the *elemental forward curvature* of the string submodular function by

$$\eta = \max_{a_i, a_j \in \mathbb{A}, M \in \mathbb{A}^*} \frac{f(M \oplus (a_i) \oplus (a_j)) - f(M \oplus (a_i))}{f(M \oplus (a_j)) - f(M)}. \tag{6.7}$$

Moreover, we define the K*-elemental forward curvature* as follows:

$$\hat{\eta} = \max_{a_i, a_j \in \mathbb{A}, M \in \mathbb{A}^*, |M| \le 2K-2} \frac{f(M \oplus (a_i) \oplus (a_j)) - f(M \oplus (a_i))}{f(M \oplus (a_j)) - f(M)}. \tag{6.8}$$

For a forward-monotone function, we have $\eta \ge 0$, and the diminishing-return is equivalent to the condition $\eta \le 1$. By the definitions, we know that $\hat{\eta} \le \eta$ for all K.

The definitions of $\sigma(M)$, $\epsilon(M)$, and $\hat{\eta}$ depend on the length constraint K of the optimal control problem (6.1), whereas σ, ϵ, and η are independent of K. In other words, σ, ϵ, and η can be treated as the *universal* upper bounds for $\sigma(M)$, $\epsilon(M)$, and $\hat{\eta}$, respectively.

6.2.3 Strategies

We will consider the following two strategies.

1) Optimal strategy: Consider the problem (6.1) of finding a string that maximizes f under the constraint that the string length is not larger than K. We call a solution of this problem an *optimal strategy* (a term we already have used repeatedly before). Note that because the function f is forward monotone, it suffices to just find the optimal strategy subject to the stronger constraint that the string length is equal to K. In other words, if there exists an optimal strategy, then there exists one with length K.

2) Greedy strategy: A string $G_k = (a_1^*, a_2^*, \ldots, a_k^*)$ is called *greedy* if

$$a_i^* = \arg\max_{a_i \in \mathbb{A}} f((a_1^*, a_2^*, \ldots, a_{i-1}^*, a_i)) - f((a_1^*, a_2^*, \ldots, a_{i-1}^*))$$

$\forall i = 1, 2, \dots, k.$

Notice that the greedy strategy only maximizes the step-wise gain in the objective function. In general, the greedy strategy (also called the greedy string) is not an optimal solution to (6.1). In this chapter, we establish theorems which state that the greedy strategy achieves at least a factor of the performance of the optimal strategy, and therefore serves in some sense to *approximate* an optimal strategy.

6.3 Uniform Structure

Let I consist of those elements of $\mathbb{A}^*$ with maximal length K: $I = \{A \in \mathbb{A}^* : |A| \leq K\}$. We call I a *uniform structure*. Note that the way we define uniform structure is similar to the way we define independent sets associated with uniform matroids. We will investigate the case of non-uniform structure in the next section. Now (6.1) can be rewritten as

$$\begin{aligned}&\text{maximize } f(M)\\ &\text{subject to } M \in I.\end{aligned}$$

We first consider the relationship between the total curvatures and the approximation bounds for the greedy strategy.

Theorem 6.3.1. *Consider a string submodular function f. Let O be a solution to* (6.1). *Then, any greedy string G_K satisfies*

(i)

$$\begin{aligned} f(G_K) &\geq \frac{1}{\sigma(O)}\left(1 - \left(1 - \frac{\sigma(O)}{K}\right)^K\right) f(O) \\ &> \frac{1}{\sigma(O)}(1 - e^{-\sigma(O)}) f(O), \end{aligned}$$

(ii) $f(G_K) \geq (1 - \max_{i=1,\dots,K-1} \epsilon(G_i)) f(O)$.

Proof. (i) For any $M \in \mathbb{A}^*$ and any $N = (a_1, a_2, \ldots, a_{|N|}) \in \mathbb{A}^*$, we have

$$\begin{aligned} & f(M \oplus N) - f(M) \\ &= \sum_{i=1}^{|N|} (f(M \oplus (a_1, \ldots, a_i)) - f(M \oplus (a_1, \ldots, a_{i-1}))) \end{aligned}$$

Therefore, using the forward-monotone property, there exists an element $a_j \in \mathbb{A}$ such that

$$f(M \oplus (a_1, \ldots, a_j)) - f(M \oplus (a_1, \ldots, a_{j-1})) \geq \frac{1}{|N|}(f(M \oplus N) - f(M)).$$

Moreover, the diminishing-return property implies that

$$\begin{aligned} & f(M \oplus (a_j)) - f(M) \\ &\geq f(M \oplus (a_1, \ldots, a_j)) - f(M \oplus (a_1, \ldots, a_{j-1})) \\ &\geq \frac{1}{|N|}(f(M \oplus N) - f(M)). \end{aligned}$$

Now let us consider the optimization problem (6.1). Using the property of the greedy strategy and the above inequality (substitute $M = G_{i-1}$ and $N = O$), for each $i = 1, 2, \ldots, K$ we have

$$\begin{aligned} f(G_i) - f(G_{i-1}) &\geq \frac{1}{K}(f(G_{i-1} \oplus O) - f(G_{i-1})) \\ &\geq \frac{1}{K}(f(O) - \sigma(O) f(G_{i-1})). \end{aligned}$$

Therefore, we have

$$\begin{aligned} f(G_K) &\geq \frac{1}{K} f(O) + \left(1 - \frac{\sigma(O)}{K}\right) f(G_{K-1}) \\ &= \frac{1}{K} f(O) \sum_{i=0}^{K-1} \left(1 - \frac{\sigma(O)}{K}\right)^i \\ &= \frac{1}{\sigma(O)} \left(1 - \left(1 - \frac{\sigma(O)}{K}\right)^K\right) f(O). \end{aligned}$$

Note that

$$\frac{1}{\sigma(O)} \left(1 - \left(1 - \frac{\sigma(O)}{K}\right)^K\right) \to \frac{1}{\sigma(O)}(1 - e^{-\sigma(O)})$$

from above as $K \to \infty$. This achieves the desired result.

(ii) Using a similar argument to that in (i), we have

$$\begin{aligned}
f(G_i) - f(G_{i-1}) &\geq \frac{1}{K}(f(G_{i-1} \oplus O) - f(G_{i-1})) \\
&\geq \frac{1}{K}(f(G_{i-1}) + (1 - \epsilon(G_{i-1}))f(O) - f(G_{i-1})) \\
&= \frac{1}{K}(1 - \epsilon(G_{i-1}))f(O).
\end{aligned}$$

Therefore, by recursion we have

$$\begin{aligned}
f(G_K) &= \sum_{i=1}^{K}(f(G_i) - f(G_{i-1})) \\
&\geq \sum_{i=1}^{K} \frac{1}{K}(1 - \epsilon(G_{i-1}))f(O) \\
&\geq \frac{1}{K}K(1 - \max_{i=1,\ldots,K-1} \epsilon(G_i))f(O) \\
&= (1 - \max_{i=1,\ldots,K-1} \epsilon(G_i))f(O).
\end{aligned}$$

□

Under the framework of maximizing submodular set functions, similar results are reported in [79]. However, the forward and backward algebraic structures are not exposed in [79] because the total curvature there does not depend on the order of the elements in a set. In the setting of maximizing string submodular functions, the above theorem exposes the roles of forward and backward algebraic structures in bounding the greedy strategy. To explain further, let us state the results in a symmetric fashion. Suppose that the diminishing-return property is stated in a backward way: $f((a) \oplus M) - f(M) \geq f((a) \oplus N) - f(N)$ for all $a \in \mathbb{A}$ and $M, N \in \mathbb{A}^*$ such that $N = (a_1, \ldots, a_k) \oplus M$. Moreover, a string $\hat{G}_k = (a_1^*, a_2^*, \ldots, a_k^*)$ is called *backward-greedy* if

$$a_i^* = \arg\max_{a_i \in \mathbb{A}} f((a_i, a_{i-1}^*, \ldots, a_2^*, a_1^*)) - f((a_{i-1}^*, \ldots, a_1^*)) \quad \forall i = 1, 2, \ldots, k.$$

Then, we can derive bounds in the same way as Theorem 6.3.1, and the results are symmetric.

The results in Theorem 6.3.1 implies that for a string submodular function, we have $\sigma(O) \geq 0$. Otherwise, part (i) of Theorem 6.3.1 would imply that $f(G_K) \geq f(O)$, which is absurd. Recall too that if the function is backward monotone, then $\sigma(O) \leq \sigma \leq 1$. From these facts and part (i) of Theorem 6.3.1, we obtain the following result, also derived in [82].

Corollary 6.3.1. *Suppose that f is string submodular and backward monotone. Then,*

$$f(G_K) \geq (1-(1-\frac{1}{K})^K)f(O) > (1-e^{-1})f(O).$$

Another immediate result follows from the facts that $\sigma(O) \leq \sigma$ and $\epsilon(G_i) \leq \epsilon$ for all i.

Corollary 6.3.2. *Suppose that f is string submodular and backward monotone. Then,*

(i)

$$\begin{aligned} f(G_K) &\geq \frac{1}{\sigma}\left(1-\left(1-\frac{\sigma}{K}\right)^K\right)f(O) \\ &> \frac{1}{\sigma}(1-e^{-\sigma})f(O), \end{aligned}$$

(ii) $f(G_K) \geq (1-\epsilon)f(O)$.

We note that the bounds $\frac{1}{\sigma}(1-e^{-\sigma})$ and $(1-\epsilon)$ are independent of the length constraint K. Therefore, the above bounds can be treated as universal lower bounds of the greedy strategy for all possible length constraints.

Next, we use elemental forward curvature to generalize the diminishing-return property and we investigate the approximation bound using the elemental forward curvature.

Theorem 6.3.2. *Consider a forward-monotone function f with K-elemental forward curvature $\hat{\eta}$ and elemental forward curvature η. Let O be an optimal solution to* (6.1)*. Suppose that $f(G_i \oplus O) \geq f(O)$ for $i = 1, 2, \ldots, K-1$. Then, any greedy string G_K satisfies*

$$\begin{aligned} f(G_K) &\geq f(O)\left(1-(1-\frac{1}{K_{\hat{\eta}}})^K\right) \\ &\geq f(O)\left(1-(1-\frac{1}{K_{\eta}})^K\right), \end{aligned}$$

where $K_{\hat{\eta}} = (1-\hat{\eta}^K)/(1-\hat{\eta})$ *if* $\hat{\eta} \neq 1$ *and* $K_\eta = K$ *if* $\hat{\eta} = 1$*;* $K_\eta = (1-\eta^K)/(1-\eta)$ *if* $\eta \neq 1$ *and* $K_\eta = K$ *if* $\eta = 1$.

Proof. For any $M, N \in \mathbb{A}^*$ such that $|M| \leq K$ and $|N| \leq K$, by the definition of K-elemental forward curvature, there exists $a \in \mathbb{A}$ such that

$$\begin{aligned}
f(M \oplus N) - f(M) &= \sum_{i=1}^{|N|} (f(M \oplus (a_1, \ldots, a_i)) - f(M \oplus (a_1, \ldots, a_{i-1}))) \\
&\leq \sum_{i=1}^{|N|} \hat{\eta}^{i-1} (f(M \oplus a_i) - f(M)) \\
&\leq (1 + \hat{\eta} + \hat{\eta}^2 + \ldots + \hat{\eta}^{|N|-1})(f(M \oplus a) - f(M)) \\
&= K_{\hat{\eta}}(f(M \oplus a) - f(M)).
\end{aligned}$$

Now let us consider the optimization problem (6.1) with length constraint K. Using the property of the greedy strategy and the assumptions, we have for $i = 1, 2, \ldots, K$,

$$\begin{aligned}
f(G_i) - f(G_{i-1}) &\geq \frac{1}{K_{\hat{\eta}}}(f(G_{i-1} \oplus O) - f(G_{i-1})) \\
&\geq \frac{1}{K_{\hat{\eta}}}(f(O) - f(G_{i-1})).
\end{aligned}$$

Therefore, by recursion, we have

$$\begin{aligned}
f(G_K) &\geq \frac{1}{K_{\hat{\eta}}} f(O) + (1 - \frac{1}{K_{\hat{\eta}}}) f(G_{K-1}) \\
&= \frac{1}{K_{\hat{\eta}}} f(O) \sum_{i=0}^{K-1} (1 - \frac{1}{K_{\hat{\eta}}})^i \\
&= f(O) \left(1 - (1 - \frac{1}{K_{\hat{\eta}}})^K \right).
\end{aligned}$$

Because $1 - (1 - \frac{1}{K_{\hat{\eta}}})^K$ is decreasing as a function of $\hat{\eta}$ and $\hat{\eta} \leq \eta$ by definition, we obtain

$$f(O) \left(1 - (1 - \frac{1}{K_{\hat{\eta}}})^K \right) \geq f(O) \left(1 - (1 - \frac{1}{K_\eta})^K \right).$$

□

Recall that $\hat{\eta}$ depends on the length constraint K, whereas η does not. Therefore, the lower bound using K_η can be treated as a universal lower bound of the greedy strategy.

Suppose that f is string submodular. Then, we have $\eta \leq 1$. Because $1 - (1 - \frac{1}{K_\eta})^K$ is decreasing as a function of η, we obtain the following result, which is reported in [71].

Corollary 6.3.3. *Consider a string submodular function f. Let O be a solution to* (6.1)*. Suppose that $f(G_i \oplus O) \geq f(O)$ for $i = 1, 2, \ldots, K-1$. Then, any greedy string G_K satisfies*

$$f(G_K) \geq (1 - (1 - \frac{1}{K})^K) f(O) > (1 - e^{-1}) f(O).$$

The second inequality in the above corollary is given by the fact that $1 - (1 - \frac{1}{K})^K \rightarrow 1 - e^{-1}$ from above, as K goes to infinity. Next we combine the results in Theorems 6.3.1 and 6.3.2 to yield the following result.

Proposition 6.3.1. *Consider a forward-monotone function f with elemental forward curvature η and K-elemental forward curvature $\hat{\eta}$. Let O be a solution to* (6.1)*. Then, any greedy string G_K satisfies*

(i)

$$\begin{aligned} f(G_K) &\geq \frac{1}{\sigma(O)} \left(1 - \left(1 - \frac{\sigma(O)}{K_{\hat{\eta}}}\right)^K\right) f(O) \\ &\geq \frac{1}{\sigma(O)} \left(1 - \left(1 - \frac{\sigma(O)}{K_\eta}\right)^K\right) f(O), \end{aligned}$$

(ii)

$$\begin{aligned} f(G_K) &\geq (1 - \max_{i=1,\ldots,K-1} \epsilon(G_i)) \frac{K}{K_{\hat{\eta}}} f(O) \\ &\geq (1 - \max_{i=1,\ldots,K-1} \epsilon(G_i)) \frac{K}{K_\eta} f(O). \end{aligned}$$

Proof. (i) For any $M, N \in \mathbb{A}^*$ and $|M| \leq K$, $|N| \leq K$, we have shown in the proof of Theorem 6.3.1 that, there exists $a \in \mathbb{A}$ such that $f(M \oplus N) - f(M) \leq K_{\hat{\eta}}(f(M \oplus a) - f(M))$.

Now let us consider the optimization problem (6.1) with length constraint K. Using the property of the greedy strategy and the monotone property, we have

$$\begin{aligned} f(G_i) - f(G_{i-1}) &\geq \frac{1}{K_{\hat{\eta}}}(f(G_{i-1} \oplus O) - f(G_{i-1})) \\ &\geq \frac{1}{K_{\hat{\eta}}}(f(O) - \sigma(O) f(G_{i-1})). \end{aligned}$$

Therefore, by recursion, we have

$$\begin{aligned} f(G_K) &\geq \frac{1}{K_{\hat{\eta}}} f(O) + (1 - \frac{\sigma(O)}{K_{\hat{\eta}}}) f(G_{K-1}) \\ &= \frac{1}{K_{\hat{\eta}}} f(O) \sum_{i=0}^{K-1} (1 - \frac{\sigma(O)}{K_{\hat{\eta}}})^i \\ &= \frac{1}{\sigma(O)} \left(1 - (1 - \frac{\sigma(O)}{K_{\hat{\eta}}})^K\right) f(O). \end{aligned}$$

The second inequality simply follows from the facts that $\frac{1}{\sigma(O)}\left(1 - (1 - \frac{\sigma(O)}{K_{\hat{\eta}}})^K\right)$ is a monotone decreasing function of $\hat{\eta}$ and $\hat{\eta} \leq \eta$ by definition.

(ii) Using a similar argument as part (i), we have

$$\begin{aligned} f(G_i) - f(G_{i-1}) &\geq \frac{1}{K_{\hat{\eta}}}(f(G_{i-1} \oplus O) - f(G_{i-1})) \\ &\geq \frac{1}{K_{\hat{\eta}}}(f(G_{i-1}) - f(G_{i-1}) + (1 - \epsilon(G_{i-1})) f(O)). \end{aligned}$$

Therefore, by recursion,

$$\begin{aligned} f(G_K) &= \sum_{i=1}^{K} (f(G_i) - f(G_{i-1})) \\ &\geq \sum_{i=1}^{K} \frac{1}{K_{\hat{\eta}}}(1 - \epsilon(G_{i-1})) f(O) \\ &\geq \frac{K}{K_{\hat{\eta}}}(1 - \max_{i=1,\ldots,K-1} \epsilon(G_i)) f(O). \end{aligned}$$

The second inequality simply follows from the facts that $\frac{K}{K_{\hat{\eta}}}$ is a monotone decreasing function of $\hat{\eta}$ and $\hat{\eta} \leq \eta$ by definition.

$\square$

We note that the condition in Theorem 6.3.2, $f(G_i \oplus O) \geq f(O)$ for $i = 1, \dots, K-1$, is essentially captured by $\sigma(O)$. In other words, even if the condition $f(G_i \oplus O) \geq f(O)$ is violated, we can still provide approximation bound using $\sigma(O)$, which is larger than 1 in this case.

6.4 Non-uniform Structure

In the last section, we considered the case where I is a uniform structure. In this section, we consider the case of non-uniform structures.

We first need the following definition. Let $M = (m_1, m_2, \dots, m_{|M|})$ and $N = (n_1, n_2, \dots, n_{|N|})$ be two strings in $\mathbb{A}^*$. We write $M \prec N$ if there exists a sequence of strings $L_i \in \mathbb{A}^*$ such that

$$
\begin{aligned}
N = L_1 \oplus (m_1, \dots, m_{i_1}) \oplus L_2 \oplus (m_{i_1+1}, \dots, m_{i_2}) \oplus \dots \\
\oplus (m_{i_{k-1}+1}, \dots, m_{|M|}) \oplus L_{k+1}.
\end{aligned}
$$

In other words, we can remove some elements in N to get M. Note that $\prec$ is a weaker notion of dominance than $\preceq$ defined earlier in Section 6.2. In other words, $M \preceq N$ implies that $M \prec N$ but the converse is not necessarily true.

Now we state the definition of a non-uniform structure, analogous to the definition of independent sets in matroid theory. A subset I of $\mathbb{A}^*$ is called a *non-uniform structure* if it satisfies the following conditions:

1. I is non-empty;

2. *Hereditary*: $\forall M \in I$, $N \prec M$ implies that $N \in I$;

3. *Augmentation*: $\forall M, N \in I$ and $|M| < |N|$, there exists an element $x \in \mathbb{A}$ in the string N such that $M \oplus (x) \in I$.

By analogy with the definition of a matroid, we call the pair $(\mathbb{A}, I)$ a *string-matroid*. We assume that there exists K such that for all $M \in I$ we have $|M| \leq K$ and there exists a

$N \in I$ such that $|N| = K$. We call such a string N a *maximal* string. We are interested in the following optimization problem:

$$\begin{aligned} &\text{maximize } f(N) \\ &\text{subject to } N \in I. \end{aligned} \tag{6.9}$$

Note that if the function is forward monotone, then the maximum of the function subject to a string-matroid constraint is achieved at a maximal string in the matroid. The greedy strategy $G_k = (a_1^*, \ldots, a_k^*)$ in this case is given by

$$a_i^* = \underset{a_i \in \mathbb{A} \text{ and } (a_1^*, \ldots, a_{i-1}^*, a_i) \in I}{\arg\max} f((a_1^*, a_2^*, \ldots, a_{i-1}^*, a_i)) - f((a_1^*, a_2^*, \ldots, a_{i-1}^*)),$$

$\forall i = 1, 2, \ldots, k$. Compared with (6.1), at each stage i, instead of choosing a_i arbitrarily in $\mathbb{A}$ to maximize the step-wise gain in the objective function, we also have to choose the action a_i such that the concatenated string $(a_1^*, \ldots, a_{i-1}^*, a_i)$ is an element of the non-uniform structure I. We first establish the following theorem.

Theorem 6.4.1. *For any $N \in I$, there exists a permutation of N, denoted by $\mathcal{P}(N) = (\hat{n}_1, \hat{n}_2, \ldots, \hat{n}_{|N|})$, such that for $i = 1, 2, \ldots, |N|$ we have*

$$f(G_{i-1} \oplus (\hat{n}_i)) - f(G_{i-1}) \leq f(G_i) - f(G_{i-1}).$$

Proof. We prove this claim by induction on $i = |N|, |N| - 1, \ldots, 1$ (in descending order). If $i = |N|$, considering $G_{|N|-1}$ and N, we know from the String-Matroid Axiom 3 that there exists an element of N, denoted by $\hat{n}_{|N|}$ (we can always permute this element to the end of the string with a certain permutation), such that $G_{|N|-1} \oplus (\hat{n}_{|N|}) \in I$. Moreover, we know that the greedy way of selecting $a_{|N|}^*$ gives the largest gain in the objective function. Therefore, we obtain

$$f(G_{|N|-1} \oplus (\hat{n}_i)) - f(G_{|N|-1}) \leq f(G_{|N|}) - f(G_{|N|-1}).$$

Now let us assume that the claim holds for all $i > i_0$ and the corresponding elements are $\{\hat{n}_{i_0+1}, \ldots, \hat{n}_{|N|}\}$. Next we show that the claim is true for $i = i_0$. Let $\hat{N}_{i_0}$ be the string after we remove the elements in $\{\hat{n}_{i_0+1}, \ldots, \hat{n}_{|N|}\}$ from the original string N. We know from Axiom 2 that $\hat{N}_{i_0} \in I$ and that $|G_{i_0-1}| < |\hat{N}_{i_0}|$, therefore, there exists an element from $\hat{N}_{i_0}$, denoted by $\hat{n}_{i_0}$, such that $G_{i_0-1} \oplus (\hat{n}_{i_0}) \in I$. Using the property of the greedy strategy, we obtain

$$f(G_{i_0-1} \oplus (\hat{n}_{i_0})) - f(G_{i_0-1}) \leq f(G_{i_0}) - f(G_{i_0-1}).$$

This concludes the induction proof. □

Next we investigate the approximation bounds for the greedy strategy using the total curvatures.

Theorem 6.4.2. *Let O be an optimal strategy for* (6.9). *Suppose that f is a string submodular function. Then, a greedy strategy G_K satisfies*

(i) $f(G_K) \geq \frac{1}{1+\sigma(O)} f(O)$,

(ii) $f(G_K) \geq (1 - \epsilon(G_K)) f(O)$.

Proof. (i) By the definition of the total backward curvature, we know that

$$f(G_K \oplus O) - f(O) \geq (1 - \sigma(O)) f(G_K).$$

Therefore, we have

$$\begin{aligned} f(O) &\leq f(G_K \oplus O) - (1 - \sigma(O)) f(G_K) \\ &= f(G_K) - (1 - \sigma(O)) f(G_K) + f(G_K \oplus O) - f(G_K). \end{aligned}$$

Let $O = (o_1, o_2, \ldots, o_K)$. By the diminishing-return property, we have

$$\begin{aligned} f(G_K \oplus O) - f(G_K) &= \sum_{i=1}^{K} (f(G_K \oplus (o_1, \ldots, o_i)) - f(G_K \oplus (o_1, \ldots, o_{i-1}))) \\ &\leq \sum_{i=1}^{K} (f(G_K \oplus (o_i)) - f(G_K)). \end{aligned}$$

By Theorem 6.4.1, we know that there exists a permutation: $\mathcal{P}(O) = (\hat{o}_1, \hat{o}_2, \dots, \hat{o}_{|O|})$ such that

$$f(G_{i-1} \oplus (\hat{o}_i)) - f(G_{i-1}) \leq f(G_i) - f(G_{i-1}),$$

for $i = 1, 2, \dots, K$. Therefore, by the diminishing-return property again,

$$\begin{aligned}\sum_{i=1}^{K}(f(G_K \oplus (o_i)) - f(G_K)) &\leq \sum_{i=1}^{K}(f(G_{i-1} \oplus (\hat{o}_i)) - f(G_{i-1}))\\ &\leq \sum_{i=1}^{K}(f(G_i) - f(G_{i-1}))\\ &= f(G_K).\end{aligned}$$

From the above equations,

$$\begin{aligned}f(O) &\leq f(G_K) + f(G_K) - (1 - \sigma(O))f(G_K)\\ &= (1 + \sigma(O))f(G_K),\end{aligned}$$

and this achieves the desired result.

(ii) From the definition of total forward curvature, we have

$$f(G_K \oplus O) - f(G_K) \geq (1 - \epsilon(G_K))f(O).$$

From the proof of part (i), we also know that $f(G_K \oplus O) - f(G_K) \leq f(G_K)$. Therefore, we have $f(G_K) \geq (1 - \epsilon(G_K))f(O)$. □

The inequality in (i) above is a generalization of a result on maximizing submodular set functions with a general matroid constraint [78]. The submodular set counterpart involves total curvature, whereas the string version involves total *backward* curvature. Note that if f is backward monotone, then $\sigma(O) \leq \sigma \leq 1$. We now state an immediate corollary of Theorem 6.4.2.

Corollary 6.4.1. *Suppose that f is string submodular and backward monotone. Then, the greedy strategy achieves at least a $1/2$-approximation of the optimal strategy.*

Another immediate result follows from the facts that $\sigma(O) \le \sigma$ and $\epsilon(G_K) \le \epsilon$.

Corollary 6.4.2. *Suppose that f is string submodular and backward monotone. Then, we have*

(i) $f(G_K) \ge \frac{1}{1+\sigma} f(O)$,

(ii) $f(G_K) \ge (1-\epsilon) f(O)$.

Next we generalize the diminishing-return property using the elemental forward curvature.

Theorem 6.4.3. *Suppose that f is a forward-monotone function with elemental forward curvature η and K-elemental forward curvature $\hat{\eta}$. Suppose that $f(G_K \oplus O) \ge f(O)$. If $\hat{\eta} \le 1$, then*

$$f(G_K) \ge \frac{1}{1+\hat{\eta}} f(O) \ge \frac{1}{1+\eta} f(O).$$

If $\hat{\eta} > 1$, then

$$f(G_K) \ge \frac{1}{1+\hat{\eta}^{2K-1}} f(O) \ge \frac{1}{1+\eta^{2K-1}} f(O).$$

Proof. Let $O = (o_1, o_2, \ldots, o_K)$. From the definition of K-elemental forward curvature,

$$\begin{aligned}
f(G_K \oplus O) - f(G_K) &= \sum_{i=1}^{K} (f(G_K \oplus (o_1, \ldots, o_i)) - f(G_K \oplus (o_1, \ldots, o_{i-1}))) \\
&\le \sum_{i=1}^{K} (f(G_{K-1} \oplus (o_i)) - f(G_{K-1}))\hat{\eta}^i \\
&\le \begin{cases} \hat{\eta} \sum_{i=1}^{K} (f(G_{K-1} \oplus (o_i)) - f(G_{K-1})), & \text{if } \hat{\eta} \le 1 \\ \hat{\eta}^K \sum_{i=1}^{K} (f(G_{K-1} \oplus (o_i)) - f(G_{K-1})), & \text{if } \hat{\eta} > 1. \end{cases}
\end{aligned}$$

From Theorem 6.4.1, there exists a permutation $\mathcal{P}$ of O: $\mathcal{P}(O) = (\hat{o}_1, \ldots, \hat{o}_K)$, such that

$$f(G_{i-1} \oplus (\hat{o}_i)) - f(G_{i-1}) \le f(G_i) - f(G_{i-1}),$$

for $i = 1, \ldots, K$.

Moreover, by the definition of K-elemental forward curvature,

$$\begin{aligned}
&\sum_{i=1}^{K}(f(G_{K-1}\oplus(o_i))-f(G_{K-1}))\\
&=\sum_{i=1}^{K}(f(G_{K-1}\oplus(\hat{o}_i))-f(G_{K-1}))\\
&\leq\sum_{i=1}^{K}\hat{\eta}^{K-i}(f(G_{i-1}\oplus(\hat{o}_i))-f(G_{i-1}))\\
&\leq\begin{cases}\sum_{i=1}^{K}(f(G_{i-1}\oplus(\hat{o}_i))-f(G_{i-1})), & \text{if } \hat{\eta}\leq 1\\ \hat{\eta}^{K-1}\sum_{i=1}^{K}(f(G_{i-1}\oplus(\hat{o}_i))-f(G_{i-1})), & \text{if } \hat{\eta}>1.\end{cases}\\
&\leq\begin{cases}f(G_K), & \text{if } \hat{\eta}\leq 1\\ \hat{\eta}^{K-1}f(G_K), & \text{if } \hat{\eta}>1.\end{cases}
\end{aligned}$$

Therefore, we have

$$f(O)\leq\begin{cases}(1+\hat{\eta})f(G_K), & \text{if } \hat{\eta}\leq 1\\ (1+\hat{\eta}^{2K-1})f(G_K), & \text{if } \hat{\eta}>1.\end{cases}$$

Since $\hat{\eta} \leq \eta$ and $\frac{1}{1+\hat{\eta}}$ and $\frac{1}{1+\hat{\eta}^{2K-1}}$ are monotone decreasing functions of $\hat{\eta}$, we obtain the desired results. □

This result is similar to that in [81]. However, the second bound in Theorem 6.4.3 is different from that in [81]. This is because the proof in [81] uses the fact that the value of a set function at a set is independent of the order of elements in the set, whereas this is not the case for a string. Recall that the elemental forward curvature for a string submodular function is not larger than 1. We obtain the following result.

Corollary 6.4.3. *Suppose that f is a string submodular function and $f(G_K \oplus O) \geq f(O)$. Then, the greedy strategy achieves at least a $1/2$-approximation of the optimal strategy.*

Now we combine the results for total and elemental curvatures to get the following.

Proposition 6.4.1. *Suppose that f is a forward-monotone function with K-elemental forward curvature $\hat{\eta}$ and elemental forward curvature η. Then, a greedy strategy G_K satisfies* *(i)* $f(G_K) \geq \frac{1}{\sigma(O)+h(\hat{\eta})} f(O) \geq \frac{1}{\sigma(O)+h(\eta)} f(O)$,

(ii) $f(G_K) \geq \frac{1-\epsilon(G_K)}{h(\hat{\eta})} f(O) \geq \frac{1-\epsilon(G_K)}{h(\eta)} f(O)$,

where $h(\hat{\eta}) = \hat{\eta}$ *and* $h(\eta) = \eta$ *if* $\hat{\eta} \leq 1$*;* $h(\hat{\eta}) = \hat{\eta}^{2K-1}$ *and* $h(\eta) = \eta^{2K-1}$ *if* $\hat{\eta} > 1$.

Proof. (i) Using the definition of total backward curvature, we have $f(G_K \oplus O) - f(O) \geq (1 - \sigma(O))f(G_K)$, which implies that $f(G_K \oplus O) - f(G_K) \geq f(O) - \sigma(O)f(G_K)$. Using a similar argument as that of Theorem 6.4.3, we know that

$$f(G_K \oplus O) - f(G_K) \leq h(\hat{\eta})f(G_K).$$

Therefore, we have

$$f(G_K) \geq \frac{1}{h(\hat{\eta}) + \sigma(O)} f(O).$$

The second inequality follows from $h(\hat{\eta}) \leq h(\eta)$.

(ii) Using the definition of total forward curvature, we have

$$f(G_K \oplus O) - f(G_K) \geq (1 - \epsilon(G_K))f(O).$$

Using a similar argument as that of Theorem 6.4.3, we know that $f(G_K \oplus O) - f(G_K) \leq h(\hat{\eta})f(G_K)$. Therefore, we have

$$f(G_K) \geq \frac{1 - \epsilon(G_K)}{h(\hat{\eta})} f(O).$$

The second inequality follows from $h(\hat{\eta}) \leq h(\eta)$. □

From these results, we know that when f is string submodular, $\hat{\eta} \in [0, 1]$ and we must have $\sigma(O) + \hat{\eta} \geq 1$ and $\epsilon(G_K) + \hat{\eta} \geq 1$. From Theorems 6.3.1, 6.3.2, 6.4.1, and 6.4.2, we see that the performance of the greedy strategy relative to the optimal improves as the total forward/backward curvature or the elemental forward curvature decreases to 0. On the other hand, the inequalities above indicate that this performance improvement with forward and elemental curvature constraints cannot become arbitrarily good simultaneously. When equality in either case holds, the greedy strategy is optimal. A special case for this scenario is when the objective function is *string-linear*: $f(M \oplus N) = f(M) + f(N)$ for all $M, N \in \mathbb{A}^*$, i.e., $\eta = 1$ and $\sigma = \epsilon = 0$. Recall that $0 \leq \sigma(O) \leq \sigma$, $0 \leq \epsilon(G_K) \leq \epsilon$, and $0 \leq \hat{\eta} \leq \eta$. Therefore, we have $\sigma(O) = \epsilon(G_k) = 0$ and $\hat{\eta} = 1$.

Remark 6.4.1. The above proposition and the discussions afterward easily generalize to the framework of submodular set functions.

6.5 Applications

In this section, we investigate three applications of string submodular functions with curvature constraints.

6.5.1 Learning in Balanced Binary Relay Trees

We consider the problem of testing binary hypothesis between H_0 and H_1 in a balanced binary relay tree, with structure shown in Fig. 3.1. Let p be any fusion node (i.e., p is a nonleaf node). We say that p is at level k if there are k hops between this node and the closest sensor (leaf node) in the tree. We denote by $C(p)$ the set of child nodes of p. Suppose that p receives binary messages $Y_c \in \{0, 1\}$ from every $c \in C(p)$ (i.e., from its child nodes), and then summarizes the two received binary messages into a new binary message $Y_p \in \{0, 1\}$ using a fusion rule λ^p:

$$Y_p = \lambda^p(\{Y_c : c \in C(p)\}).$$

The new message Y_p is then communicated to the parent node (if any) of p. Ultimately, the fusion center makes an overall decision.

It turns out that the only meaningful rules to aggregate two binary messages in this case are simply ‘AND’ and ‘OR’ rules defined as follows:

- AND rule (denoted by $\mathcal{A}$): a parent node decides 1 if and only if both its child nodes send 1;

- OR rule (denoted by $\mathcal{O}$): a parent node decides 0 if and only if both its child nodes send 0.

Henceforth, we only consider the case where each fusion node in the tree chooses a fusion rule from $\mathcal{Y} := \{\mathcal{A}, \mathcal{O}\}$.

We assume that all sensors are independent and the binary messages associated with these sensors have identical Type I error probability α_0 and identical Type II error probability β_0. Moreover, we assume that all the fusion nodes at level k ($k \in \{1, 2, \ldots, h\}$) use the same fusion rule λ_k; i.e., for each node p that lies at the kth level of the tree, $\lambda^p = \lambda_k$. In this case, all the output binary messages for nodes at level k have the same Type I and Type II error probabilities, which we denote by α_k and β_k respectively. Given a fusion rule λ_k, we can show that the error probabilities evolve as follows:

$$(\alpha_k, \beta_k) := \begin{cases} (1 - (1 - \alpha_{k-1})^2, \beta_{k-1}^2), & \text{if } \lambda_k = \mathcal{A}, \\ (\alpha_{k-1}^2, 1 - (1 - \beta_{k-1})^2), & \text{if } \lambda_k = \mathcal{O}. \end{cases}$$

Remark 6.5.1. Note that the evolution of the error probability pair (α_k, β_k) is symmetric with respect to the line $\alpha + \beta = 1$. Hence, it suffices to consider the case where the initial pair satisfies $\alpha_0 + \beta_0 < 1$. We can derive similar result for the case where $\alpha_0 + \beta_0 > 1$ (e.g., by only flipping the decision at the fusion center). In the case where $\alpha_0 + \beta_0 = 1$, the Type I and II error probabilities add up to one regardless of the fusion rule used. Hence, this case is not of interest.

Notice that the ULRT fusion rule is either the $\mathcal{A}$ rule or the $\mathcal{O}$ rule, depending on the values of the Type I and Type II error probabilities at a particular level of the tree. More precisely, we have

- If $\beta_k > \alpha_k$, then the ULRT fusion rule is $\mathcal{A}$;
- If $\beta_k < \alpha_k$, then the ULRT fusion rule is $\mathcal{O}$;

- If $\beta_k = \alpha_k$, then the total error probability remains unchanged after using $\mathcal{A}$ or $\mathcal{O}$. Moreover, the error probability pairs at the next level $(\alpha_{k+1}, \beta_{k+1})$ after using $\mathcal{A}$ or $\mathcal{O}$ are symmetric about the line $\beta = \alpha$. Therefore, we call both $\mathcal{A}$ and $\mathcal{O}$ the ULRT fusion rule in this case.

We define a fusion strategy as a string of fusion rules $\lambda_j \in \mathcal{Y}$ used at levels $j = 1, 2, \ldots, h$, denoted by $\pi = (\lambda_1, \lambda_2, \ldots, \lambda_h)$. Note that h denotes the height of the tree. Let the collection of all possible fusion strategies with length h be $\mathcal{Y}^h$:

$$\mathcal{Y}^h := \{\pi = (\lambda_1, \lambda_2, \ldots, \lambda_h) | \lambda_j \in \mathcal{Y} \text{ for } j = 1, 2, \ldots, h\}.$$

For a given initial error probability pair (α_0, β_0) at the sensor level, the pair (α_h, β_h) at the fusion center (level h) is a function of (α_0, β_0) and the specific fusion strategy π used. We consider the Bayesian criterion in this chapter, under which the objective is to minimize the total error probability $\pi_0\alpha_h + \pi_1\beta_h$ at the fusion center, where π_0 and π_1 are the prior probabilities of the two hypotheses, respectively. Equivalently, we can find a strategy that maximizes the reduction of the total error probability between the sensors and the fusion center. We call this optimization problem an *h-optimal problem*. Without loss of generality, we assume that the prior probabilities are equal; i.e., $\pi_0 = \pi_1 = 1/2$, in which case the h-optimal problem (ignoring a factor of $1/2$) can be written as:

$$\begin{aligned} &\text{maximize } \alpha_0 + \beta_0 - (\alpha_h + \beta_h) \\ &\text{subject to } \pi \in \mathcal{Y}^h. \end{aligned} \tag{6.10}$$

A fusion strategy that maximizes (6.10) is called the h-optimal strategy:

$$\begin{aligned} \pi^o(\alpha_0, \beta_0) &= \operatorname*{arg\,max}_{\pi \in \mathcal{Y}^h} (\alpha_0 + \beta_0 - (\alpha_h + \beta_h)) \\ &= \operatorname*{arg\,max}_{\pi \in \mathcal{Y}^h} \sum_{j=0}^{h-1} (\alpha_j + \beta_j - (\alpha_{j+1} + \beta_{j+1})). \end{aligned}$$

In contrast, the ULRT fusion rule only minimizes the step-wise reduction in the total error probability:

$$\text{ULRT} = \operatorname*{arg\,max}_{\lambda_i \in \mathcal{Y}} (\alpha_i + \beta_i - (\alpha_{i+1} + \beta_{i+1})) \quad \forall i.$$

Because of the equal prior probability assumption, a *maximum a posteriori* (MAP) fusion rule is the same as the ULRT fusion rule. In this context, we call a fusion strategy consisting of repeated ULRT fusion rules at all levels a ULRT (greedy) strategy.

In the next section, we derive the h-optimal fusion strategy for balanced binary relay trees with height h using a dynamic programming approach. More specifically, we express the solution using Bellman's equations. We then show that the 2-optimal strategy is equivalent to the ULRT strategy. Moreover, we show that the reduction of the total error probability is a string submodular function, which implies that the greedy strategy is close to the optimal fusion strategy in terms of the reduction in the total error probability.

6.5.1.1 Dynamic programming formulation

In this section, we formulate the problem of finding the optimal fusion strategy using a deterministic dynamic programming model. First we define the necessary elements of this dynamic model.

I. *Dynamic System*: We define the error probability pair at the kth level (α_k, β_k) as the system state, denoted by s_k. Notice that α_k and β_k can only take values in the interval $[0, 1]$. Therefore, the set of all possible states is $\{(\alpha, \beta) > 0 | \alpha + \beta < 1\}$. Moreover, given the fusion rule, the *state transition function* is deterministic. If we choose $\lambda_k = \mathcal{A}$, then

$$(\alpha_k, \beta_k) = (1 - (1 - \alpha_{k-1})^2, \beta_{k-1}^2).$$

On the other hand, if we choose $\lambda_k = \mathcal{O}$, then

$$(\beta_k, \alpha_k) = (1 - (1 - \beta_{k-1})^2, \alpha_{k-1}^2).$$

II. *Rewards*: At each level k, we define the instantaneous reward to be the reduction of the total error probability after fusing with λ_k:

$$r(s_{k-1}, \lambda_k) = (\alpha_{k-1} + \beta_{k-1}) - (\alpha_k + \beta_k),$$

where α_k and β_k are functions of the previous state s_{k-1} and the fusion rule λ_k.

Let $v_{h-k}(s_k)$ be the cumulative reduction of the total error probability if we start the system at state s_k at level k and the strategy $(\lambda_{k+1}, \lambda_{k+2} \ldots, \lambda_h) \in \mathcal{Y}^{h-k}$ is used. Following the above definitions, we have

$$v_{h-k}(s_k) = \sum_{j=k+1}^{h} r(s_{j-1}, \lambda_j).$$

If we let $k = 0$, that is, we start calculating the reduction from the sensor level, then the above cumulative reward function is the same as the global objective function defined in (6.10). Therefore, for given initial state s_0, we have to solve the following optimization problem to find the global optimal strategy over horizon h:

$$v_h^o(s_0) = \max_{\pi \in \mathcal{Y}^h} \sum_{j=1}^{h} r(s_{j-1}, \lambda_j).$$

The globally optimal strategy π^o is

$$\pi^o(s_0) = \arg\max_{\pi \in \mathcal{Y}^h} \sum_{j=1}^{h} r(s_{j-1}, \lambda_j).$$

Notice that s_k depends on the previous state s_{k-1} and the fusion rule λ_k. We will write the state at level k to be $s_k|_{s_{k-1},\lambda_k}$. The solution of the above optimization problem can be characterized using Bellmam's equations, which state that

$$v_h^o(s_0) = \max_{\lambda_1 \in \mathcal{Y}} \left[r(s_0, \lambda_1) + v_{h-1}^o(s_1|_{s_0,\lambda_1}) \right]$$

$$\lambda_1^o(s_0) = \arg\max_{\lambda_1 \in \mathcal{Y}} \left[r(s_0, \lambda_1) + v_{h-1}^o(s_1|_{s_0,\lambda_1}) \right],$$

where $\lambda_1^o(s_0)$ is the first element of the optimal strategy $\pi^o(s_0)$. Recursively, the solution of the optimization problem is given by

$$v_{h-(k-1)}^o(s_{k-1}) = \max_{\lambda_k \in \mathcal{Y}} \left[r(s_{k-1}, \lambda_k) + v_{h-k}^o(s_k|_{s_{k-1},\lambda_k}) \right].$$

Moreover, the kth element of the optimal strategy $\pi^o(s_0)$ is given by

$$\lambda_k^o(s_{k-1}) = \arg\max_{\lambda_k \in \mathcal{Y}} \left[r(s_{k-1}, \lambda_k) + v_{h-k}^o(s_k|_{s_{k-1},\lambda_k}) \right].$$

Remark 6.5.2. The above formulation can easily be generalized to the case where the nodes and links in the tree fails with certain probabilities and even more complicated network architectures simply by changing the state transition functions and the set of all possible fusion rules. Also, we can easily generalize the above formulation to non-equal prior probability scenario.

The complexity of the explicit solution to Bellman's equations grows exponentially with respect to the horizon h. Therefore, it is usually intractable to compute the h-optimal strategy if h is sufficiently large. An alternative strategy is the ULRT strategy, which consists of repeating ULRT fusion rule at all levels. We have shown in Chapter 3 that the decay rate of the total error probability with this strategy is $\sqrt{N}$. Next we study whether the ULRT strategy is the same as the h-optimal strategy. If not, does the ULRT strategy provide a reasonable approximation of the h-optimal strategy?

6.5.1.2 2-optimal strategy

In this section, we show that the 2-optimal strategy is the same as the ULRT strategy. Moreover, we give an counterexample which shows that the ULRT strategy is not 3-optimal.

Consider the 2-optimal problem in the balanced binary relay tree with height 2:

$$v_2^o(s_0) = \max_{\pi \in \mathcal{Y}^2} \sum_{j=1}^{2} r(s_{j-1}, \lambda_j),$$

where $\mathcal{Y}^2 = \{(\mathcal{A}, \mathcal{A}), (\mathcal{A}, \mathcal{O}), (\mathcal{O}, \mathcal{O}), (\mathcal{O}, \mathcal{A})\}$. The 2-optimal strategy in this case is

$$\pi^o(s_0) = \arg\max_{\pi \in \mathcal{Y}^2} \sum_{j=1}^{2} r(s_{j-1}, \lambda_j).$$

We have the following theorem.

Theorem 6.5.1. *A strategy π is 2-optimal if and only if π is the ULRT strategy.*

Proof. First consider the special cases where $\beta_k = \alpha_k$ in the 2-optimal problem for $k = 0$ or $k = 1$. We know that when $\beta_k = \alpha_k$, then both $\mathcal{A}$ and $\mathcal{O}$ do not change the total error probability after fusion and the next states after using $\mathcal{A}$ and $\mathcal{O}$ are symmetric with respect to $\beta = \alpha$ line. Moreover, both $\mathcal{A}$ and $\mathcal{O}$ are called the ULRT strategy. Consequentially in this case, the 2-optimal problem reduces to a 1-optimal problem. Hence, if π is 2-optimal, then we can show that π is the ULRT strategy. On the other hand, if π is the ULRT strategy and $\beta_k = \alpha_k$ for $k = 0$ or $k = 1$, then it is easy to show that π is always 2-optimal.

Now we show the theorem for the case where $\beta_k \neq \alpha_k$ for $k = 0$ and $k = 1$. First note that the total error probability decreases strictly after fusing with the ULRT fusion rule. However, if we apply the fusion rule other than the ULRT fusion rule in $\mathcal{Y}$, then the total error probability increases strictly after fusion. For example, if $\beta_k > \alpha_k$ and we apply the $\mathcal{O}$ fusion rule, then the total error probability increases strictly; i.e.,

$$\alpha_{k+1} + \beta_{k+1} = \alpha_k^2 + 1 - (1 - \beta_k)^2 > \alpha_k + \beta_k,$$

in other words, the instantaneous reward in this case is negative,

$$r(s_k, \mathcal{O}) < 0.$$

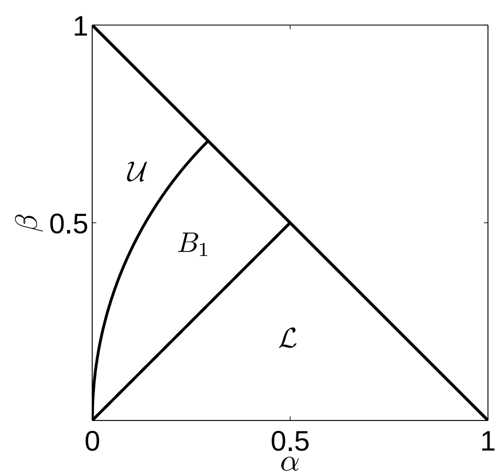


Figure 6.1: Regions $\mathcal{U}$, $\mathcal{L}$, and B_1 in the (α, β) plane.

Because of symmetry, it suffices to prove this theorem in the upper triangular region $\mathcal{U}$ defined as follows (see Fig. 6.1):

$$\mathcal{U} := \{(\alpha, \beta) \geq 0 | \alpha + \beta < 1 \text{ and } \beta > \alpha\}.$$

We define the reflection of $\mathcal{U}$ with respect to $\beta = \alpha$ line to be $\mathcal{L}$. Recall that if $(\alpha_k, \beta_k) \in B_1$, where

$$B_1 := \{(\alpha, \beta) \in \mathcal{U} | (1 - \alpha)^2 + \beta^2 \leq 1\},$$

then the next state $(\alpha_{k+1}, \beta_{k+1}) \in \mathcal{L}$. See Fig. 6.1 for the region B_1. Also recall that if (α_0, β_0) lies on the boundary of B_1, then the next state (α_1, β_1) lies on $\beta = \alpha$ line. Hence, this boundary is not considered.

We divide the proof into two cases:

- *Case I*: $(\alpha_0, \beta_0) \in B_1$, in which case the ULRT strategy is $(\mathcal{A}, \mathcal{O})$;
- *Case II*: $(\alpha_0, \beta_0) \in \mathcal{U} \setminus B_1$, in which case the ULRT strategy is $(\mathcal{A}, \mathcal{A})$.

For Case I where $(\alpha_0, \beta_0) \in B_1$, it is easy to see that strategy $(\mathcal{A}, \mathcal{O})$ achieves a larger reduction than that of $(\mathcal{A}, \mathcal{A})$, because using $\mathcal{A}$ rule for the second level increases the total error probability. Moreover, the total error probability after using $(\mathcal{O}, \mathcal{O})$ increases with respect to the initial total error probability. Hence, this fusion rule cannot be 2-optimal. It suffices to show that the strategy $(\mathcal{A}, \mathcal{O})$ achieves a larger reduction than that of $(\mathcal{O}, \mathcal{A})$:

$$r(s_0, \mathcal{A}) + r(s_1, \mathcal{O}) > r(s_0, \mathcal{O}) + r(s_1, \mathcal{A}),$$

which is equivalent with the following inequality

$$\begin{aligned}
&r(s_0, \mathcal{A}) + r(s_1, \mathcal{O}) - (r(s_0, \mathcal{O}) + r(s_1, \mathcal{A})) = \\
&(1 - (1 - \beta_0)^2)^2 + 1 - (1 - \alpha_0^2)^2 - \\
&((1 - (1 - \alpha_0)^2)^2 + 1 - (1 - \beta_0^2)^2) > 0.
\end{aligned}$$

The above inequality can be reduced to

$$\beta_0^2(1-\beta_0)^2 - \alpha_0^2(1-\alpha_0)^2 > 0,$$

which holds for all $(\alpha_0, \beta_0) \in B_1$. Hence, the 2-optimal fusion strategy in this case is also $(\mathcal{A}, \mathcal{O})$. We conclude that if $(\alpha_0, \beta_0) \in B_1$, then a strategy is 2-optimal if and only if it is the ULRT strategy.

For Case II where $(\alpha_0, \beta_0) \in \mathcal{U} \setminus B_1$, it is easy to see that strategy $(\mathcal{A}, \mathcal{A})$ achieves a larger reduction than that of $(\mathcal{A}, \mathcal{O})$. Moreover, the total error probability after using $(\mathcal{O}, \mathcal{O})$ increases with respect to the initial total error probability. Hence, this fusion rule cannot be 2-optimal. It suffices to show that the strategy $(\mathcal{A}, \mathcal{A})$ achieves a larger reduction than that of $(\mathcal{O}, \mathcal{A})$:

$$r(s_0, \mathcal{A}) + r(s_1, \mathcal{A}) > r(s_0, \mathcal{O}) + r(s_1, \mathcal{A}),$$

which reduces to

$$\begin{aligned}
&r(s_0, \mathcal{A}) + r(s_1|_{s_0}, \mathcal{A}) - (r(s_0, \mathcal{O}) + r(s_1|_{s_0}, \mathcal{A})) = \\
&(1-(1-\beta_0)^2)^2 + 1 - (1-\alpha_0^2)^2 - \\
&(1-(1-\alpha_0)^4 + \beta_0^4) > 0.
\end{aligned}$$

The above inequality is equivalent to

$$\beta_0(1-\beta_0)(1+\beta_0) - \alpha_0(1-\alpha_0)(1-\alpha_0) > 0,$$

which holds for all $(\alpha_0, \beta_0) \in \mathcal{U} \setminus B_1$. Therefore, the 2-optimal fusion strategy in this case is also $(\mathcal{A}, \mathcal{A})$. We conclude that if $(\alpha_0, \beta_0) \in \mathcal{U} \setminus B_1$, then a strategy is 2-optimal if and only if it is the ULRT strategy.

□

This result also applies to any sub-tree with height 2 within a balanced binary relay tree with arbitrary height $h > 2$. However, the ULRT strategy is not in general optimal for multiple levels; i.e., $h > 2$, as the following counter-example for $h = 3$ shows.

Let the initial state be $(\alpha_0, \beta_0) = (0.2, 0.3)$, in which case the ULRT strategy is $(\mathcal{A}, \mathcal{O}, \mathcal{A})$. As shown in Fig. 6.2, the solid (red) line denotes the total error probabilities at each level up to 3. However, the 3-optimal strategy in this case is $(\mathcal{O}, \mathcal{A}, \mathcal{A})$. The total error probability curve of this strategy is shown as a dashed (green) line in Fig. 6.2. Similar counterexamples can be found for cases where $h > 3$. Hence, the ULRT strategy is not in general h-optimal for $h \geq 3$. In the next section, we will introduce and employ the notion of string submodularity to quantify the gap in performances between optimal and ULRT strategies for $h \geq 3$.

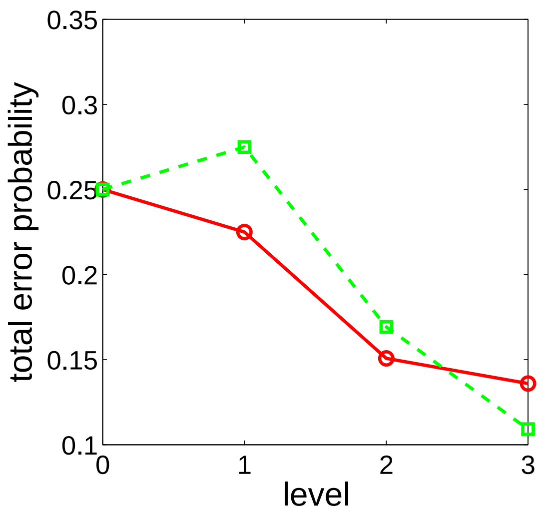


Figure 6.2: Comparison of the ULRT strategy and the 3-optimal strategy. The solid (red) line represents the error probability curve using the ULRT strategy. The dashed (green) line represents the error probability curve using the 3-optimal strategy.

6.5.1.3 String submodularity

We now apply the theory of string subnormality to learning in balanced binary relay trees with even heights. Again we assume that the nodes at the same level use the same fusion rule. Moreover, we assume that two fusion rules Λ of consecutive levels k and $k+1$ (without loss of generality, we assume that k is an even number) are chosen from the following set $\mathcal{Z} = \{(\mathcal{A}, \mathcal{O}), (\mathcal{O}, \mathcal{A})\}$. Let $\Pi = (\Lambda_1, \Lambda_2, \dots, \Lambda_h)$ be a fusion strategy, where $\Lambda_i \in \mathcal{Z}$ for all i. Let $\mathcal{Z}^*$ be the set of all possible strategies (strings); i.e., $\mathcal{Z}^* = \{(\Lambda_1, \Lambda_2, \dots, \Lambda_h) | h =$

$0, 1, \ldots$ and $\Lambda_i \in \mathcal{Z}\ \forall i\}$. Here we only prove the case where the prior probabilities are equally likely. The following analysis easily generalizes to non-equal prior probabilities. Given the two types of error probability (α_0, β_0) at level 0, the reduction of the total error probability after applying a strategy Π is

$$u(\Pi) = \alpha_0 + \beta_0 - (\alpha_{2h}(\Pi) + \beta_{2h}(\Pi)),$$

where α_{2h} and β_{2h} represent the Type I and II error probabilities at level $2h$ after fusion with Π.

Next we show that u is a string submodular function.

Proposition 6.5.1. *For sufficiently small (α_0, β_0), the function u: $\mathcal{Z}^* \to \mathbb{R}$ is string submodular.*

Proof. First we show that the function u is a monotone function. It suffices to show the following:

$$u((\Lambda_1, \ldots, \Lambda_k) \oplus (\Lambda^*)) \geq u((\Lambda_1, \ldots, \Lambda_k)),$$

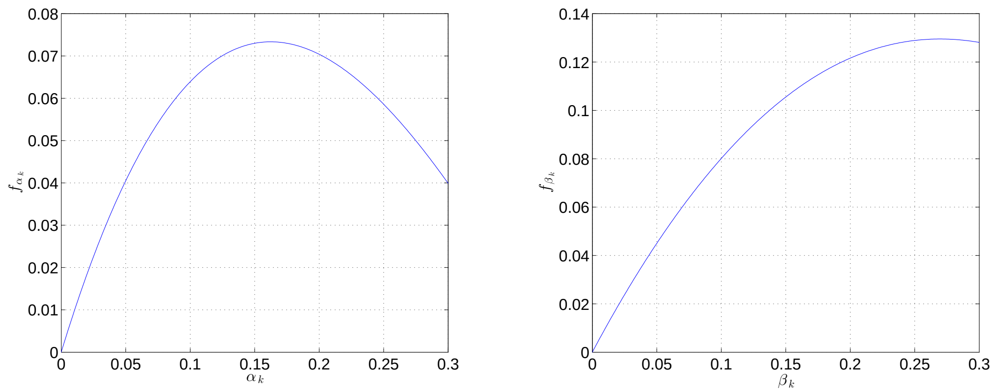
for all $\Lambda_i, \Lambda^* \in \mathcal{Z}$, where $i = 1, 2, \ldots, k$. We will use (α_k, β_k) to denote the error probabilities after using $(\Lambda_1, \ldots, \Lambda_k)$. If $\Lambda^* = (\mathcal{A}, \mathcal{O})$, then we need to show that

$$\begin{aligned}
&u((\Lambda_1, \ldots, \Lambda_k) \oplus (\Lambda^*)) - u((\Lambda_1, \ldots, \Lambda_k)) \\
&= \alpha_k + \beta_k - (1 - (1 - \alpha_k)^2)^2 - (1 - (1 - \beta_k^2)^2) \\
&= f_{\alpha_k} + f_{\beta_k} \geq 0,
\end{aligned}$$

where $f_{\alpha_k} = \alpha_k - (1 - (1 - \alpha_k)^2)^2$ and $f_{\beta_k} = \beta_k - (1 - (1 - \beta_k^2)^2)$. It is evident that f_α and f_β are non-negative if α_k and β_k are sufficiently small. More precisely, if $\alpha_k \leq 0.3$ and $\beta_k \leq 0.3$, then the function u is monotone increasing. Therefore, if the initial error probabilities α_0 and β_0 are sufficiently small, u is monotone increasing. See Fig. 6.3 for plots of f_{α_k} and f_{β_k} versus α_k and β_k, respectively. If $\Lambda^* = (\mathcal{O}, \mathcal{A})$, then

$$u(\Lambda^*) = \alpha_k + \beta_k - (1 - (1 - \alpha_k^2)^2) - (1 - (1 - \beta_k)^2)^2 \geq 0,$$

which also holds for sufficiently small α_k and β_k. This can also be proved using the symmetry property of the problem.



(a) (b)

Figure 6.3: (a) Values of f_{α_k} versus α_k. (b) Values of f_{β_k} versus β_k.

Next we show the diminishing return property of u, that is,

$$u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m)) \geq$$

$$u((\Lambda_1, \Lambda_2, \ldots, \Lambda_n) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_n))$$

for all $m \leq n$, where $\Lambda_i \in \mathcal{Z}$ for all i and $\Lambda^* \in \mathcal{Z}$. First let us consider the simplest case where $m = 0$ and $n = 1$; i.e.,

$$u((\Lambda^*)) - u(\emptyset) \geq u((\Lambda_1, \Lambda^*)) - u((\Lambda_1)), \tag{6.11}$$

for all $\Lambda_1, \Lambda^* \in \mathcal{Z}$. We know that $u(\emptyset) = 0$ because the error probabilities do not change without any fusion. Because of symmetry, it suffices to show the above inequality for the cases where $(\Lambda_1, \Lambda^*) = (\mathcal{A}, \mathcal{O}) \oplus (\mathcal{A}, \mathcal{O})$ and $(\Lambda_1, \Lambda^*) = (\mathcal{A}, \mathcal{O}) \oplus (\mathcal{O}, \mathcal{A})$. The error probabilities evolves as follows:

$$(\alpha_k, \beta_k) \xrightarrow{\Lambda_1} (\alpha_{k+2}, \beta_{k+2}) \xrightarrow{\Lambda^*} (\alpha_{k+4}, \beta_{k+4}).$$

Again because of symmetry, we only consider the evolution for α_k.

$$
\begin{aligned}
&u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m)) \\
&\geq u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m, \Lambda_{m+1}) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m, \Lambda_{m+1})) \\
&\geq u((\Lambda_1, \Lambda_2, \ldots, \Lambda_{m+1}, \Lambda_{m+2}) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_{m+1}, \Lambda_{m+2})).
\end{aligned}
\tag{6.12}
$$

Case i: If $(\Lambda_1, \Lambda^*) = (\mathcal{A}, \mathcal{O}) \oplus (\mathcal{A}, \mathcal{O})$, then we can show the following

$$
\begin{aligned}
&\alpha_{k+4} - \alpha_{k+2} - (\alpha_{k+2} - \alpha_k) = \\
&-\alpha_k^{16} + 8\alpha_k^{14} - 24\alpha_k^{12} + 32\alpha_k^{10} \\
&-14\alpha_k^{8} - 8\alpha_k^{6} + 10\alpha_k^{4} - 4\alpha_k^{2} + \alpha_k \geq 0,
\end{aligned}
$$

which holds for sufficiently small α_k. See Fig. 6.4(a) for a plot of $\alpha_{k+4} - \alpha_{k+2} - (\alpha_{k+2} - \alpha_k)$ versus α_k. Notice that if $\alpha_k < 0.3$, then the above inequality holds. This analyze easily generalizes to the inequality for the Type II error probability by symmetry.

Case ii: If $(\Lambda_1, \Lambda^*) = (\mathcal{A}, \mathcal{O}) \oplus (\mathcal{O}, \mathcal{A})$, then we have

$$
\begin{aligned}
&\alpha_{k+4} - \alpha_{k+2} - (\bar{\alpha}_{k+2} - \alpha_k) = \\
&-\alpha_k^{16} + 2\alpha_k^{8} + 2\alpha_k^{4} - 4\alpha_k^{2} + \alpha_k \geq 0,
\end{aligned}
$$

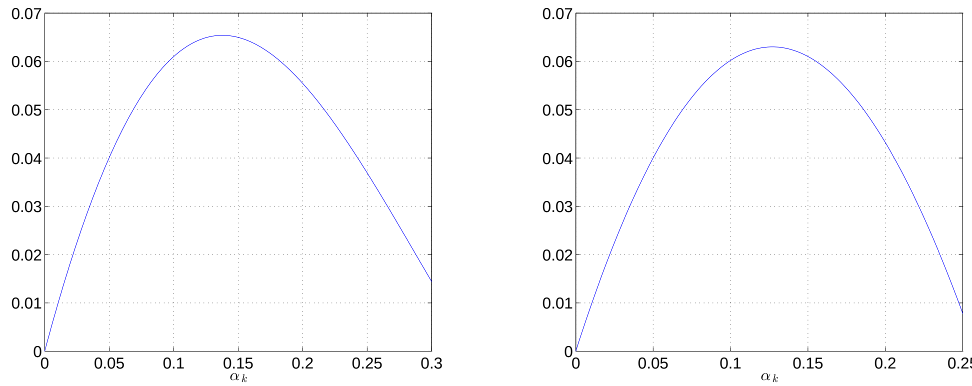
which holds for sufficiently small α_k. We note that $\bar{\alpha}_{k+2}$ denotes the Type I error probability after using Λ^*. See Fig. 6.4(b) for a plot of $\alpha_{k+4} - \alpha_{k+2} - (\bar{\alpha}_{k+2} - \alpha_k)$ versus α_k. Notice that if $\alpha_k < 0.25$, then the above inequality holds. Therefore, the inequality (6.11) for the simplest case holds.

From this case, it is easy to show (6.12). Then by recursion, we have

$$
\begin{aligned}
&u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_m)) \\
&\geq u((\Lambda_1, \Lambda_2, \ldots, \Lambda_n) \oplus (\Lambda^*)) - u((\Lambda_1, \Lambda_2, \ldots, \Lambda_n))
\end{aligned}
$$

for all $m \leq n$, where $\Lambda_i \in \mathcal{Z}$ for all i and $\Lambda^* \in \mathcal{Z}$.

$\square$



(a) (b)

Figure 6.4: (a) Values of $\alpha_{k+4} - \alpha_{k+2} - (\alpha_{k+2} - \alpha_k)$ versus α_k. (b) Values of $\alpha_{k+4} - \alpha_{k+2} - (\bar{\alpha}_{k+2} - \alpha_k)$ versus α_k.

For a balanced binary relay trees with height $2K$, the global optimization problem is to find a strategy $\Pi \in \mathcal{Z}^*$ with length K such that the above reduction is maximized; that is

$$\begin{array}{l} \text{maximize } u(\Pi) \\ \text{subject to } \Pi \in \mathcal{Z}^*, |\Pi| = K. \end{array} \tag{6.13}$$

We have shown that the reduction of the total error probability u is a string submodular function. Moreover, we know that the total error probability does not change if there is no fusion; i.e.,

$$u(\emptyset) = 0.$$

Therefore, we can employ Corollary 6.3.3 to the above maximization problem (6.13).

Consider a balanced binary relay tree with height $2K$. We denote by $u(G_K)$ the reduction of the total error probability after using the greedy strategy. We have shown that the ULRT strategy is 2-optimal. Moreover, we have also shown in Chapter 3 that the ULRT strategy only allows at most two identical consecutive fusion rules after the error probability pair enters a certain regime in the (α, β) plane. Hence, we can conclude that a strategy is the ULRT strategy if and only if it is the greedy strategy. We denote by $u(O)$ the reduction of the total error probability using the optimal strategy. We have the following theorem.

Theorem 6.5.2. *Consider a balanced binary relay tree with height $2K$. We have*

$$(1 - e^{-1})u(O) < u(G_K) \leq u(O).$$

Proof. The inequality on the right hand side holds because

$$u(\Pi) \leq u(O)$$

for all $\Pi \in \Pi^*$ where $|\Pi| = K$.

For the inequality on the left hand side, we have shown that u is a string submodular function with $u(\emptyset) = 0$. For any greedy string G, we have $u(G) \geq 0$ because of the monotone property. Moreover, we can show that the Type I and II error probabilities both decrease after applying the string G. Therefore, the Type I error probability after G is not larger than the initial Type I error probability.

We know that the mapping $s_\Pi : \alpha_k \to \alpha_{k+2}$ is a monotone non-decreasing function respect to α_k for any fusion rule $\Pi \in \mathcal{Z}$. Moreover, the optimal string O is simply a composite of several such monotone non-decreasing functions. Hence, the Type I error probability after applying O is a monotone non-decreasing function with respect to the initial Type I error probability. With these, we conclude that $u(G \oplus O) \geq u(O)$ for any greedy string G.

We know that u is a string submodular function from Proposition 6.5.1. We also know that $u(\emptyset) = 0$. Therefore, after applying Corollary 6.3.1 to this problem we complete the proof.

□

Remark 6.5.3. Recall that the fusion strategy is a string of fusion rules chosen from $\mathcal{Z} = \{(\mathcal{A}, \mathcal{O}), (\mathcal{O}, \mathcal{A})\}$. Thus, the strategies we considered in this section have at most two consecutive repeated fusion rules. For example, the strategy $(\mathcal{A}, \mathcal{A}, \mathcal{A}, \ldots)$ is not considered. It is easy to show that with repeating identical fusion rule, the total error probability goes to 1/2. Therefore, it is reasonable to rule out this situation.

$$(\alpha_{k+1}, \beta_{k+1}) := \begin{cases} \left(\frac{(1-q_k)(1-(1-\alpha_k)^2)+2q_k\alpha_k}{1+q_k}, \frac{(1-q_k)\beta_k^2+2q_k\beta_k}{1+q_k}\right), & \text{if } \lambda = \mathcal{A}, \\ \left(\frac{(1-q_k)\alpha_k^2+2q_k\alpha_k}{1+q_k}, \frac{(1-q_k)(1-(1-\beta_k)^2)+2q_k\beta_k}{1+q_k}\right), & \text{if } \lambda = \mathcal{O}. \end{cases} \tag{6.14}$$

6.5.1.4 Node and link failures

We now consider balanced binary relay trees with node and link failures, in which case the decay rate of the total error probability has been considered in Chapter 4. We assume that each node at level k fails with identical node failure probability n_k (a failed node cannot transmit any message upward). We model each link as a *binary erasure channel*. With a certain probability, the input message X (either 0 or 1) gets erased and the receiver does not get any data. We assume that the links between nodes at height k and height $k+1$ have identical probability of erasure ℓ_k.

Consider a node $\mathcal{N}_k$ at level k connected to its parent node $\mathcal{N}_{k+1}$ at level $k+1$. We define several probabilities as follows:

- Local failure probability p_k: the probability that either the node $\mathcal{N}_k$ fails or the link from $\mathcal{N}_k$ to $\mathcal{N}_{k+1}$ fails.

- Silence probability q_k: the probability that $\mathcal{N}_{k+1}$ does not receive a message from $\mathcal{N}_k$.

From the above definition, we have

$$p_k = n_k + \ell_k - n_k\ell_k.$$

By the law of total probability, we have

$$q_{k+1} = q_k^2 + p_{k+1}(1 - q_k^2).$$

We can view q_k as the exogenous input of the tree network. In this case, the evolution of the Type I and II error probabilities are give in (6.14).

Again, we consider a balanced binary relay tree with even height $2h$ and the consecutive fusion rules are chosen from $\mathcal{Z}$.

$$\alpha_{k+2} = \frac{(1-q_k)(1-q_{k+1})}{(1+q_k)(1+q_{k+1})}(1-(1-\alpha_k)^2)^2 + \frac{2q_k(1-q_{k+1})}{(1+q_k)(1+q_{k+1})}\alpha_k^2 + \frac{2q_{k+1}(1-q_k)}{(1+q_k)(1+q_{k+1})}(1-(1-\alpha_k)^2) + \frac{4q_kq_{k+1}}{(1+q_k)(1+q_{k+1})}\alpha_k. \tag{6.15}$$

Proposition 6.5.2. *Suppose that the silence probability sequence is upper bounded by* $1/8$*; that is,* $q_k < 1/8$ *for all* k*. Then, if the initial error probabilities* (α_0, β_0) *are sufficiently small, then the function* u*:* $\Pi^* \to \mathbb{R}$ *is string submodular.*

Proof. We first show that u is non-decreasing. It suffices to show that $u((\mathcal{A}, \mathcal{O})) \geq 0$ starting from (α_k, β_k). We can decompose $(\alpha_{k+2}, \beta_{k+2})$ into different components, for example, the expression for α_{k+2} is given in (6.15). Notice that in (6.15), the coefficient for α_k is $\frac{4q_k}{(1+q_k)(1+q_{k+1})}$. Therefore, if $q_k < 1/8$ for all k, then by the coefficient of α_k in (6.15) is less than 1, which implies that $\alpha_{k+2} \leq \alpha_k$ for sufficiently small α_k.

To show the diminishing return property, it suffice to consider the situation where we use the rule $(\mathcal{A}, \mathcal{O}) \oplus (\mathcal{A}, \mathcal{O})$. In this case, we need to show $\alpha_k + \alpha_{k+4} - 2\alpha_{k+2} \geq 0$. Again, we can write the expression for α_{k+2} in (6.15). Notice that if we write $\alpha_k + \alpha_{k+4} - 2\alpha_{k+2}$ as a function of α_k, then the coefficient for α_k is not smaller than $1 - \frac{8q_k}{(1+q_k)(1+q_{k+1})}$, notice that $q_k \leq 1/8$ by assumption, then the coefficient for α_k is positive. Therefore, we have $\alpha_k + \alpha_{k+4} - 2\alpha_{k+2} \geq 0$ for sufficiently small α_k. We note that our argument rely on the fact that the initial error probabilities (α_0, β_0) are both sufficiently small.

□

Note that if $q_k \geq 1/8$, then the function is not strictly monotone increasing. However, this is a fair assumption considering that the failure probability usually is not as large as $1/8$.

Using similar analysis as the non-failure case, we obtain the following bounds, which capture the performance of greedy strategy compared with the optimal strategy.

Corollary 6.5.1. *Consider a balanced binary relay tree with height $2K$ with node and link failures. Let G_K and O denote the greedy and optimal strategies, respectively. We have*

$$(1-e^{-1})u(O) < u(G_K) \leq u(O).$$

6.5.2 Strategies for Accomplishing Tasks

Consider an objective function of the following form:

$$f((a_1,\ldots,a_k)) = \frac{1}{n}\sum_{i=1}^{n}\left(1-\prod_{j=1}^{k}(1-p_i^j(a_j))\right).$$

We can interpret this objective function as follows. We have n subtasks, and by choosing action a_j at stage j there is a probability $p_i^j(a_j)$ of accomplishing the ith subtask. Therefore, the objective function is the expected fraction of subtasks that are accomplished after performing $(a_1,\ldots,a_k)$. Suppose that p_i^j is independent of j for all i; i.e., the probability of accomplishing the ith subtask by choosing an action does not depend on the stage at which the action is chosen. Then, it is obvious that the objective function does not depend on the order of actions. In this special case, the objective function is a submodular set function and therefore the greedy strategy achieves at least a $(1-e^{-1})$-approximation of the optimal strategy. Moreover, it turns out that this special case is closely related to several previously studied problems, such as min-sum set cover [102], pipelined set cover [103], social network influence [104], and coverage-aware self scheduling in sensor networks [105]. In this section, we generalize the special case to the situation where p_i^j depends on j. Applications of this generalization include designing campaign strategy for political voting, etc. Without loss of generality, we will consider the special case where $n=1$ (our analysis easily generalizes to arbitrary n). In this case, we have

$$f((a_1,\ldots,a_k)) = 1-\prod_{j=1}^{k}(1-p^j(a_j)).$$

For each $a \in \mathbb{A}$, we assume that $p^j(a)$ takes values in $[L(a),U(a)]$, where $0<L(a)<U(a)<1$. Moreover, let

$$c(a) = \frac{1-U(a)}{1-L(a)}.$$

Obviously, $c(a) \in (0,1)$. The forward-monotone property is easy to check: For any $M, N \in \mathbb{A}^*$, the statement that $f(M \oplus N) \geq f(M)$ is obviously true.

6.5.2.1 Uniform structure

We first consider the maximization problem under the uniform structure constraint. The elemental forward curvature in this case is

$$\eta = \max_{a_i, a_j} \frac{(1 - p^i(a_i))p^j(a_j)}{p^i(a_j)}.$$

Suppose that $\hat{U} = \max_{a \in \mathbb{A}} U(a)$ and $\hat{L} = \min_{a \in \mathbb{A}} L(a)$. Then, we have

$$\eta \leq \frac{(1 - \hat{L})\hat{U}}{\hat{L}},$$

for all possible combinations of probability values p^j, $j = 1, 2, \ldots$. Note that the function is submodular if and only if $\eta \leq 1$. From the above equation, we conclude that f is submodular if

$$\frac{(1 - \hat{L})\hat{U}}{\hat{L}} \leq 1.$$

Therefore, a sufficient condition for f to be a string submodular function is

$$\hat{L}^{-1} - \hat{U}^{-1} \leq 1.$$

To apply Theorem 6.3.1, instead of calculating the total backward curvature with respect to the optimal strategy, we calculate the total backward curvature for $K \leq |M| < 2K$:

$$\hat{\sigma} = \max_{a \in \mathbb{A}, K \leq |M| < 2K} \left\{ 1 - \frac{f((a) \oplus M) - f(M)}{f((a)) - f(\varnothing)} \right\} \tag{6.16}$$

$$= 1 - \min_{a \in \mathbb{A}, K \leq |M| < 2K} \left\{ \frac{f((a) \oplus M) - f(M)}{f((a)) - f(\varnothing)} \right\}. \tag{6.17}$$

We have

$$\frac{f((a) \oplus M) - f(M)}{f((a)) - f(\varnothing)} = \frac{\prod_{j=1}^{|M|}(1 - p^j(a_j)) - (1 - p^1(a)) \prod_{j=1}^{|M|}(1 - p^{j+1}(a_j))}{p^1(a)}.$$

We then provide an upper bound for the total backward curvature for all possible combination of p^j. The minimum of the above term is achieved at $p^j(a_j) = \hat{U}$ and $p^{j+1}(a_j) = \hat{L}$:

$$\begin{aligned}\min_{a\in\mathbb{A}, K\leq|M|<2K}\left\{\frac{f((a)\oplus M)-f(M)}{f((a))-f(\varnothing)}\right\} &\geq \min_{a\in\mathbb{A},K\leq k<2K}\frac{(1-\hat{U})^k-(1-p^1(a))(1-\hat{L})^k}{p^1(a)}\\ &\geq \min_{K\leq k<2K}\frac{(1-\hat{U})^k-(1-\hat{L})^{k+1}}{\hat{L}}.\end{aligned}$$

From this we can derive an upper bound for the total backward curvature and use the upper bound in Theorem 6.3.1. For example, suppose that $(1-\hat{U})^k/(1-\hat{L})^k \geq 1-\hat{L}$ for all $K \leq k < 2K$. Then, we can provide an upper bound for $\hat{\sigma}$:

$$\begin{aligned}\hat{\sigma} &\leq 1-\min_{K\leq k<2K}\frac{(1-\hat{U})^k-(1-\hat{L})^{k+1}}{\hat{L}}\\ &= 1-\frac{(1-\hat{U})^{2K-1}-(1-\hat{L})^{2K}}{\hat{L}}.\end{aligned}$$

Moreover, it is easy to verify that $\sigma(O) \leq \hat{\sigma}$. Therefore, we can substitute the above upper bound of $\hat{\sigma}$ to Theorem 6.3.1 to derive a lower bound for the approximation of the greedy strategy.

Instead of calculating the total forward curvature with respect to the greedy strategy G_i, we calculate

$$\begin{aligned}\hat{\epsilon}_i &= \max_{a\in\mathbb{A},i\leq|M|<i+K}\left\{1-\frac{f(M\oplus(a))-f(M)}{f((a))-f(\varnothing)}\right\} &(6.18)\\ &= 1-\min_{a\in\mathbb{A},i\leq|M|<i+K}\left\{\frac{f(M\oplus(a))-f(M)}{f((a))-f(\varnothing)}\right\} &(6.19)\\ &= 1-\min_{a\in\mathbb{A},i\leq|M|<i+K}\frac{\prod_{j=1}^{|M|}(1-p^j(a_j))p^1(a)}{p^1(a)} &(6.20)\\ &\leq 1-(1-\hat{U})^{i+K-1}. &(6.21)\end{aligned}$$

It is easy to show that $\epsilon(G_i) \leq \hat{\epsilon}_i$. Moreover,

$$\max_{i=1,\ldots,K-1}\epsilon(G_i) \leq \max_{i=1,\ldots,K-1}\hat{\epsilon}_i \leq 1-(1-\hat{U})^{2K-2}.$$

We can substitute this upper bound in Theorem 6.3.1 and get a lower bound for the approximation of the optimal strategy that the greedy strategy is guaranteed to achieve.

In Theorem 6.3.2, we need the additional assumption that $f(G_i \oplus O) \leq f(O)$ for $i = 1, \dots, K-1$, which can be written as

$$\prod_{j=1}^{K}(1 - p^j(o_j)) \geq \prod_{t=1}^{i}(1 - p^j(a_j^*)) \prod_{j=1}^{K}(1 - p^{j+i}(o_j)). \tag{6.22}$$

We know that

$$\prod_{j=1}^{K}(1 - p^j(o_j)) \geq \prod_{j=1}^{K}(1 - U(o_j))$$

and

$$\prod_{t=1}^{i}(1 - p^j(a_j^*)) \prod_{j=1}^{K}(1 - p^{j+i}(o_j)) \leq \prod_{j=1}^{K}(1 - L(o_j))(1 - p^1(a_1^*)).$$

Therefore, a sufficient condition for (6.22) is

$$1 - p^1(a_1^*) \leq \frac{\prod_{j=1}^{K}(1 - U(o_j))}{\prod_{j=1}^{K}(1 - L(o_j))} = \prod_{j=1}^{K} c(o_j).$$

Let $c = \min_{a \in \mathbb{A}} c(a)$. Suppose that we have

$$p^1(a_1^*) \geq 1 - c^K.$$

Then, $f(G_i \oplus O) \leq f(O)$ for $i = 1, \dots, K-1$.

6.5.2.2 Non-uniform structure

The calculation for the case of non-uniform structure uses a similar analysis. For example, in Theorem 6.4.2, the calculation of the total backward curvature can be calculated in the same way as the case of uniform structure.

Now let us consider the backward monotone property required in Theorem 6.4.3: $f(G_K \oplus O) \geq f(O)$. This condition is much weaker than that in Theorem 6.3.2, and can be rewritten as

$$\prod_{j=1}^{K}(1 - p^j(o_j)) \geq \prod_{t=1}^{K}(1 - p^j(a_j^*)) \prod_{j=1}^{K}(1 - p^{j+K}(o_j)).$$

A sufficient condition for the above inequality is $1 - \hat{U} \geq (1 - \hat{L})^2$. Recall that the function is string submodular if

$$\eta \leq \frac{(1 - \hat{L})\hat{U}}{\hat{L}} \leq 1.$$

Combining the above two inequalities, we have

$$\begin{aligned}\eta &\leq \frac{(1-\hat{L})\hat{U}}{\hat{L}} \leq \frac{(1-\hat{L})(1-(1-\hat{L})^2)}{\hat{L}} \\ &= (1-\hat{L})(2-\hat{L}) \leq 1.\end{aligned}$$

Therefore, we obtain

$$\hat{L} \geq 1 - \frac{1}{\alpha} \text{ and } \hat{U} \leq \frac{1}{\alpha},$$

where $\alpha = \frac{1+\sqrt{5}}{2}$ is the *golden ratio*.

Now let us consider the special case where $p^j(a)$ is non-increasing over j for each $a \in \mathbb{A}$. It is easy to show that the function is string submodular. Moreover, the elemental forward curvature is

$$\begin{aligned}\eta &= \max_{a_i, a_j} \frac{(1-p^i(a_i))p^j(a_j)}{p^i(a_j)} \\ &\leq \max_{a_i}(1-p^i(a_i)) \\ &\leq 1-\hat{L}.\end{aligned}$$

Therefore, using this upper bound of the elemental forward curvature, we can provide a better approximation than $(1-e^{-1})$ for the greedy strategy.

Consider the special case where $p^j(a)$ is non-decreasing over j for each $a \in \mathbb{A}$. In this case, we have

$$\begin{aligned}\sigma(O) &\leq \hat{\sigma} \\ &= 1 - \min_{a \in \mathbb{A}, K \leq |M| < 2K} \left\{ \frac{f((a) \oplus M) - f(M)}{f((a)) - f(\varnothing)} \right\} \\ &\leq 1 - \prod_{j=1}^{|M|}(1-p^j(a_j)) \\ &\leq 1 - (1-\hat{U})^{2K-1}.\end{aligned}$$

Therefore, we can provide a better approximation than $(1-e^{-1})$ for the greedy strategy using this upper bound for $\sigma(O)$.

6.5.3 Maximizing the Information Gain

In this part, we present an application of our results on string submodular functions to sequential Bayesian estimation. Consider a signal of interest $x \in \mathbb{R}^N$ with normal prior distribution $\mathcal{N}(\mu, P_0)$. In our example, we assume that $N = 2$ for simplicity; our analysis easily generalizes to dimensions larger than 2. Let $\mathbb{D}$ denote the set of diagonal positive-semidefinite 2×2 matrices with unit Frobenius norm:

$$\mathbb{D} = \{\text{Diag}(\sqrt{e}, \sqrt{1-e}) : e \in [0, 1]\}.$$

At each stage i, we choose a measurement matrix $A_i \in \mathbb{D}$ to get an observation y_i, which is corrupted by additive zero-mean Gaussian noise $\omega_i \sim \mathcal{N}(0, R_{\omega_i \omega_i})$:

$$y_i = A_i x + \omega_i.$$

Let us denote the posterior distribution of x given $(y_1, y_2, \ldots, y_k)$ by $\mathcal{N}(x_k, P_k)$. The recursion for the posterior covariance P_k is given by

$$\begin{aligned} P_k^{-1} &= P_{k-1}^{-1} + A_k^T R_{\omega_k \omega_k}^{-1} A_k \\ &= P_0^{-1} + \sum_{i=1}^{k} A_i^T R_{\omega_i \omega_i}^{-1} A_i. \end{aligned}$$

The entropy of the posterior distribution of x given $(y_1, y_2, \ldots, y_k)$ is $H_k = \frac{1}{2} \log \det P_k + \log(2\pi e)$. The information gain given $(A_1, A_2, \ldots, A_k)$ is

$$f((A_1, A_2, \ldots, A_k)) = H_0 - H_k = \frac{1}{2}(\log \det P_0 - \log \det P_k).$$

The objective is to choose a string of measurement matrices subject to a length constraint K such that the information gain is maximized.

The optimality of the greedy strategy and the measurement matrix design problem are considered in [106] and [107], respectively. Suppose that the additive noise sequence is independent and identically distributed. Then, it is easy to see that

$$f((A_1, A_2, \ldots, A_k)) = f(\mathcal{P}(A_1, A_2, \ldots, A_k))$$

for all permutations $\mathcal{P}$. Moreover, the information gain is a submodular set function and $f(\emptyset) = 0$; see [108]. Therefore, the greedy strategy achieves at least a $(1-e^{-1})$-approximation of the optimal strategy.

Consider the situation where the additive noise sequence is independent but *not* identically distributed. Moreover, let us assume that $R_{\omega_i\omega_i} = \sigma_i^2\mathcal{I}$, where $\mathcal{I}$ denotes the identity matrix. In other words, the noise at each stage is white but the variances σ_i depend on i. The forward-monotone property is easy to see: We always gain by adding extra (noisy) measurements.

Now we investigate the sensitivity of string submodularity with respect to the varying noise variances. We claim that the function is string submodular if and only if σ_i is monotone non-decreasing with respect i. The sufficiency part is easy to understand: The information gain at a later stage certainly cannot be larger than the information gain at an earlier stage because the measurement y_i becomes more noisy as i increases. We show the necessity part by contradiction. Suppose that the function is string submodular and there exists k such that $\sigma_k \geq \sigma_{k+1}$. Suppose that the posterior covariance at stage $k-1$ is Diag(s_{k-1}, t_{k-1}) and we choose $A_k =$ Diag$(1, 0)$, $A_{k+1} =$ Diag$(0, 1)$. We have

$$\begin{aligned} f(A_k \oplus A_{k+1}) - f(A_k) &= \log(1 + t_k\sigma_{k+1}^{-2}) \\ &= \log(1 + t_{k-1}\sigma_{k+1}^{-2}) \\ &\geq \log(1 + t_{k-1}\sigma_k^{-2}) \\ &= f(A_{k+1}) - f(\varnothing). \end{aligned}$$

This contradicts the diminishing-return property and completes the argument.

In fact, it is easy to show that $\hat{\eta} \leq \eta \leq 1$ if and only if the sequence of noise variance is non-decreasing. In this case, the greedy strategy achieves at least a factor (better than $(1 - e^{-1})$) of the optimal strategy.

For general cases where the noise variance sequence is not non-decreasing, we will provide an upper bound for the K-elemental forward curvature $\hat{\eta}$. For simplicity, let $P_0 =$ Diag(s_0, t_0).

$$
\begin{aligned}
& f(M \oplus (A_i) \oplus (A_j)) - f(M \oplus (A_i)) \\
&= \log(1 + s_{|M|+1}\sigma_{|M|+2}^{-2} e_j)(1 + t_{|M|+1}\sigma_{|M|+2}^{-2}(1 - e_j)) \\
&= \log(s_{|M|+1}^{-1} + \sigma_{|M|+2}^{-2} e_j)(t_{|M|+1}^{-1} + \sigma_{|M|+2}^{-2}(1 - e_j)) + \log s_{|M|+1} t_{|M|+1} \\
&\leq \log \left(\frac{s_{|M|+1}^{-1} + \sigma_{|M|+2}^{-2} e_j + t_{|M|+1}^{-1} + \sigma_{|M|+2}^{-2}(1 - e_j)}{2} \right)^2 + \\
& \quad \max(-\log(s_0^{-1} + \sum_{i=1}^{|M|+1} \sigma_i^{-2}) t_0^{-1}, -\log s_0^{-1}(t_0^{-1} + \sum_{i=1}^{|M|+1} \sigma_i^{-2})) \\
&= \log \left(\frac{s_0^{-1} + t_0^{-1} + \sum_{i=1}^{|M|+2} \sigma_i^{-2}}{2} \right)^2 - \log s_0^{-1}(t_0^{-1} + \sum_{i=1}^{|M|+1} \sigma_i^{-2}) \qquad (6.23) \\
&= \log \left(\frac{1 + s_0 t_0^{-1} + s_0 \sum_{i=1}^{|M|+2} \sigma_i^{-2}}{2} \right) + \log \left(\frac{s_0^{-1} + t_0^{-1} + \sum_{i=1}^{|M|+2} \sigma_i^{-2}}{2(t_0^{-1} + \sum_{i=1}^{|M|+1} \sigma_i^{-2})} \right) \\
&\leq \log \left(\frac{1 + s_0 t_0^{-1} + s_0 \sum_{i=1}^{2K} \sigma_i^{-2}}{2} \right) + \log \left(\frac{1}{2} \left(1 + \frac{s_0^{-1} + \max_{i=1,\ldots,2K} \sigma_i^{-2}}{(t_0^{-1} + \sigma_1^{-2})} \right) \right).
\end{aligned}
$$

Without loss of generality, we assume that $s_0 \geq t_0$. Let $M = (A_1, A_2, \ldots, A_{|M|})$ where $A_k = \text{Diag}(\sqrt{e_k}, \sqrt{1 - e_k})$ for $k = 1, \ldots, |M|$. Let $P_{|M|} = \text{Diag}(s_{|M|}, t_{|M|})$ where

$$
s_0^{-1} \leq s_{|M|}^{-1} = s_0^{-1} + \sum_{i=1}^{|M|} \sigma_i^{-2} e_i \leq s_0^{-1} + \sum_{i=1}^{|M|} \sigma_i^{-2},
$$

$$
t_0^{-1} \leq t_{|M|}^{-1} = t_0^{-1} + \sum_{i=1}^{|M|} \sigma_i^{-2}(1 - e_i) \leq t_0^{-1} + \sum_{i=1}^{|M|} \sigma_i^{-2},
$$

and

$$
s_{|M|}^{-1} + t_{|M|}^{-1} = s_0^{-1} + t_0^{-1} + \sum_{i=1}^{|M|} \sigma_i^{-2}.
$$

Next we derive an upper bound for $\hat{\eta}$. We first derive an upper bound for the numerator in (6.8) (definition of K-elemental forward curvature), which is given by (6.23) on the next page.

We now derive a lower bound of the denominator in (6.8) by calculating the minimum value of the denominator over all possible A_j. It is easy to show that the minimum is achieved at $A_j = \text{Diag}(1,0)$ or $A_j = \text{Diag}(0,1)$:

$$\begin{aligned} f(M \oplus (A_j)) - f(M) &\geq \min(\log(1 + t_{|M|}\sigma_{|M|+1}^{-2}), \log(1 + s_{|M|}\sigma_{|M|+1}^{-2})) \\ &\geq \log(1 + \min(s_{|M|}\sigma_{|M|+1}^{-2}, t_{|M|}\sigma_{|M|+1}^{-2})) \\ &\geq \log(1 + (t_0^{-1} + \sum_{i=1}^{2K-2} \sigma_i^{-2})^{-1} \min_{i=1,\ldots,2K} \sigma_i^{-2}). \end{aligned}$$

Therefore, we can derive an upper bound for the K-elemental forward curvature as follows:

$$\hat{\eta} \leq \frac{\log \frac{1}{4}(1 + s_0 t_0^{-1} + s_0 \sum_{i=1}^{2K} \sigma_i^{-2})(1 + \frac{s_0^{-1} + \max_{i=1,\ldots,2K} \sigma_i^{-2}}{(t_0^{-1} + \sigma_1^{-2})})}{\log(1 + (t_0^{-1} + \sum_{i=1}^{2K-2} \sigma_i^{-2})^{-1} \min_{i=1,\ldots,2K} \sigma_i^{-2})}.$$

Using this upper bound, we can provide an approximation bound for the greedy strategy. We note that this upper bound is not extremely tight in the sense that it does not increase significantly with K only if s_0 or σ_i^{-2} are sufficiently small.

Now consider the case where σ_i takes an arbitrary value from $[a, b]$, where $0 < a < b$. In this case, by substituting either a or b appropriately in the inequality above, we get an upper bound for $\hat{\eta}$:

$$\hat{\eta} \leq \frac{\log \frac{1}{4}(1 + s_0 t_0^{-1} + 2 s_0 K a^{-2})(1 + \frac{s_0^{-1} + a^{-2}}{(t_0^{-1} + b^{-2})})}{\log(1 + t_0(1 + t_0(2K-2)a^{-2})^{-1} b^{-2})}.$$

With the above lower bounds for $\hat{\eta}$, we can use Theorem 6.3.2 to provide a bound for the greedy strategy. To apply Theorem 6.3.2, we need to have $f(G_i \oplus O) \geq f(O)$ for $i = 1, 2, \ldots, K-1$. Let $A^* \in \mathbb{D}$ be a greedy action. We will provide a sufficient condition such that $f((A^*) \oplus M) \geq f(M)$ for all M with length K. Suppose that $\sigma_i \in [a, b]$ for all i. Let $A^* = \text{Diag}(\sqrt{e^*}, \sqrt{1 - e^*})$ and $M = (A_1, \ldots, A_K)$, where $A_t = \text{Diag}(\sqrt{e_t}, \sqrt{1 - e_t})$ for all t. The inequality we need to verify can be written as

$$\begin{aligned} &\log(1 + s_0(\sigma_1^{-2} e^* + \sum_{t=1}^{K} \sigma_{t+1}^{-2} e_t))(1 + t_0(\sigma_1^{-2}(1 - e^*) + \sum_{t=1}^{K} \sigma_{t+1}^{-2}(1 - e_t))) \\ &\geq \log(1 + s_0(\sum_{t=1}^{K} \sigma_t^{-2} e_t))(1 + t_0(\sum_{t=1}^{K} \sigma_t^{-2}(1 - e_t))). \end{aligned} \tag{6.24}$$

We first calculate the value of e^*. It is easy to show that the objective function after applying (A^*) achieves the maximum when

$$e^* = \frac{1 + \frac{t_0^{-1} - s_0^{-1}}{\sigma_1^{-1}}}{2}.$$

Because e^* can only take values in $[0, 1]$, in the case where $(t_0^{-1} - s_0^{-1})/\sigma_1^{-1} \geq 1$, the maximum is achieved at $e^* = 1$. We will present our analysis only for this case—the analysis for the case where $(t_0^{-1} - s_0^{-1})/\sigma_1^{-1} < 1$ is similar and omitted. To show the above inequality (6.24), it suffices to show that

$$\begin{aligned}
&\log(1 + s_0\sigma_1^{-2} + (s_0 \sum_{t=1}^{K} \sigma_{t+1}^{-2} e_t))(1 + t_0(\sum_{t=1}^{K} \sigma_{t+1}^{-2}(1 - e_t))) \\
&\geq \log(1 + s_0(\sum_{t=1}^{K} \sigma_t^{-2} e_t))(1 + t_0(\sum_{t=1}^{K} \sigma_t^{-2}(1 - e_t))).
\end{aligned}$$

Removing the $\log$ on both sides of the inequality, we obtain

$$\begin{aligned}
&(1 + s_0 \sum_{t=1}^{K} \sigma_{t+1}^{-2} e_t)(1 + t_0 \sum_{t=1}^{K} \sigma_{t+1}^{-2}(1 - e_t)) + s_0\sigma_1^{-2}(1 + t_0 \sum_{t=1}^{K} \sigma_{t+1}^{-2}(1 - e_t)) \\
&\geq (1 + s_0 \sum_{t=1}^{K} \sigma_t^{-2} e_t)(1 + t_0 \sum_{t=1}^{K} \sigma_t^{-2}(1 - e_t)).
\end{aligned}$$

Rearranging terms, we obtain the following:

$$\begin{aligned}
& s_0 \sum_{t=1}^{K} e_t(\sigma_{t+1}^{-2} - \sigma_t^{-2}) + t_0 \sum_{t=1}^{K}(1 - e_t)(\sigma_{t+1}^{-2} - \sigma_t^{-2}) + s_0\sigma_1^{-2}(1 + t_0 \sum_{t=1}^{K} \sigma_{t+1}^{-2}(1 - e_t)) \\
& + s_0 t_0(\sum_{t=1}^{K} \sigma_{t+1}^{-2} e_t)(\sum_{t=1}^{K} \sigma_{t+1}^{-2}(1 - e_t)) - s_0 t_0(\sum_{t=1}^{K} \sigma_t^{-2} e_t)(\sum_{t=1}^{K} \sigma_t^{-2}(1 - e_t)) \\
\geq\; & s_0 \sum_{t=1}^{K}(\sigma_{t+1}^{-2} - \sigma_t^{-2})\mathbb{I}_t + t_0 \sum_{t=1}^{K}(\sigma_{t+1}^{-2} - \sigma_t^{-2})(1 - \mathbb{I}_t) \qquad (6.25) \\
& + s_0\sigma_1^{-2}(1 + t_0 \sum_{t=1}^{K} \sigma_{t+1}^{-2}(1 - e_t)) + s_0 t_0(b^{-4} - a^{-4})(\sum_{t=1}^{K} e_t)(\sum_{t=1}^{K}(1 - e_t)) \\
\geq\; & s_0(b^{-2} - a^{-2}) + s_0 b^{-2} + \frac{K^2}{4} s_0 t_0(b^{-4} - a^{-4}),
\end{aligned}$$

where $\mathbb{I}_t = 1$ if $\sigma_{t+1}^{-2} \leq \sigma_t^{-2}$ and $\mathbb{I}_t = 0$ if $\sigma_{t+1}^{-2} > \sigma_t^{-2}$.

From this we obtain a sufficient condition for $f((A^*) \oplus M) \geq f(M)$ to hold:

$$\frac{b^{-2}}{a^{-2} - b^{-2}} \geq \frac{K^2}{4} t_0(a^{-2} + b^{-2}) + 1.$$

We have shown before that the elemental forward curvature is not larger than 1 if and only if the noise variance is non-decreasing. Moreover, if the above inequality holds, which requires that either the length of the variance interval $[a, b]$ or K is sufficiently small, then we can get the $(1 - e^{-1})$ bound.

CHAPTER 7

CONCLUDING REMARKS

We have studied the binary hypothesis testing problem in the context of feedforwad and hierarchical tree networks. In feedfoward networks, we have considered two types of broadcast failures: erasure and flipping. In both cases, if the memory sizes are bounded, then there does not exist a decision strategy such that the error probability converges to 0. In the case of random erasure, if the memory size goes to infinity, then there exists a decision strategy such that the error probability converges to 0, even if the erasure probability converges to 1. We also characterize explicitly the relationship between the convergence rate of the error probability and the convergence rate of the memory. In the case of random flipping, if each node observes all the previous decisions, then with the myopic decision strategy, the error probability converges to 0, when the flipping probabilities are bounded away from $1/2$. In the case where the flipping probability converges to $1/2$, we derive a necessary condition on the convergence rate of the flipping probability such that the error probability converges to 0. We also characterize explicitly the relationship between the convergence rate of the flipping probability and the convergence rate of the error probability. Finally, we have derived a necessary condition such that the event herding has nonzero probability.

In hierarchical tree networks, we precisely describe the evolution of error probabilities in the (α, β) plane as we move up the tree. This allows us to deduce error probability bounds at the fusion center as functions of N under several different scenarios. These bounds show that the total error probability converges to 0 sub-exponentially, with an exponent that is essentially $\sqrt{N}$. In addition, we allow all sensors to be asymptotically crummy, in which case we deduce the necessary and sufficient conditions for the total error probability to converge to 0. All our results apply not only to the fusion center, but also to any other node in the tree network. In other words, we can similarly analyze a sub-tree inside the original tree network. We have also studied the social learning problem in the context of M-ary relay trees. We have analyzed the step-wise reductions of the Type I and Type II error probabilities and derived upper and lower bounds for each error probability at the root as explicit functions of N, which characterize the convergence rates for Type I, Type II, and the total error probabilities. We have shown that the majority dominance rule is not better than the Bayesian likelihood ratio test in terms of convergence rate. We have studied the convergence rate using the alternative majority dominance strategy. Last, we have provided a message-passing scheme which increases the convergence rate of the total error probability. We have shown quantitatively how the convergence rate varies with respect to the message alphabet sizes. This scheme is very efficient in terms of the average message size used for communication.

Our analysis leads to several open questions. We expect that our results can be extended to multiple hypotheses testing problem, paralleling a similar extension in tandem networks [6]. In the case of random flipping, we have not studied the case where the memory size goes to infinity but each node cannot observe all the previous decisions. We also want to generalize the techniques used in this thesis to more general network topologies. Moreover, besides erasure and flipping failures, we expect that our techniques can be used in the additive Gaussian noise scenario. With finite signal-to-noise ratios (SNR), the martingale convergence proof in Lemma 2.3.2 easily generalizes to this scenario. However, if SNR goes to 0 (e.g., the fading coefficient goes to 0, the noise variance goes to infinity, or the broadcasting signal power goes to 0), it is obvious that the convergence of error probability is not always true. We want to derive necessary and sufficient conditions on the convergence rate of SNR such that the error probability still converges to 0.

Social networks usually involve very complex topologies. For example, the degree of branching may vary among different agents in the network. The convergence rate analysis for general complex structures is still wide open.

Another question involves the assumption that the agent measurements are conditionally independent. It is of interest to study the scenario where these agent measurements are correlated. This scenario has been studied in the star configuration [109]–[111] but not in any other structures yet.

In the second topic of this thesis, we have introduced the notion of total forward, total backward, and elemental forward curvature for functions defined on strings. We have derived several variants of lower performance bounds, in terms of these curvature values, for the greedy strategy with respect to the optimal strategy. Our results contribute significantly to our understanding of the underlying algebraic structure of string submodular functions. Moreover, we have investigated two applications of string submodular functions with curvature constraints.

BIBLIOGRAPHY

[1] T. M. Cover, "Hypothesis testing with finite statistics," *Ann. Math. Statist.*, vol. 40, no. 3, pp. 828–835, Jun. 1969.

[2] M. E. Hellman and T. M. Cover, "Learning with finite memory," *Ann. Math. Statist.*, vol. 41, no. 3, pp. 765–782, Jun. 1970.

[3] P. Swaszek, "On the performance of serial networks in distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 254–260, Jan. 1993.

[4] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: Part I—Tandem structures," *IEEE Trans. Syst., Man and Cybern.*, vol. 21, no. 5, pp. 1044–1059, Sept./Oct. 1991.

[5] R. Viswanathan, S. C. A. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 366–376, Jul. 1988.

[6] J. Koplowitz, "Necessary and sufficient memory size for m-hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 44–46, Jan. 1975.

[7] J. D. Papastravrou and M. Athans, "Distributed detection by a large team of sensors in tandem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 639–653, Jul. 1992.

[8] V. V. Veeravalli, "Topics in decentralized detection," Ph.D. dissertation, Univ. Illinois, Urbana-Champaign, 1992.

[9] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the sub-exponential decay of detection error probabilities in long tandems," *IEEE Trans. Inform. Theory*, vol. 54, no. 10, pp. 4767–4771, Oct. 2008.

[10] I. Lobel, D. Acemoglu, M. A. Dahleh, and A. Ozdaglar, "Lower bounds on the rate of learning in social networks," in *Proc. IEEE American Control Conf.*, Hyatt Regency Riverfront, St. Louis, MO,pp. 2825–2830, Jun. 2009.

[11] K. Drakopoulos, A. Ozdaglar, and J. N. Tsitsiklis, "On learning with finite memory," *preprint*, available in arXiv:1209.1122, 2012.

[12] P. M. Djuric and Y. Wang, "Distributed Bayesian learning in multiagent systems: Improving our understanding of its capabilities and limitations," *IEEE Signal Process. Magazine*, vol. 29, no. 2, pp. 65–76, Mar. 2012.

[13] A. V. Banarjee, "A simple model for herd behavior," *Quart. J. Econ.*, vol. 107, no. 3, pp. 797–817, Aug. 1992.

[14] S. Bikchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as information cascades," *J. Political Econ.*, vol. 100, no. 5, pp. 992–1026, Oct. 1992.

[15] L. Smith and P. Sorensen, "Pathological outcomes of observational learning," *Econometrica*, vol. 68, no. 2, pp. 371–398, Mar. 2000.

[16] V. Krishnamurthy, "Bayesian sequential detection with phase-distributed change time and nonlinear penalty–A POMDP lattice programming approach," *IEEE Trans. Inform. Theory*, vol. 57, no. 10, pp. 7096–7124, Oct. 2011.

[17] V. Krishnamurthy, "Quickest detection POMDPs with social learning: Interaction of local and global decision makers," *IEEE Trans. Inform. Theory*, vol. 58, no. 8, pp. 5563–5587, Aug. 2012.

[18] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *Review Economic Studies*, vol. 78, no. 4, pp. 1201–1236, 2011.

[19] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games and Economic Behavior*, vol. 76, no. 1, pp. 210–225, Sep. 2012.

[20] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Tran. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, May. 2011.

[21] K. A. Jameson, M. C. Appleby, and L. C. Freeman, "Finding an appropriate order for a hierarchy based on probabilistic dominance," *Animal Behaviour*, vol. 57, no. 5, pp. 991–998, May. 1999.

[22] S. Wasserman, K. Faust, and D. Iacobucci, "Social network analysis: Methods and applications (structural analysis in the social sciences)," *Cambridge University Press*, Nov. 1994.

[23] D. J. Watts, P. S. Dodds, and M. E. J. Newman, "Identity and search in social networks," *Science*, vol. 296, no. 5571, pp. 1302–1305, May. 2002.

[24] A. Clauset, C. Moore, and M. E. J. Newman, "Hierarchical structure and the prediction of missing links in networks," *Nature*, vol. 453, pp. 98-101, May. 2008.

[25] A. S. Maiya and T. Y. Berger-Wolf, "Inferring the maximum likelihood hierarchy in social networks," in *Proc. 12th IEEE Intl. Conf. on Computational Science and Engineering (CSE '09)*, Vancouver, CA, Aug. 2009.

[26] Y. Kanoria and O. Tamuzy, "Tractable Bayesian social learning," in *Proc. IEEE Intl. Symp. on Inform. Theory*, Boston, MA, Jul. 2012.

[27] L. Lopez, J. F. F. Mendes, and M. A. F. Sanjuan, "Hierarchical social networks and information flow," *Phys. A*, vol. 316, no. 1–4, pp. 695–708, Dec. 2002.

[28] C. C. Eckel and R. K. Wilson, "Social learning in a social hierarchy: An experimental study," *Assoc. for the Advancement of Sci.*, Convention in Washington, D.C., Feb. 16–22, 2000.

[29] M. Gupte, P. Shankar, J. Li, S. Muthukrishnan, and L. Iftode, "Finding hierarchy in directed online social networks," in *Proc. of the 20th Intl. World Wide Web Conf. (WWW)*, Hyderabad, India, Mar. 28–Apr. 1, 2011.

[30] A. E. Motter, T. Nashikawa, and Y.-C. Lai, "Large-scale structural organization of social networks," *Phys. Rev. E*, vol. 68, no. 3, pp. 036105, Sep. 2003.

[31] J. D. Noh, "Exact scaling properties of a hierarchical network model," *Phys. Rev. E*, vol. 67, no. 4, pp. 045103, Apr. 2003.

[32] S. Choi, B. Celen, and S. Kariv, "Learning in networks: An experimental study," *UCLA Technical Report*, Dec. 2005.

[33] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, no. 4, pp. 501–510, Jul. 1981.

[34] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 1, pp. 98–101, Jan. 1986.

[35] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.

[36] J. N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.

[37] G. Polychronopoulos and J. N. Tsitsiklis, "Explicit solutions for some simple decentralized detection problems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 2, pp. 282–292, Mar. 1990.

[38] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Asymptotic performance of a censoring sensor network," *IEEE Trans. Inform. Theory*, vol. 53, no. 11, pp. 4191–4209, Nov. 2007.

[39] P. Willett and D. Warren, "The suboptimality of randomized tests in distributed and quantized detection systems," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 355–361, Mar. 1992.

[40] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I—Fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[41] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II—Advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.

[42] T. M. Duman and M. Salehi, "Decentralized detection over multiple-access channels," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 2, pp. 469–476, Apr. 1998.

[43] B. Chen and P. K. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 693–699, Feb. 2005.

[44] B. Liu and B. Chen, "Channel-optimized quantizers for decentralized detection in sensor networks," *IEEE Trans. Inform. Theory*, vol. 52, no. 7, pp. 3349–3358, Jul. 2006.

[45] B. Chen and P. K. Varshney, "A Bayesian sampling approach to decision fusion using hierarchical models," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1809–1818, Aug. 2002.

[46] A. Kashyap, "Comments on on the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1274–1275, Mar. 2006.

[47] H. Chen, B. Chen, and P. K. Varshney, "Further results on the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 2, pp. 828–832, Feb. 2009.

[48] G. Fellouris and G. V. Moustakides, "Decentralized Sequential Hypothesis Testing Using Asynchronous Communication," *IEEE Trans. Inform. Theory*, vol. 57, no. 1, pp. 534–548, Jan. 2011.

[49] J. A. Gubner, L. L. Scharf, and E. K. P. Chong, "Exponential error bounds for binary detection using arbitrary binary sensors and an all-purpose fusion rule in wireless sensor networks," in *Proc. IEEE International Conf. on Acoustics, Speech, and Signal Process.*, Taipei, Taiwan, Apr. 19-24 2009, pp. 2781–2784.

[50] P. K. Varshney, *Distributed detection and data fusion*, Springer-Verlag, New York, NY, 1997.

[51] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: Part II—Tree structures," *IEEE Trans. Syst., Man and Cybern.*, vol. 23, no. 1, pp. 211–221, Jan./Feb. 1993.

[52] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Data fusion trees for detection: Does architecture matter?," *IEEE Trans. Inform. Theory*, vol. 54, no. 9, pp. 4155–4168, Sept. 2008.

[53] A. R. Reibman and L. W. Nolte, "Design and performance comparison of distributed detection networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 6, pp. 789–797, Nov. 1987.

[54] W. P. Tay and J. N. Tsitsiklis, "Error exponents for decentralized detection in tree networks," in *Networked Sensing Information and Control*, V. Saligrama, Ed., Springer-Verlag, New York, NY, 2008, pp 73–92.

[55] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Bayesian detection in bounded height tree networks," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4042–4051, Oct. 2009.

[56] A. Pete, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks with multiple event structures," *IEEE Trans. Autom. Control*, vol. 39, no. 8, pp. 1702–1707, Aug. 1994.

[57] O. P. Kreidl and A. S. Willsky, "An efficient message-passing algorithm for optimizing decentralized detection networks," *IEEE Trans. Autom. Control*, vol. 55, no. 3, pp. 563–578, Mar. 2010.

[58] S. Alhakeem and P. K. Varshney, "A unified approach to the design of decentralized detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 9–20, Jan. 1995.

[59] Y. Lin, B. Chen, and P. K. Varshney, "Decision fusion rules in multi-hop wireless sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 2, pp. 475–488, Apr. 2005.

[60] J. A. Gubner, E. K. P. Chong, and L. L. Scharf, "Aggregation and compression of distributed binary decisions in a wireless sensor network," in *Proc. Joint 48th IEEE Conf. on Decision and Control and 28th Chinese Control Conf.*, Shanghai, P. R. China, Dec. 16-18 2009, pp. 909–913.

[61] Y. Kanoria and A. Montanari, "Subexponential convergence for information aggregation on regular trees," in *Proc. Joint 50th IEEE Conf. on Decision and Control and European Control Conf.*, Orlando, FL, Dec. 12-15 2011, pp. 5317–5322.

[62] H. Ebel, L.-I. Mielsch, and S. Bornholdt, "Scale-free topology of e-mail networks," *Phys. Rev. E*, vol. 66, no. 3, pp. 035103, Sep. 2002.

[63] S.-H. Yook, H. Jeong, and A.-L. Barabasi, "Modeling the Internet's large-scale topology," *Proc. Natl. Acad. Sciences*, vol. 99, no. 21, pp. 13382–13386, Oct. 2002.

[64] M. E. J. Newman, D. J. Watts, and S. H. S. Strogatz, "Random graph models of social networks," *Proc. Natl. Acad. Sciences*, vol. 99, no. Suppl 1, pp. 2566–2572, Feb. 2002.

[65] T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley and Sons, New York, NY, 2006.

[66] C. P. Chamley, *Rational herds*, New York: Cambridge, 2004.

[67] J. Surowiecki, *The wisdom of crowds*, New York: Doubleday, 2005.

[68] R. Durrett, *Probability: theory and examples*, Duxbury Advanced Series, 1996.

[69] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley and Sons, New York, NY, 1968.

[70] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the impact of node failures and unreliable communications in dense sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2535–2546, Jun. 2008.

[71] Z. Zhang, E. K. P. Chong, A. Pezeshki, W. Moran, and S. D. Howard, "Submodularity and optimality of fusion rules in balanced binary relay trees," in *Proc. IEEE 51st Annual Conference on Decision and Control*, Maui, HI, Dec. 2012, pp. 3802–3807.

[72] D. P. Bertsekas, *Dynamic programming and optimal control*, Athena Scientific, 2000.

[73] T. M. Cover, "Universal portfolios," *Math. Finance*, vol. 1, no. 1, pp. 1–29, Jan. 1991.

[74] E. K. P. Chong, C. M. Kreucher, and A. O. Hero III, "Partially observable markov decision process approximations for adaptive sensing," *Discrete Event Dynamic Systems*, vol. 19, no. 3, pp. 377–422, Sep. 2009.

[75] E. Mossel and S. Roch, "On the submodularity of influence in social networks," in *Proc. 39th Ann. ACM Symp. Theory Computing*, San Diego, California, USA, Jun. 2007, pp. 128–134.

[76] W. Tutte, "Lectures on matroids," *J. Res. Nat. Bur. Standards Sect. B*, vol. 69, no. 1-47, p. 468, 1965.

[77] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions—i," *Math. Program.*, vol. 14, no. 1, pp. 265–294, 1978.

[78] M. L. Fisher, G. L. Nemhauser, and L. A. Wolsey, "An analysis of approximations for maximizing submodular set functions—ii," *Math. Program.*, vol. 8, pp. 73–87, 1978.

[79] M. Conforti and G. Cornuejols, "Submodular set functions, matroids and the greedy algorithm: tight worst-case bounds and some generalizations of the rado-edmonds theorem," *Discrete Applied Mathematics*, vol. 7, no. 3, pp. 251–274, 1984.

[80] J. Vondrák, "Submodularity and curvature: The optimal algorithm," *RIMS Kokyuroku Bessatsu B*, vol. 23, pp. 253–266, 2010.

[81] Z. Wang, W. Moran, X. Wang, and Q. Pan, "Approximation for maximizing monotone non-decreasing set functions with a greedy method," *preprint*.

[82] M. Streeter and D. Golovin, "An online algorithm for maximizing submodular functions," in *Proc. 22nd Ann. Conf. Neural Inform. Processing Systems*, Dec. 2008.

[83] S. Alaei and A. Malekian, "Maximizing sequence-submodular functions and its application to online advertising," *arXiv preprint arXiv:1009.4153*, 2010.

[84] D. Golovin and A. Krause, "Adaptive submodularity: Theory and applications in active learning and stochastic optimization," *J. Artificial Intel. Research*, vol. 42, no. 1, pp. 427–486, Sep. 2011.

[85] N. Buchbinder, M. Feldman, J. S. Naor, and R. Schwartz, "A tight linear time (1/2)-approximation for unconstrained submodular maximization," in *Proc. IEEE 53rd Ann. Symp. Foundations of Computer Science*, New Brunswick, NJ, 2012, pp. 649–658.

[86] G. Calinescu, C. Chekuri, M. Pál, and J. Vondrák, "Maximizing a monotone submodular function subject to a matroid constraint," *SIAM J. Computing*, vol. 40, no. 6, pp. 1740–1766, 2011.

[87] D. Chakrabarty and G. Goel, "On the approximability of budgeted allocations and improved lower bounds for submodular welfare maximization and gap," *SIAM J. Computing*, vol. 39, no. 6, pp. 2189–2211, 2010.

[88] J. Vondrák, C. Chekuri, and R. Zenklusen, "Submodular function maximization via the multilinear relaxation and contention resolution schemes," in *Proc. 43rd Ann. ACM Symp. Theory of Computing*, 2011, pp. 783–792.

[89] S. Dobzinski and M. Schapira, "An improved approximation algorithm for combinatorial auctions with submodular bidders," in *Proc. 17th ACM-SIAM Symp. on Discrete Algorithm*, 2006, pp. 1064–1073.

[90] U. Feige, "A threshold of $\ln n$ for approximating set cover," *J. ACM*, vol. 45, no. 4, pp. 634–652, Jul. 1998.

[91] U. Feige and J. Vondrak, "Approximation algorithms for allocation problems: Improving the factor of $1 - 1/e$," in *Proc. 47th IEEE Symposium on Foundations of Computer Science*, 2006, pp. 667–676.

[92] U. Feige and J. Vondrák, "The submodular welfare problem with demand queries," *Theory of Computing*, vol. 6, pp. 247–290, 2010.

[93] Y. Filmus and J. Ward, "A tight combinatorial algorithm for submodular maximization subject to a matroid constraint," in *Proc. IEEE 53rd Annual Symposium on Foundations of Computer Science*, New Brunswick, NJ, 2012, pp. 659–668.

[94] A. Kulik, H. Shachnai, and T. Tamir, "Maximizing submodular set functions subject to multiple linear constraints," in *Proc. 20th ACM-SIAM Symposium on Discrete Algorithms*, 2009, pp. 545–554.

[95] J. Lee, V. S. Mirrokni, V. Nagarajan, and M. Sviridenko, "Non-monotone submodular maximization under matroid and knapsack constraints," in *Proc. ACM Symposium on Theory of Computing*, 2009, pp. 323–332.

[96] J. Lee, M. Sviridenko, and J. Vondrák, "Submodular maximization over multiple matroids via generalized exchange properties," *Math. Operations Research*, vol. 35, no. 4, pp. 795–806, 2010.

[97] G. L. Nemhauser and L. A. Wolsey, "Best algorithms for approximating the maximum of a submodular set function," *Math. Operations Research*, vol. 3, no. 3, pp. 177–188, 1978.

[98] M. Sviridenko, "A note on maximizing a submodular set function subject to a knapsack constraint," *Operations Research Letters*, vol. 32, no. 1, pp. 41–43, Jan. 2004.

[99] J. Vondrak, "Optimal approximation for the submodular welfare problem in the value oracle model," in *Proc. 40th ACM Symposium on Theory of Computing*, 2008, pp. 67–74.

[100] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *Proc. 49th IEEE Conference on Decision and Control*, 2010, pp. 2572–2577.

[101] A. A. Ageev and M. I. Sviridenko, "Pipage rounding: A new method of constructing algorithms with proven performance guarantee," *Journal of Combinatorial Optimization*, vol. 8, no. 3, pp. 307–328, 2004.

[102] U. Feige, L. Lovász, and P. Tetali, "Approximating min sum set cover," *Algorithmica*, vol. 40, no. 4, pp. 219–234, 2004.

[103] K. Munagala, S. Babu, R. Motwani, and J. Widom, "The pipelined set cover problem," in *Proc. 10th International Conference on Database Theory*, 2005, pp. 83–98.

[104] D. Kempe, J. Kleinberg, and É. Tardos, "Influential nodes in a diffusion model for social networks," in *Proc. 32nd International Conference on Automata, Languages and Programming*, 2005, pp. 1127–1138.

[105] J. Lu and T. Suda, "Coverage-aware self-scheduling in sensor networks," in *Proc. IEEE 18th Annual Workshop on Computer Communications*, 2003, pp. 117–123.

[106] E. Liu, E. K. P. Chong, and L. L. Scharf, "Greedy adaptive measurements with signal and measurement noise," in *Proc. 46th Asilomar Conference on Signals, Systems and Computers*, 2012, pp. 1229–1232.

[107] W. R. Carson, M. Chen, M. R. Rodrigues, R. Calderbank, and L. Carin, "Communications-inspired projection design with application to compressive sensing," *SIAM Journal on Imaging Sciences*, vol. 5, no. 4, pp. 1185–1212, 2012.

[108] A. Krause and C. Guestrin, "Near-optimal observation selection using submodular functions," in *Proc. National Conference on Artificial Intelligence*, vol. 22, no. 2, 2007, pp. 1650–1654.

[109] J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?" *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5099–5106, Nov. 2006.

[110] W. Li and H. Dai, "Distributed detection in large-scale sensor networks with correlated sensor observations," in *Proc. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 2005.

[111] H. Chen, B. Chen, and P. K. Varshney, "A new framework for distributed detection with conditionally dependent observations," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1409–1419, Mar. 2012.