1-1-2015

Online and Statistical Learning in Networks

Shahin Shahrampour
University of Pennsylvania, shahin.shahrampour@gmail.com

Recommended Citation

Shahrampour, Shahin, "Online and Statistical Learning in Networks" (2015). *Publicly Accessible Penn Dissertations*. 1999.
http://repository.upenn.edu/edissertations/1999

Online and Statistical Learning in Networks

Abstract

Learning, prediction and identification has been a main topic of interest in science and engineering for many years. Common in all these problems is an agent that receives the data to perform prediction and identification procedures. The agent might process the data individually, or might interact in a network of agents. The goal of this thesis is to address problems that lie at the interface of statistical processing of data, online learning and network science with a focus on developing distributed algorithms. These problems have wide-spread applications in several domains of systems engineering and computer science. Whether in individual or group, the main task of the agent is to understand how to treat data to infer the unknown parameters of the problem. To this end, the first part of this thesis addresses statistical processing of data. We start with the problem of distributed detection in multi-agent networks. In contrast to the existing literature which focuses on asymptotic learning, we provide a finite-time analysis using a notion of Kullback-Leibler cost. We derive bounds on the cost in terms of network size, spectral gap and relative entropy of data distribution. Next, we turn to focus on an inverse-type problem where the network structure is unknown, and the outputs of a dynamics (e.g. consensus dynamics) are given. We propose several network reconstruction algorithms by measuring the network response to the inputs. Our algorithm reconstructs the Boolean structure (i.e., existence and directions of links) of a directed network from a series of dynamical responses. The second part of the thesis centers around online learning where data is received in a sequential fashion. As an example of collaborative learning, we consider the stochastic multi-armed bandit problem in a multi-player network. Players explore a pool of arms with payoffs generated from player-dependent distributions. Pulling an arm, each player only observes a noisy payoff of the chosen arm. The goal is to maximize a global welfare or to find the best global arm. Hence, players exchange information locally to benefit from side observations. We develop a distributed online algorithm with a logarithmic regret with respect to the best global arm, and generalize our results to the case that availability of arms varies over time. We then return to individual online learning where one learner plays against an adversary. We develop a fully adaptive algorithm that takes advantage of a regularity of the sequence of observations, retains worst-case performance guarantees, and performs well against complex benchmarks. Our method competes with dynamic benchmarks in which regret guarantee scales with regularity of the sequence of cost functions and comparators. Notably, the regret bound adapts to the smaller complexity measure in the problem environment.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Electrical & Systems Engineering

First Advisor

Ali Jadbabaie

ONLINE AND STATISTICAL LEARNING IN NETWORKS

Shahin Shahrampour

A DISSERTATION

in

Electrical and Systems Engineering

Presented to the Faculties of the University of Pennsylvania
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

2015

Ali Jadbabaie, Supervisor of Dissertation
Professor of Electrical and Systems Engineering

Alexander Rakhlin, Co-Supervisor of Dissertation
Associate Professor of Statistics

Alejandro Ribeiro, Graduate Group Chairperson
Associate Professor of Electrical and Systems Engineering

Dissertation Committee:

Angelia Nedich, Associate Professor of Industrial and Enterprise Systems Engineering

Victor M. Preciado, Assistant Professor of Electrical and Systems Engineering

Karthik Sridharan, Assistant Professor of Computer Science

ONLINE AND STATISTICAL LEARNING IN NETWORKS

Acknowledgments

Starting this part of the thesis, I realized that I must thank so many people. As I have a short memory, I would like to apologize in advance to anyone whom I forgot to mention in the coming text. I will thank them in person, and buy them a Nutella (I believe chocolates make us happy).

First and foremost, I would like to thank my advisors, Professors Ali Jadbabaie and Alexander Rakhlin, for their support and encouragement during these years. Ali helped me to become an independent researcher while allowing me to explore several research areas to find my interest. He taught me how to think about research problems, and to present my ideas. Whether about science or not, I always learned a lot from our discussions which helped me to grow as a person. I was exposed to the world of online learning by Sasha. In every single meeting we had, he patiently taught me a lot of new things. He helped me to understand the right way to approach problems, and to have high standards in research. This thesis could not have been written without the guidance of Ali and Sasha.

I would like to thank my committee members, Professors Angelia Nedich, Victor M. Preciado and Karthik Sridharan for their thoughtful comments and suggestions. I am greatly indebted to Victor who helped me to develop my research skills. It was only under his supervision that Chapter 3 could be written. I am also indebted to Karthik for helpful discussions over various topics of learning theory. Researching alongside him was a valuable experience for me, and Chapter 5 would have not been possible without his guidance. I would also like to thank Mohammad Amin Rahimian for his contributions in Section 2.5.2.

I start this paragraph with a routine sentence (a fact that we sometimes take for granted): true friends are pearls beyond price. There are so many friends to whom I owe credit, so I avoid the controversy with an alphabetical order. Special thanks to Afshin, Aryan, Behnam, Mahyar, Mehdi, Nikolay, Sara, Shermin, Solmaz, Tahmineh and many others. They all helped me a lot to have fun and enjoy my life in Philly. I already know how much I will miss them all.

It is now time to go back to where everything started. I express my heartfelt gratitude to my parents, Soraya and Ali, who have given me their unconditional love and support throughout my life. I cannot thank them enough for being such caring parents, and am sure I will have a hard time mimicking their path if I ever become a father. I am also very grateful to my sisters, Shabnam and Shiva, for their helps and encouragements during all these years.

Last but not least, I must thank my wonderful wife, Mahsa Zarei. I am extremely lucky to have met Mahsa in the right place at the right time. She has given me nothing but love, all the time we have spent together. Student life is not always fun, but she managed to put up with everything to make sure we do not spare a minute not enjoying our moments. Mahsa, thank you for everything that we achieved, and will achieve together.

ABSTRACT

ONLINE AND STATISTICAL LEARNING IN NETWORKS

Shahin Shahrampour

Ali Jadbabaie

Alexander Rakhlin

Learning, prediction and identification has been a main topic of interest in science and engineering for many years. Common in all these problems is an agent that receives the data to perform prediction and identification procedures. The agent might process the data individually, or might interact in a network of agents. The goal of this thesis is to address problems that lie at the interface of statistical processing of data, online learning and network science with a focus on developing distributed algorithms. These problems have wide-spread applications in several domains of systems engineering and computer science. Whether in individual or group, the main task of the agent is to understand how to treat data to infer the unknown parameters of the problem. To this end, the first part of this thesis addresses statistical processing of data. We start with the problem of distributed detection in multi-agent networks. In contrast to the existing literature which focuses on asymptotic learning, we provide a finite-time analysis using a notion of Kullback-Leibler cost. We derive bounds on the cost in terms of network size, spectral gap and relative entropy of data distribution. Next, we turn to focus on an inverse-type problem where the network structure is unknown, and the outputs of a dynamics (e.g. consensus dynamics) are given. We propose several network reconstruction algorithms by measuring the network response to the inputs. Our algorithm reconstructs the Boolean structure (i.e., existence and directions of links) of a directed network from a series of dynamical responses. The second part of the thesis centers around online learning where data is received in a sequential fashion. As an example of collaborative learning, we consider the stochastic multi-armed bandit problem in a multi-player network. Players explore a pool of arms with payoffs generated from player-dependent distributions. Pulling an arm, each player only observes a noisy payoff of the chosen arm. The goal is to maximize a global welfare or to find the best global arm. Hence, players exchange information locally to benefit from side observations. We develop a distributed online algorithm with a logarithmic regret with respect to the best global arm, and generalize our results to the case that availability of arms varies over time. We then return to individual online learning where one learner plays against an adversary. We develop a fully adaptive algorithm that takes advantage of a regularity of the sequence of observations, retains worst-case performance guarantees, and performs well against complex benchmarks. Our method competes with dynamic benchmarks in which regret guarantee scales with regularity of the sequence of cost functions and comparators. Notably, the regret bound adapts to the smaller complexity measure in the problem environment.

Contents

List of Tables

List of Figures

Chapter 1

Overview

In recent years learning, prediction and identification have become a main topic of interest in science and engineering. In these fields, it is important to understand data, and process it efficiently. Many scenarios in everyday life could be categorized as such. How can one make a smart choice when buying a product? How can one beat the traffic to commute to work on time? How can one guess the friendship network of a few individuals based on their behavior? These questions and many mores motivate us to better understand the signals around us to be able to make a wise decision or prediction. Common in all these problems is a learner or designer that attempts to incorporate available data in a sensible way to develop prediction and identification procedures. Data could either be generated arbitrarily or follow some statistical model, where in the latter the leaner can take advantage of statistical properties to design more efficient algorithms.

Of particular interest in these problems are those in which the learner should interact in a group to obtain missing data dispersed throughout a network. Network science has gained a growing popularity over the past few years [1–3]. This discipline serves the goal of studying interactions among individuals (e.g. using graphs to model networks mathematically). Interesting problems arising in sensor, social and economic networks have attracted attention of scientists in many different fields.

The goal of this thesis is to address problems that lie at the interface of *statistical processing* of data, *online learning* and *network science* with a focus on developing *distributed* algorithms. These problems have wide-spread applications in many domains of engineering such as distributed estimation, optimization and machine learning. Whether in individual or group, the main challenge is to understand how to treat data to infer the unknown of the problem. We contribute to these emerging fields by providing algorithms that use the *statistical* or *online* nature of data to tackle the problem.

To this end, the first part of this thesis addresses statistical processing of data. We start with the problem of distributed detection in multi-agent networks. Distributed detection has gained a considerable attention in the past few decades. The problem has wide range of applications from sensor networks to social and economic networks. We propose an information aggregation scheme where agents collaborate with each other to perform a team task. More formally, agents receive private signals about an unknown state of the world. The underlying state is globally identifiable, yet informative signals may be dispersed throughout the network. Using an optimization-based framework, we develop an iterative local strategy for agents. To measure the efficiency of our local update, we compare it to its global counter part using a notion of Kullback-Leibler cost.

In contrast to the existing literature which focuses on asymptotic learning, we provide a finite-time analysis. We derive bounds on the cost in terms of network size, spectral gap, centrality of each agent and relative entropy of agents' signal structures. We further prove convergence of beliefs in fixed and switching network topologies.

Next, we turn to focus on an inverse-type problem where the network structure is unknown, but the outputs of a dynamics (e.g. consensus dynamics) are given. We propose several reconstruction algorithms by measuring the cross-power spectral densities of the network response to the inputs. Our algorithm reconstructs the Boolean structure (i.e., existence and directions of links) of a directed network from a series of dynamical responses. Moreover, we propose a second algorithm to recover the exact structure of the network (including edge weights), when an eigenvalue-eigenvector pair of the connectivity matrix is known (for example, Laplacian connectivity matrices).

Finally, for the particular cases of nonreciprocal networks (i.e., networks with no directed edges pointing in opposite directions) and undirected networks, we propose specialized algorithms that result in a lower computational cost.

The second part of the thesis centers around online learning where the learner receives data in a sequential fashion. As an example of collaborative learning, we consider the stochastic multi-armed bandit problem in a multi-player network. Players explore a pool of arms with payoffs generated from player-dependent distributions. Pulling an arm, each player only observes a noisy payoff of the chosen arm. The goal is to maximize a global welfare in the sense of competing with the arm with highest average payoff among players, i.e. to find the best global arm. To achieve this goal, players (confined to a network structure) exchange information locally to benefit from side observations. We use this model to develop a distributed online algorithm with a logarithmic regret with respect to the best global arm. The algorithm can be generalized to deal with sleeping bandits where availability of arms varies over time. The regret in that context is defined with respect to the best-ordering benchmark. Our algorithms are optimal in the sense that in a complete network they scale down the regret of their single-player counterpart by network size. We demonstrate the application of the results in the context of distributed detection in sensor networks.

We then return to individual online learning where one learner plays against an adversary. We develop a fully adaptive algorithm that takes advantage of a regularity of the sequence of observations, retains worst-case performance guarantees, and performs well against complex benchmarks. Our method competes with dynamic benchmarks in which regret guarantee scales with regularity of the sequence of cost functions and comparators. Notably, the regret bound adapts to the smaller complexity measure in the problem environment. Finally, we apply our results to drifting zero-sum, two-player games where both players achieve no regret guarantees against best sequences of actions in hindsight.

1.1 Statistical Processing of Data in Networks

In many learning and identification problems, the learner deals with data samples that have certain *statistical* characteristics. For instance, the leaner encounters a stream of i.i.d. signals, or has knowledge about the power spectrum of signals. Then, these properties can be used to infer about the *unknown* of the problem. There are many problems that can be statistically modeled with wide-range of applications.

Consider a group of friends in which one person wants to buy a cell phone. The person can always obtain information through ads, websites and other sources (private signals). Using these signals, she might not be able to figure out the best option in isolation. However, as a part of the friendship group, she can discuss her options with her friends to benefit from side observations, and find out the best cell phone.

In the scenario we just described, we assumed that the network structure is *given*, and the outcome is *unknown*. What if we *know* of output properties, and the network structure is *unknown*? Can we use this information to identify the topology of the friendship network? Inverse problems have a long history going back to several decades ago. In these problems, the goal is to characterize the network structure given some input-output measurements. Complex dynamical networks have attracted considerable attention in recent years. The power grid, the Internet, the World Wide Web, as well as many other biological, social and economic networks, are examples of networked dynamical systems that motivate this interest.

Distributed Detection in Fixed and Switching Topologies

Decentralized detection, optimization and observational social learning has been an intense focus of research over the past three decades with applications ranging from sensor networks to social and economic networks [4–10]. In these frameworks the computation burden is distributed among agents of a network, allowing them to achieve a *global* task using *local* information. Developments in distributed optimization [9–13] have opened new venues to investigate principled distributed detection. Viewing with an optimization lens, one can think of the problem as minimizing a network loss that is a sum of individual losses. Using linear losses, the problem coincides with distributed detection, where the goal is to identify an *unknown* true state of the world.

We formalize this idea in Chapter 2, and propose a distributed detection algorithm. Observing private signals (individual stochastic gradients), agents use purely local interactions to detect the true state of the world which belongs to a finite state space. The main objective of the chapter is to address the *finite-time* analysis of distributed detection and the impact of *network topology*.

To characterize the efficiency, we compare our algorithm to its centralized counterpart. More specifically, consider an individual agent i that forms a probability distribution $\mu_{i,t}$ over the state space at time t. Also, let μ_t be the probability distribution that agent would have formed, had it had access to observations of others (at time t). Our goal is to analyze the following objective

$$\mathbf{Cost}_{i,T} = \sum_{t=1}^{T} D_{KL}(\mu_{i,t} \| \mu_t),$$

where $D_{KL}(\cdot\|\cdot)$ denotes the KL-divergence. We show that the cost can be bounded uniformly in time in terms of *relative entropy of signals*, *agents centralities* and *network spectral gap*. Benefiting from this fact, we show how one can speed up learning by designing an optimal network. We further prove convergence results about following *time-varying* networks:

- Gossip protocol: in this scenario, the underlying topology of the network varies based on a gossip communication rule between agents. At each time one agent is picked randomly, and the selected agent communicates with a random agent in its neighborhood.

- Information-based switching: we study a more communication-efficient version of switching rules where agents need not interact with each other at every round. In fact, they only communicate when their private signals are not informative enough. We measure the signal information by total variation distance of the *posterior* belief from the *prior*. The hardness of the problem stems from the fact that the communication protocol is signal-dependent.

In both cases, we prove the *almost sure* convergence of the beliefs to the true state, and characterize the asymptotic rate in terms of relative entropy of signals.

Inverse Problem : Network Identification

In many distributed, information aggregation procedures, the focus is more on the aggregation method rather than the network structure. In fact, the network structure is usually *given*, and the algorithm outputs an *estimate* or *prediction* accordingly. However, one can also consider an inverse problem where the outputs of an update (say a consensus algorithm where individuals eventually converge to a common opinion) are *given*, and the network structure is *unknown*. The question is whether we can reconstruct the topology of the network based on output measurements.

To this end, Chapter 3 addresses topology identification of networked dynamical systems. We consider reconstruction of *directed* networks in the presence of *intrinsic noise* with *unknown* power spectral density. We propose a method which builds on the grounding procedure. When a node is grounded, it broadcasts zero without being removed from the network. Sequentially grounding the nodes, the cross-power spectral densities for each pair of nodes are measured. The relationship between the power spectra ends up being a function of network structure, allowing us to identify the network topology. While our method can reconstruct directed networks, it incurs a lower computational cost when the network is undirected or nonreciprocal. In particular, this work can solve the reconstruction of LTI systems running a consensus dynamics.

1.2 Individual and Collaborative Online Learning

The term *online* is roughly used when a learner, predictor or designer, performs the corresponding task in a sequential fashion. The main challenge in online settings is to develop efficient methods that take advantage of the past history to predict the future. One can immediately observe that the strength of these algorithms is their functionality without accessing the entire data set. Given their power, it is not a surprise that online learning algorithms have received a considerable attention in computer science, machine learning and statistics over the past few years.

To better understand the application of online learning we start with a few motivating examples. Consider a person (learner) whose job is to place ads on a website, say, for a particular type of user. The learner attempts to place an ad in which the user is interested. Indeed, before placing it on the website, the learner does not know whether the user will click on the ad. However, after a few rounds, the learner might get a sense of user's interests. For example, the learner can notice whether the user is interested in sports, music, traveling and etc., and offer those contexts to the user. The game between the learner and user is an instance of individual online learning.

As another example, one can think of a person (learner) who commutes to work every day.

Using past experience, she decides on a route every morning, and the goal is to get to office in the shortest possible time (in the long run). Indeed, the learner does not receive any information about the unchosen paths. In this problem the traffic pattern might not follow any specific distribution, but the learner has the chance to predict the future traffic patterns based on the past. This problem can be modeled as an online shortest path problem.

On the other hand, online learning could also be studied in multi-player frameworks. Consider a group of sensors (players) that measure the location of a finite number of targets. Each sensor contacts one target per time step, and can only measure a specified coordinate of its position. The target reveals a noisy version of the coordinate to the sensor, and the noise characteristics are different among sensors. They aim to track the closest target to the origin, and with one coordinate at hand, sensors must communicate with each other to supplement their imperfect observations.

Motivated by these examples and many others, we dedicate the second part of this thesis to problems in the domain of online learning. We address both one-player and multi-player settings in different contexts.

Multi-Armed Bandits in Multi-Agent Networks

The multi-armed bandit (MAB) problem has been extensively studied in the literature [14–18]. The problem, defined by a set of arms or actions, captures the exploration-exploitation dilemma for a learner. At each time step, the learner chooses an arm and receives its corresponding payoff or reward. In stochastic MAB, the reward sequence (the data) is assumed to be iid (non-iid rewards have also been addressed in the literature). The objective is to maximize the total payoff obtained from sequentially selecting the arms. Equivalently, the learner aims to minimize regret when competing with the best single arm in hindsight. While early studies on MAB dates back to nine decades ago, the problem has received considerable attention due to its modern applications. MAB could be an instance of sequential decision making for ad placement, website optimization or packet routing.

In Chapter 4, we address the *stochastic* MAB in a *multi-player* network. Players explore a pool of arms with payoffs generated from *player-dependent* distributions. Pulling an arm, each player only observes a noisy payoff of the chosen arm. The goal is to maximize a global welfare in the sense of competing with the arm with highest average payoff among players, i.e. to find the best *global* arm. To achieve this goal, players (confined to a network structure) exchange information locally to benefit from side observations.

The main contribution of the chapter is to develop a distributed online algorithm with a logarithmic regret with respect to the best global arm. The method is a variant of the celebrated `UCB1` algorithm in which the confidence bound relies on the network characteristics. The algorithm can be generalized to deal with sleeping bandits where availability of arms varies over time. The regret in that context is defined with respect to the best-ordering benchmark. Proposed algorithms are optimal in the sense that in a complete network they scale down the regret of their single-player counterpart by network size. We demonstrate the application of the results in the context of distributed detection in sensor networks.

Online Optimization in Dynamic Environments

Apart from distributed (online) detection, *one-player* online learning is also a popular area of interest in the literature of learning theory [19–21]. The problem models *sequential* decision making used in wide spectrum of real-world applications. Early works on online learning started with the problem of prediction with expert advice, and the topic has been expansively studied ever since.

Online learning / optimization is modeled as a game between a *learner* and an *adversary* where the learner sequentially chooses an action at each round, and the adversary in turn reveals a loss to the learner. Typically, the goal is to minimize the *static* regret defined with respect to the best single action in hindsight. In other words, the static regret is the difference between the accumulated loss versus the smallest possible loss (achieved with one single action) had the learner been aware of the entire loss sequence *a priori*. The literature has witnessed a series of works developing no-(static)regret algorithms. Perhaps less well-known, is the notion of *dynamic* regret where the learner competes against the best action of each round. Indeed, aiming for this stringent benchmark is only possible under certain regularity conditions.

In Chapter 5, we pose an online learning problem where the learner selects action x_t and incurs the loss $f_t(x_t)$ at time t. The goal is to compete with the best action of each round, or to minimize the *dynamic* regret

$$\mathbf{Reg}_T^d = \sum_{t=1}^{T} f_t(x_t) - f_t(x_t^*),$$

where x_t is the minimizer of $f_t(\cdot)$ over a convex set $\mathcal{X}$. Our main tools to bound the dynamic regret are three complexity measures: temporal variability of the loss sequence V_T, deviation of gradients from a predictable sequence D_T and regularity in the pattern of minima sequence C_T. We then prove the following bound on the dynamic regret,

$$\mathbf{Reg}_T^d \leq \tilde{\mathcal{O}}\left(\sqrt{D_T+1}\right) + \tilde{\mathcal{O}}\left(\min\left\{\sqrt{(D_T+1)C_T}, (D_T+1)^{1/3}T^{1/3}V_T^{1/3}\right\}\right).$$

The algorithm is *adaptive* in the sense that the learner needs no prior knowledge of the environment. Unlike the stationary setting, the algorithm uses a *non-monotone*, adaptive step size tuned based on a *doubling trick*. The intuition behind the choice of non-monotone step size lies under non-stationarity of the environment where the learner needs to discard some information from the past. Interestingly, combining these complexity measures allows the learner to adapt to the best measure.

Part I

Statistical Processing of Data in Networks

Chapter 2

Distributed Detection in Fixed and Switching Network Topologies

Recent years have witnessed an intense interest on *distributed* detection, estimation, prediction and optimization [4–10]. Decentralizing the computation burden among agents has been widely studied in networks ranging from sensor and robot to social and economic networks [22–25]. In this broad class of problems, agents in a network need to perform a global task for which they only have *partial* information. Therefore, they recursively exchange information with their neighbors, and the global dispersion of information in the network provides them with adequate data to accomplish the task. In the big picture, many of these schemes can also be embedded in the context of *consensus* protocols which have gained a growing popularity over the past three decades [26–28].

In this chapter, we develop a distributed detection algorithm using the model of learning and detection proposed by Jadbabaie et al. [29]. In this framework, the world is governed by a fixed true *state* or *hypothesis* that is aimed to be recovered by a network of agents. The state belongs to a *finite* set, and might represent a decision, an opinion, the price of a product or any quantity of interest.

Each agent observes a stream of *private* signals generated by a marginal of the global likelihood *conditioned* on the true state. However, the signals might *not* be informative enough for the agent to distinguish the underlying state of the world. Therefore, agents use *local* diffusion to compensate for their imperfect knowledge about the environment.

In the literature, a host of schemes build on the model in [29] to describe distributed learning. Despite the wealth of results on the asymptotic behavior of these methods, the *finite-time* analysis remains elusive. Though appealing in certain cases, asymptotic analysis only describes the dominant factors that influence learning in the long run. In real world applications, however, the *decision* on the true state has to be made in a *finite* time. Therefore, it is crucial to study the finite-time variant of these schemes to gain insight into the interplay of *network parameters* which affect learning. For instance, let us think of a social network where individuals need to choose a product which best suits the network. Individuals might value the product differently, and they need to reach consensus in a few rounds of opinion exchange. Agents do not have an infinite horizon to make a decision; therefore, one needs to view this scenario as a finite-time problem. To this end, following up on the work of Duchi et al. [30] on distributed dual averaging, we propose an optimization-based algorithm for distributed detection.

The rest of the chapter is organized as follows: we provide a summary of our results and the related literature to the problem in Section 2.1. We describe the formal statement of the problem, and flesh out the distributed detection scheme in Section 2.2. Section 2.3 is devoted to the finite-time analysis of the algorithm, whereas Section 2.4 elaborates on the impact of network characteristics on the convergence rate. We discuss briefly about asymptotic learning in time-varying network topologies in Section 2.5, and provide our numerical experiments in Section 2.6. The contents of this chapter are mainly from the works of [31–33].

2.1 Contribution and Related Literature

Our main goal is to address the non-asymptotic behavior of the algorithm. To this end, we introduce the notion of *Kullback-Leibler (KL)* cost to measure the learning rate of an individual agent versus an *expert* who has all available information for learning. The KL decentralization cost simply compares the performance of distributed algorithm to its centralized counterpart. We derive an upper bound on the cost which proves the *spectral gap* of the network is substantial beside agents' centralities. It turns out that the upper bound scales inversely in the spectral gap, and logarithmically with the network size and number of states. The rate also scales with the inverse of the relative entropy of the conditional marginals. More specifically, the KL cost grows when signals do not provide enough evidence in favor of the true state versus some other state of the world.

Assuming that the network is realized with a *default* communication structure, each agent has a fixed measure of influence or *centrality*. We establish that allocating more informative signals to more central agents can expedite learning. More interestingly, the importance of spectral gap opens new venues for *optimal* network design to facilitate agents' interactions. Each agent assigns different weights to its neighbors' information while communicating with them. We demonstrate how agents can modify these weights to achieve a faster learning rate. The key idea is to find the Markov chain with the best mixing behavior that is consistent with the network structure and agents' centralities. On the other hand, as a natural conjecture, we expect a more rapid learning rate in well-connected networks. We study the ramification of *link failures* in the network, and prove that in symmetric networks, less connectivity amounts to a sluggish learning process. We further apply our results on star, cycle and two-dimensional grid network, and observe that in each case the effect of spectral gap can be translated to the network *diameter*. Intuitively, a larger diameter makes the information propagation difficult around the network.

We also prove convergence of beliefs in two types of *time-varying* network topologies. First, we restrict out attention to *gossip* protocols as an instance of stochastically switching networks. Next, we develop a switching rule where agents communicate to their neighbors only if their private signals are not informative. The latter is motivated by social network applications in which individuals only communicate when they need to obtain information from human resources. We finally present several examples on binary signal detection which perfectly match our theoretical findings.

2.1.1 Related Literature

Earlier works on decentralized detection have considered scenarios where each agent sends its observations to a fusion center that decides over the true value of a parameter [4, 5, 22]. In these situations, the fusion center faces a classical hypothesis testing (centralized detection) problem after collecting the data from agents. Distributed detection has been widely regarded in various works providing the asymptotic analysis. Cattivelli et al. [34] propose a fully distributed algorithm where no fusion center is necessary. The methodology builds on the connection of Neyman-Pearson detection and minimum-variance estimation to solve the problem. Jakovetić et al. [35] develop a consensus+innovations algorithm for detection under *Gaussian* observations. The method achieves an *asymptotic* exponential error rate even when communications of agents are noisy. In [36], the authors extend the consensus+innovations method to *generic* (non-Gaussian) observations over random networks.

We now elaborate on several works inspired by the learning model in [29]. The authors in [29] propose a non-Bayesian update rule in the context of social networks. Each individual averages her Bayesian posterior belief with the opinion of her neighbors. It is then shown that, under mild technical assumptions, agents' beliefs converge to the true state almost surely. Lalitha et al. [37] introduce another strategy where agents perform a local Bayesian update, and geometrically average the posteriors in their neighborhood. The authors then provide the convergence and rate analysis of their method. In [38, 39], a learning without recall approach is considered where each agent performs Bayesian update sequentially using the prior of one particular neighbor. Nedić et al. [40–42] address the finite-time analysis of a similar problem in deterministically switching networks. In their setting, the prior is geometrically averaged among neighbors of each agent. On the other hand, Rahnama Rad et al. [43] present a distributed estimation algorithm for *continuous* state space. They prove the convergence of the algorithm, and characterize the asymptotic efficiency of the method in compare to any centralized estimator. In [29, 37], the convergence occurs exponentially fast, and the *asymptotic* rate is characterized in terms of the *relative entropy* of individuals' signal structures and their *eigenvector centralities* (see [44] for the rate analysis of [29]).

2.2 The Problem Description and Algorithm

In this section, we describe the observation and network model, and outline the centralized setting for the problem. Then, we provide a formal statement of the distributed setting, and characterize the decentralization cost.

2.2.1 Notation

We adhere to the following notation in the exposition of our results:

$[n]$	The set $\{1, 2, ..., n\}$ for any integer n
$x(k)$	The k-th element of vector x
$x_{[k]}$	The k-th largest element of vector x
I_m	Identity matrix of size m
Δ_m	The m-dimensional probability simplex
$\mathbf{e}_k$	Delta distribution on k-th component
$\|\cdot\|_p$	p-norm operator
$\mathbb{1}$	Vector of all ones
$\|\mu - \pi\|_{\text{TV}}$	Total variation distance between $\mu, \pi \in \Delta_m$
$D_{KL}(\mu\|\pi)$	KL-divergence of $\pi \in \Delta_m$ from $\mu \in \Delta_m$
$\lambda_i(W)$	The i-th largest eigenvalue of matrix W

Table 2.1: Notation

For any $f \in \mathbb{R}^m$ and $\mu \in \Delta_m$, we let $\mathbb{E}_\mu[\cdot]$ represent the expectation of f under the measure μ, i.e., we have $\mathbb{E}_\mu[f] = \sum_{j=1}^{m} \mu(j)f(j)$. Throughout, all the vectors are assumed to be column vectors.

2.2.2 Observation Model

The signal and observation model of this work closely follows the framework proposed in [29]. We consider an environment in which $\Theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$ denotes a finite set of *states* of the world. We have a network of n agents that seek the *unique*, true state of the world $\theta_1 \in \Theta$. At each time $t \in [T]$, the belief of agent i is denoted by $\mu_{i,t} \in \Delta_m$, where Δ_m is a probability distribution over the set Θ. In particular, $\mu_{i,0} \in \Delta_m$ denotes the prior belief of agent $i \in [n]$ about the states of the world, and it is assumed to be uniform[1].

The learning model is given by a conditional likelihood function $\ell(\cdot|\theta_k)$ which is governed by a state of the world $\theta_k \in \Theta$. For each $i \in [n]$, let $\ell_i(\cdot|\theta_k)$ denote the i-th marginal of $\ell(\cdot|\theta_k)$, and we use the vector representation $\ell_i(\cdot|\theta) = [\ell_i(\cdot|\theta_1), ..., \ell_i(\cdot|\theta_m)]^\top$ to stack all states. At each time $t \in [T]$, the signal $s_t = (s_{1,t}, s_{2,t}, \ldots, s_{n,t}) \in \mathcal{S}_1 \times \cdots \times \mathcal{S}_n$ is generated based on the true state θ_1. Therefore, for each $i \in [n]$, the signal $s_{i,t} \in \mathcal{S}_i$ is a sample drawn according to the likelihood $\ell(\cdot|\theta_1)$ where $\mathcal{S}_i$ is the sample space.

The signals are i.i.d. over time, and also the marginals are independent, i.e., $\ell(\cdot|\theta_k) = \Pi_{i=1}^n \ell_i(\cdot|\theta_k)$ for any $k \in [m]$. For the sake of convenience, we define $\psi_{i,t} := \log \ell_i(s_{i,t}|\theta)$ which is a sample corresponding to $\Psi_i := \log \ell_i(\cdot|\theta)$ for any $i \in [n]$.

A1. We assume that all log-marginals are uniformly *bounded* such that $\|\psi_{i,t}\|_\infty \leq B$ for any $s_{i,t} \in \mathcal{S}_i$, i.e., we have $|\log \ell_i(\cdot|\theta_k)| \leq B$ for any $i \in [n]$ and $k \in [m]$.

Based on assumption **A1**, every private signal has bounded information content. The assumption can also be interpreted as Radon-Nikodym derivative of every private signal (likelihood ratio) being bounded [45]. This bound can be found, for instance, when the signal space is discrete and provides a full support for distribution. Let us define $\bar{\Theta}_i$ as the set of states that are observationally equivalent to θ_1 for agent $i \in [n]$; in other words, $\bar{\Theta}_i = \{\theta_k \in \Theta : \ell_i(s_i|\theta_k) = \ell_i(s_i|\theta_1) \;\; \forall s_i \in \mathcal{S}_i\}$ with probability one. As evident from the definition, any state $\theta_k \neq \theta_1$ in the set $\bar{\Theta}_i$ is not distinguishable from the true state by observation of samples from the i-th marginal. Let $\bar{\Theta} = \cap_{i=1}^n \bar{\Theta}_i$ be the set of states that are observationally equivalent to θ_1 from all agents perspective.

A2. We assume that no state in the world is observationally equivalent to the true state from the

[1] The assumption of uniform prior only lets us avoid notational clutter. The analysis holds for any prior with full support.

standpoint of the network, i.e., the true state is *globally identifiable*, and we have $\bar{\Theta} = \{\theta_1\}$.

Assumption **A2** guarantees that the global likelihood provides sufficient information to make the true state uniquely identifiable. In other words, for any false state $\theta_k \neq \theta_1$, there must exist an agent who is able to distinguish θ_1 from θ_k.

Let $\mathcal{F}_t$ be the smallest σ-field containing the information about all agents up to time t. Then, when the learning process continues for T rounds, the probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ is defined as follows: the sample space $\Omega = \otimes_{t=1}^{T}(\otimes_{i=1}^{n}\mathcal{S}_i)$, the σ-field $\mathcal{F} = \cup_{t=1}^{T}\mathcal{F}_t$, and the true probability measure $\mathbb{P} = \otimes_{t=1}^{T}\ell(\cdot|\theta_1)$. Finally, the operator $\mathbb{E}$ denotes the expectation with respect to $\mathbb{P}$.

2.2.3 Network Model

The interaction between agents is captured by a directed graph $G = ([n], E)$, where $[n]$ is the set of nodes corresponding to agents, and E is the set of edges. Agent i receives information from j only if the pair $(i, j) \in E$. We let $\mathcal{N}_i = \{j \in [n] : (i, j) \in E\}$ be the set of neighbors of agent i. Throughout the learning process agents truthfully report their information to their neighbors. We represent by $[W]_{ii} > 0$ the *self-reliance* of agent i, and by $[W]_{ij} > 0$ the weight that agent i assigns to information received from agent j in its neighborhood. Then, the matrix W is constructed such that $[W]_{ij}$ denotes the entry in its i-th row and j-th column. Therefore, W has nonnegative entries, and $[W]_{ij} > 0$ only if $(i, j) \in E$. For normalization purposes, we further assume that W is stochastic; hence,

$$\sum_{j=1}^{n}[W]_{ij} = \sum_{j \in \mathcal{N}_i}[W]_{ij} = 1.$$

A3. We assume that the network is *strongly connected*, i.e., there exists a directed path from any agent $i \in [n]$ to any agent $j \in [n]$. We further assume for simplicity that W is diagonalizable[2].

[2] Note that diagonalizability is not necessary for convergence analysis, and it only simplifies the results by avoiding

The strong connectivity constraint in assumption **A3** guarantees the information flow in the network. The assumption implies that $\lambda_1(W) = 1$ is unique, and the other eigenvalues of W are strictly less than one in magnitude [46]. Given the matrix of social interactions W, the eigenvector centrality is a non-negative vector π such that for all $i \in [n]$,

$$\pi(i) = \sum_{j=1}^{n} [W]_{ji}\pi(j). \tag{2.2.1}$$

for $\|\pi\|_1 = 1$. Then, $\pi(i)$ denoting the i-th element of π is the eigenvector centrality of agent i. In the matrix form, the preceding relation takes the form $\pi^\top W = \pi^\top$, which means π is the stationary distribution of W. Assumption **A3** entails that the Markov chain W is irreducible and aperiodic, and the unique stationary distribution π has strictly positive components [46].

2.2.4 Centralized Detection

To motivate the development of distributed scheme, we commence by introducing centralized detection[3]. In this case, the scenario could be described as a two player repeated game between Nature and a *centralized* agent (expert) that has *global* information to learn the true state. More specifically, the expert observes the sequence of signals $\{s_t\}_{t=1}^T$ that are in turn revealed by Nature, and knows the entire network characteristics. At any round $t \in [T]$, the expert accumulates a *weighted average* of log-marginals, and forms the *belief* $\mu_t \in \Delta_m$ about the states, where $\Delta_m = \{\mu \in \mathbb{R}^m \mid \mu \succeq 0,\ \sum_{k=1}^m \mu(k) = 1\}$ denotes the m-dimensional probability simplex. Letting

$$\psi_t := \sum_{i=1}^{n} \pi(i)\psi_{i,t} = \sum_{i=1}^{n} \pi(i) \log \ell_i(s_{i,t}|\theta), \tag{2.2.2}$$

Jordan blocks. In the absence of this assumption, our theoretical results will depend on the size of the largest Jordan block of W, which only complicates the message of the problem.

[3] The method can be cast as special cases of *Follow the Regularized Leader* [47] and *Mirror Descent* [48] algorithm.

the sequence of interactions could be depicted in the form of the following algorithm:

Centralized Detection

Input : A uniform prior belief μ_0, a learning rate $\eta > 0$.

Initialize : Let $\phi_0(k) = 0$ for all $k \in [m]$.

At time $t = 1, ..., T$: Observe the signal $s_t = (s_{1,t}, s_{2,t}, \dots, s_{n,t})$, update the vector function ϕ_t, and form the belief μ_t as follows,

$$\phi_t = \phi_{t-1} + \psi_t \quad , \quad \mu_t = \text{argmin}_{\mu \in \Delta_m} \left\{ -\mu^\top \phi_t + \frac{1}{\eta} D_{KL}(\mu \| \mu_0) \right\}. \tag{2.2.3}$$

Weighting the marginals based on the eigenvector centrality (2.2.2), the centralized detector aggregates a geometric average of marginals in ϕ_t. At each time $t \in [T]$, the goal is to maximize the expected sum while sticking to the default belief μ_0, i.e., minimizing the divergence. The trade-off between the two behavior is tuned with the *learning rate* η.

Let us note that according to Jensen's inequality for the concave function $\log(\cdot)$, we have for every $i \in [n]$ and $k \in [m]$ that

$$-D_{KL}\left(\ell_i(\cdot|\theta_1) \| \ell_i(\cdot|\theta_k)\right) = \mathbb{E}\left[\log \frac{\ell_i(\cdot|\theta_k)}{\ell_i(\cdot|\theta_1)}\right] \leq \log \mathbb{E}\left[\frac{\ell_i(\cdot|\theta_k)}{\ell_i(\cdot|\theta_1)}\right] = 0,$$

where the inequality turns to equality if and only if $\ell_i(\cdot|\theta_1) = \ell_i(\cdot|\theta_k)$, i.e., iff $\theta_k \in \bar{\Theta}_i$. Therefore, it holds that $\mathbb{E}[\log \ell_i(\cdot|\theta_k)] \leq \mathbb{E}[\log \ell_i(\cdot|\theta_1)]$, and recalling that the stationary distribution π consists of positive elements, we have for any $k \neq 1$ that,

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^{n} \pi(i)\Psi_i(k)\right] &= \mathbb{E}\left[\sum_{i=1}^{n} \pi(i) \log \ell_i(\cdot|\theta_k)\right] \\ &< \mathbb{E}\left[\sum_{i=1}^{n} \pi(i) \log \ell_i(\cdot|\theta_1)\right] = \mathbb{E}\left[\sum_{i=1}^{n} \pi(i)\Psi_i(1)\right], \end{aligned}$$

where the strict inequality is due to uniqueness of the true state θ_1, and the fact that $\bar{\Theta} = \cap_{i=1}^{n} \bar{\Theta}_i = \{\theta_1\}$ based on assumption **A2**. In the sequel, without loss of generality, we assume the following descending order, i.e.

$$\mathbb{E}\left[\sum_{i=1}^{n}\pi(i)\Psi_i(1)\right] > \mathbb{E}\left[\sum_{i=1}^{n}\pi(i)\Psi_i(2)\right] \geq \cdots \geq \mathbb{E}\left[\sum_{i=1}^{n}\pi(i)\Psi_i(m)\right]. \tag{2.2.4}$$

We shall see that the ordering will only simplify the derivation of technical results throughout the chapter.

2.2.5 Distributed Detection

We now extend the previous section to distributed setting modeled based on a network of agents. In the distributed scheme, each agent $i \in [n]$ only observes the stream of private signals $\{s_{i,t}\}_{t=1}^{T}$ generated based on the parametrized likelihood $\ell_i(\cdot|\theta_1)$. That is, agent $i \in [n]$ does not directly observe $s_{j,t}$ for any $j \neq i$. As a result, it gathers the local information by averaging the log-likelihoods in its neighborhood, and forms the belief $\mu_{i,t} \in \Delta_m$ at round $t \in [T]$ as follows:

Distributed Detection

Input : A uniform prior belief $\mu_{i,0}$, a learning rate $\eta > 0$.

Initialize : Let $\phi_{i,0}(k) = 0$ for all $k \in [m]$ and $i \in [n]$.

At time $t \in [T]$: Observe the signal $s_{i,t}$, update the function $\phi_{i,t}$, and form the belief $\mu_{i,t}$ as follows,

$$\phi_{i,t} = \sum_{j\in\mathcal{N}_i}[W]_{ij}\phi_{j,t-1} + \psi_{i,t} \;\;,\;\; \mu_{i,t} = \text{argmin}_{\mu\in\Delta_m}\left\{-\mu^{\top}\phi_{i,t} + \frac{1}{\eta}D_{KL}(\mu\|\mu_{i,0})\right\}. \tag{2.2.5}$$

As outlined above, each agent updates its belief using purely local diffusion. We are interested in measuring the efficiency of the distributed algorithm via a metric comparing that to its centralized counterpart. At any round $t \in [T]$, let us postulate that the cost which agent $i \in [n]$ needs to pay to have the same opinion as the expert is $D_{KL}(\mu_{i,t}\|\mu_t)$; then, the total *decentralization cost* that the agent incurs after T rounds is as follows

$$\mathbf{Cost}_{i,T} := \sum_{t=1}^{T} D_{KL}(\mu_{i,t} \| \mu_t) = \sum_{t=1}^{T} \mathbb{E}_{\mu_{i,t}} \left[\log \frac{\mu_{i,t}}{\mu_t} \right]. \tag{2.2.6}$$

At each round, the output of the centralized and decentralized algorithm is a probability distribution over state space. The KL-divergence captures the dissimilarity of two probability distributions; hence, it could be a reasonable metric to measure the difference between two algorithms. The function quantifies the difference between the agent that observes private signals $\{s_{i,t}\}_{t=1}^{T}$ and an expert that has $\{s_t\}_{t=1}^{T}$ and π available. In other words, it shows how well the decentralized algorithm copes with the partial information. Note importantly that $\mathbf{Cost}_{i,T}$ is a random quantity since the expectation is *not* taken with respect to randomness of signals.

We conclude this section with the following lemma which reiterates that both algorithms are reminiscent of the well-known *Exponential Weights* algorithm.

Lemma 2.1. *The update rules* (2.2.3) *and* (2.2.5) *have the explicit form solutions,*

$$\mu_t(k) = \frac{\exp\{\eta\phi_t(k)\}}{\langle \mathbb{1}, \exp\{\eta\phi_t\}\rangle} \quad \textit{and} \quad \mu_{i,t}(k) = \frac{\exp\{\eta\phi_{i,t}(k)\}}{\langle \mathbb{1}, \exp\{\eta\phi_{i,t}\}\rangle},$$

respectively, for any $i \in [n]$ and $k \in [m]$. Moreover,

$$\phi_{i,t} = \sum_{\tau=1}^{t} \sum_{j=1}^{n} \left[W^{t-\tau}\right]_{ij} \psi_{j,\tau}.$$

One can observe from above that

$$\sum_{i=1}^{n} \pi(i)\phi_{i,t} = \sum_{\tau=1}^{t} \sum_{j=1}^{n} \sum_{i=1}^{n} \pi(i) \left[W^{t-\tau}\right]_{ij} \psi_{j,\tau} = \sum_{\tau=1}^{t} \sum_{j=1}^{n} \pi(j)\psi_{j,\tau} = \phi_t,$$

which connects the centralized and decentralized update via eigenvector centrality (2.2.1). As explored in [37, 44], we shall see that centrality plays an important role in the convergence rate.

2.3 Finite-time Analysis of Cost Functions

In this section, we investigate the convergence of agents' beliefs to the true state in the network. Agents exchange information over time, and reach consensus about the true state. The connectivity of the network plays an important role in the learning as $W^t \to \mathbb{1}\pi^\top$ as $t \to \infty$. To examine the learning rate, we need to have knowledge about the mixture behavior of Markov chain W. The following lemma sheds light on the mixture rate, and we invoke it later for technical analysis.

Lemma 2.2. *Let the strong connectivity of network (assumption* **A3***) hold, and define* $\lambda_{\max}(W) := \max\{|\lambda_n(W)|, |\lambda_2(W)|\}$. *Then, for any* $t \in [T]$ *and* $n > 5$*, the stochastic matrix* W *satisfies*

$$\sum_{\tau=1}^{t}\sum_{j=1}^{n}\left|\left[W^{t-\tau}\right]_{ij} - \pi(j)\right| \leq \frac{4\log n}{1-\lambda_{\max}(W)},$$

for any $i \in [n]$ *where* $0 \leq \lambda_{\max}(W) < 1$.

We now establish that agents have arbitrarily close opinions in a connected network. Furthermore, the convergence rate is governed by cardinality of state space and network characteristics.

Lemma 2.3. *Let the sequence of beliefs* $\{\mu_{i,t}\}_{t=1}^T$ *for each agent* $i \in [n]$ *be generated by the Distributed Detection algorithm with the learning rate* η*. Given bounded log-marginals (assumption* **A1***), global identifiability of the true state (assumption* **A2***), and strong connectivity of the network (assumption* **A3***), for each individual agent* $i \in [n]$ *it holds that*

$$\frac{1}{\eta}\log\|\mu_{i,t} - \mathbf{e}_1\|_{TV} \leq -\mathcal{I}(\theta_1,\theta_2)t + \sqrt{2B^2 t\log\frac{m}{\delta}} + \frac{8B\log n}{1-\lambda_{\max}(W)} + \frac{\log m}{\eta},$$

with probability at least $1-\delta$*, where for* $k \geq 2$

$$\mathcal{I}(\theta_1,\theta_k) := \sum_{i=1}^{n}\pi(i)D_{KL}(\ell_i(\cdot|\theta_1)\|\ell_i(\cdot|\theta_k)).$$

In particular, we have $\|\mu_{i,t} - \mathbf{e}_1\|_{TV} \longrightarrow 0$ *almost surely.*

Beside providing an any-time bound in the high probability sense, the lemma verifies that the belief $\mu_{i,t}$ of each agent $i \in [n]$ is *strongly* consistent, i.e., it converges almost surely to a delta distribution on the true state. We also remark that the asymptotic rate of $\mathcal{I}(\theta_1, \theta_2)$ was also discovered in [31, 37, 44] for the updates under study. However, Lemma 2.3 provides a non-asymptotic version of the convergence rate. Let us proceed to the next lemma to derive a total variation bound on the decentralization cost (2.2.6).

Lemma 2.4. *The instantaneous KL cost associated to the Distributed Detection algorithm with the learning rate η satisfies for any $t \in [T]$*

$$D_{KL}(\mu_{i,t} \| \mu_t) \le 2\|\mathbf{e}_1 - \mu_t\|_{TV},$$

as long as $\eta\|q_{i,t}\|_\infty \le 1/4$ at each round, where $q_{i,t} := \phi_{i,t} - \phi_t$.

The bound in Lemma 2.4 is evocative of a reverse Pinsker's inequality. It provides a total variation bound on the cost function which is of the KL-divergence form. Let us remark that an appropriate choice of learning rate η warrants the condition $\eta\|q_{i,t}\|_\infty \le 1/4$. We now present the main result of the chapter in the following theorem.

Theorem 2.1. *Let the sequence of beliefs $\{\mu_{i,t}\}_{t=1}^T$ for each agent $i \in [n]$ be generated by the Distributed Detection algorithm with the choice of learning rate $\eta = \frac{1-\lambda_{\max}(W)}{16B\log n}$. Given bounded log-marginals (assumption* **A1***), global identifiability of the true state (assumption* **A2***), and strong connectivity of the network (assumption* **A3***), we have*

$$\mathbf{Cost}_{i,T} \le \frac{18B^2}{\mathcal{I}^2(\theta_1,\theta_2)} \max\left\{\log\left[\frac{6m}{\delta}\right], \frac{3B\sqrt{2}}{\mathcal{I}(\theta_1,\theta_2)}\right\} + \frac{48B\log n}{\mathcal{I}(\theta_1,\theta_2)}\frac{\log m + 2}{1-\lambda_{\max}(W)},$$

with probability at least $1-\delta$.

Regarding Theorem 2.1 the following comments are in order: the rate is related to the inverse of $\mathcal{I}(\theta_1, \theta_2)$ which is a weighted average of KL-divergence of observations under θ_2 (the second best alternative) from observations under θ_1 (the true state). Also, from the definition of $\mathcal{I}(\theta_1, \theta_2)$ in Lemma 2.3, the weights turn out to be agents' centralities. Intuitively, when signals hardly reveal the difference between the best two candidates for the true state, agents must make more effort to distinguish the two. In turn, this results in suffering a larger cost caused by slower learning. The decentralization cost always scales logarithmically with the number of states m. Now define

$$\gamma(W) := 1 - \lambda_{\max}(W), \tag{2.3.1}$$

as the *spectral gap* of the network. Then, Theorem 2.1 suggests that for large networks, the cost scales inversely in the spectral gap, and logarithmically with the network size n. Finally, the detection cost is time-independent and optimal with respect to time horizon (with high probability). Therefore, the average expected cost (per iteration cost) asymptotically tends to zero.

2.4 The Impact of Network Topology

The results of previous section verify that network characteristics govern the learning process. We now discuss the role of agents' centralities and the network spectral gap.

2.4.1 Effect of Agent Centrality

To examine centrality, let us return to the definition of $\mathcal{I}(\theta_1, \theta_2)$ in Lemma 2.3, and imagine that the network is *collaborative* in the sense that the network designer wants to expedite learning. Then, to have the best information dispersion, the marginal which collects the most evidence in favor of θ_1 against θ_2 should be allocated to the most central agent. By the same token, in an *adversarial* network where Nature aims to delay the learning process, such marginal should be assigned to the least central agent. To sum up, let us put forth the concept of network *regularity* as defined in [44] in the context of social learning. Recalling the definition of eigenvector centrality (2.2.1), we say a network G is more regular than G' if π' majorizes π, i.e., if for all $j \in [n]$

$$\sum_{i=1}^{j} \pi_{[i]} \leq \sum_{i=1}^{j} \pi'_{[i]}, \tag{2.4.1}$$

where $\pi_{[i]}$ denotes the i-th largest element of π. Letting

$$u := \left[D_{KL}(\ell_1(\cdot|\theta_1)\|\ell_1(\cdot|\theta_2)), \ldots, D_{KL}(\ell_n(\cdot|\theta_1)\|\ell_n(\cdot|\theta_2))\right]^\top,$$

it is a straightforward consequence of Lemma 1 proved in [44] that

$$\sum_{i=1}^{n} \pi_{[i]} u_{[i]} \leq \sum_{i=1}^{n} \pi'_{[i]} u_{[i]},$$

when π' majorizes π. Therefore, spreading more informative signals among central agents speeds up the learning procedure.

2.4.2 Optimizing the Spectral Gap

We now turn our attention to the spectral gap of network (2.3.1). Suppose that agents are given a default communication matrix W which determines their neighborhood and centrality. The problem is to find the optimal spectral gap assuming that the neighborhood and centrality of each agent are fixed. The key idea is to change the mixing behavior of the Markov chain W. It is well-known, for instance, that we could do so using lazy random walks [49] which replaces W with $\frac{1}{2}(W + I_n)$. To generalize the idea, let us define a modified communication matrix

$$W' := \alpha W + (1-\alpha) I_n \quad \alpha \in [0,1], \tag{2.4.2}$$

which has the same eigenstructure as W. Then, the eigenvalues of W' are weighted averages of those of W with one. From standpoint of network design, one can exploit the freedom in choosing α to optimize the spectral gap.

Proposition 2.1. *The optimal spectral gap of the modified communication matrix W'* (2.4.2) *is as follows,*

$$\gamma^* = \frac{2-2\lambda_2(W)}{2-\lambda_n(W)-\lambda_2(W)} \quad \textit{for} \quad \alpha^* = \frac{2}{2-\lambda_n(W)-\lambda_2(W)},$$

when $\lambda_n(W) + \lambda_2(W) < 0$

Proof. To optimize the spectral gap, we need to minimize the second largest eigenvalue of W' in magnitude, that is, to solve the min-max problem

$$\min_{\alpha\in[0,1]} \lambda_{\max}(W') = \min_{\alpha\in[0,1]} \max\left\{|\alpha\lambda_2(W)+1-\alpha|, |\alpha\lambda_n(W)+1-\alpha|\right\}. \tag{2.4.3}$$

The functions $|\alpha\lambda_2(W)+1-\alpha|$ and $|\alpha\lambda_n(W)+1-\alpha|$ are both convex with respect to α. Therefore, the point-wise maximum of the two is also convex, and achieves its minimum on a compact set. Since $\lambda_n(W) < -\lambda_2(W)$ by hypothesis, the minimum occurs at the intersection of the following lines

$$\alpha\lambda_2(W) + 1 - \alpha = -\alpha\lambda_n(W) + \alpha - 1,$$

yielding $\alpha^* = \frac{2}{2-\lambda_n(W)-\lambda_2(W)}$. Plugging α^* into the min-max problem (2.4.3), we calculate the optimal value $\lambda^*_{\max}$ as

$$\lambda^*_{\max} = \frac{\lambda_2(W)-\lambda_n(W)}{2-\lambda_n(W)-\lambda_2(W)},$$

and since $\gamma^* = 1-\lambda^*_{\max}$ the proof follows immediately. □

We remark that when the Markov chain is symmetric, the problem can be formulated as a convex optimization [50]. Moreover, for gossip protocols where the expected communication matrix is symmetric, the problem can be posed as a semidefinite program [51]. However, in our setting the chain is not necessarily symmetric and these results are not applicable.

2.4.3 Sensitivity to Link Failure

It is intuitive that in a network with more links, agents are offered more opportunities for communication. Adding links provides more avenues for spreading information, and improves the learning quality. We study this phenomenon for symmetric networks where a pair of agents assign similar weights to each other, i.e., $W^\top = W$. In particular, we explore the connection of spectral gap with the link failure. In this regard, let us introduce the following positive semi-definite matrix

$$\Delta W(i,j) := (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^\top, \tag{2.4.4}$$

where $\mathbf{e}_i$ is the i-th unit vector in the standard basis of $\mathbb{R}^n$. Then, for $i, j \in [n]$ the matrix

$$\bar{W}(i,j) := W + [W]_{ij}\Delta W(i,j), \tag{2.4.5}$$

corresponds to a new communication matrix that removes edges (i,j) and (j,i) from the network, and adds $[W]_{ij} = [W]_{ji}$ to the self-reliance of agent i and agent j.

Proposition 2.2. *Consider the communication matrix $\bar{W}(i,j)$ in* (2.4.5). *Then, for any $i, j \in [n]$ the following inequality holds*

$$\lambda_{\max}(W) \le \lambda_{\max}\left(\bar{W}(i,j)\right),$$

so long as W is positive semi-definite.

Proof. We recall that $\Delta W(i,j)$ in (2.4.4) is positive semi-definite with $\lambda_n\left(\Delta W(i,j)\right)=0$. Applying Weyl's eigenvalue inequality on (2.4.5), we obtain for any $k \in [n]$

$$\lambda_k\left(W\right) \le \lambda_k\left(\bar{W}(i,j)\right),$$

which holds in particular for $k=2$. On the other hand, the matrix W is positive semi-definite, so we have that $\lambda_{\max}\left(W\right)=\lambda_2\left(W\right)$. Combining with the fact that $\bar{W}(i,j)$ is symmetric and positive semi-definite, the proof is completed. □

The proposition immediately implies that removing a link reduces the spectral gap. In this case, in view of the bound in Theorem 2.1, the decentralization cost has more latitude to vary. Therefore, to keep the costs small, agents tend to maintain their connections. Let us take note of the delicate point that monotone increase in the upper bound does *not* necessarily imply a monotone increase in the cost; however, one can *roughly* expect such behavior. We elaborate on this issue in the numerical experiments. Notice that the positive semi-definiteness constraint on W is not strong, since it can be easily satisfied by replacing a lazy random walk $\frac{1}{2}(W+I_n)$ with W. Finally, we remark that link failures in distributed optimization [52] and consensus protocols [53] has been previously studied in the literature. We refer the interested reader to these references where the impact of *random* link failure is considered.

2.4.4 Star, Cycle and Grid Networks

We now examine the spectral gap impact for some interesting networks (Fig. 2.1), and derive explicit bounds for decentralization cost. In the star network (regardless of the network size), existence of a central agent always preserves the network diameter, and therefore, we expect a benign scaling with network size. On the other side of the spectrum lies the cycle network where the diameter grows linearly with the network size. We should, hence, observe how the poor communication in cycle network affects the learning rate. Finally, as a possible model for sensor networks, we study the grid network where the network size scales quadratically with the diameter.

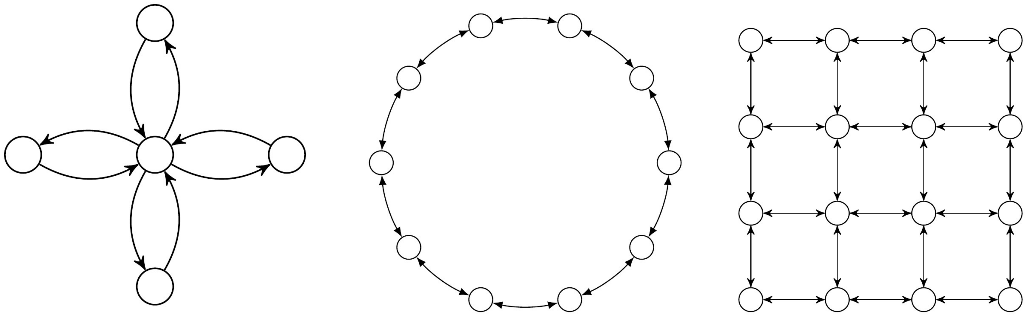

Figure 2.1: Illustration of networks : star, cycle and grid networks with n agents. For each network, each individual agent possesses a self-reliance of $\omega \in (0, 1)$.

Corollary 2.1. *Under conditions of Theorem 2.1 and the choice of learning rate* $\eta = \frac{\gamma(\cdot)}{16B \log n}$, *for* n *large enough we have the following bounds on the decentralization cost:*

(a) *For the star network in Fig. 2.1*

$$\mathbf{Cost}_{i,T} \leq \mathcal{O}\left(\frac{\log [nm]}{\min \{1-\omega, 1-|2\omega-1|\}}\right).$$

(b) *For the cycle network in Fig. 2.1*

$$\mathbf{Cost}_{i,T} \leq \mathcal{O}\left(\frac{\log [nm]}{\min \left\{1-|2\omega-1|, 2(1-\omega)\sin^2 \frac{\pi}{n}\right\}}\right).$$

(c) *For the grid network in Fig. 2.1*

$$\mathbf{Cost}_{i,T} \leq \mathcal{O}\left(\frac{\log [nm]}{\min \left\{1-|2\omega-1|, 2(1-\omega)\sin^2 \frac{\pi}{\sqrt{n}}\right\}}\right).$$

Proof. The spectrum of the Laplacian of star and cycle graphs are well-known [54]. We have the eigenvalue set corresponding to communication matrix of star and cycle graphs as

$$\Big\{1, \omega, \dots, \omega, 2\omega - 1\Big\} \quad \text{and} \quad \Big\{\omega + (1-\omega)\cos\frac{2\pi i}{n}\Big\}_{i=0}^{n-1},$$

respectively. Therefore, the proof of **(a)** and **(b)** follows immediately. The grid graph is the Cartesian product of two rings of size $\sqrt{n}$ (due to wraparounds at the edges), and hence, its eigenvalues are derived by summing the eigenvalues of two $\sqrt{n}$-rings[54]. Therefore, the eigenvalue set takes the form

$$\Big\{\omega + (1-\omega)\cos\frac{\pi(i+j)}{\sqrt{n}}\cos\frac{\pi(i-j)}{\sqrt{n}}\Big\}_{i,j=0}^{\sqrt{n}-1},$$

and the proof of **(c)** is completed. □

Let us use the notation $\tilde{\mathcal{O}}(\cdot)$ to hide the poly log factors. Then, the bounds derived in Corollary 2.1 indicate that the algorithm requires $\tilde{\mathcal{O}}(1)$ iterations to achieve a near optimal log-distance from the true state in the star network. However, the rate deteriorates to $\tilde{\mathcal{O}}(n^2)$(respectively, $\tilde{\mathcal{O}}(n)$) in the cycle (respectively, grid) network. In all cases, the rate depends on the diameter of the network which is a natural indicator of information dissemination quality.

2.5 Switching Topologies : Asymptotic Learning

We addressed the finite-time analysis in the case of fixed network topology. What would happen if the network structure varies over time? In other words, consider the following variant of (2.2.5) with $\eta = 1$,

$$\phi_{i,t} = \sum_{j\in\mathcal{N}_i}[W(t)]_{ij}\phi_{j,t-1} + \psi_{i,t}, \qquad \mu_{i,t}(k) = \frac{\exp\{\phi_{i,t}(k)\}}{\langle \mathbb{1}, \exp\{\phi_{i,t}\}\rangle}, \tag{2.5.1}$$

in which $W(t)$ is a *time-varying* communication matrix. We would like to discuss two interesting switching rules with applications in sensor and social networks. The two protocols are different in nature, though they both guarantee asymptotic learning, i.e., in both scenarios the beliefs converge to the true state asymptotically (in almost sure sense). For the rest of this section, we assume $W^\top = W$.

2.5.1 Stochastic Links

Random link failures are unavoidable in many wireless sensor networks. Sensors might fail to establish connection with each other at some time periods. Therefore, randomized communication protocols are interesting subject of study in many engineering applications. In general, we can discuss convergence of beliefs for any *time-varying* random sequence $\{W(t)\}_{t=1}^{\infty}$ for which

$$W(t)W(t-1)\cdots W(1) \longrightarrow \frac{1}{n}\mathbb{1}\mathbb{1}^\top,$$

almost surely. However, we particularize our discussion to an invariant gossip protocol studied extensively in [51]. In this scenario, each node has a clock which ticks according to a rate 1 Poisson process. Equivalently, there is a single global clock which ticks according to a rate n Poisson process at times $\mathscr{T}_t$, where $\{\mathscr{T}_t - \mathscr{T}_{t-1}\}$ are i.i.d. exponential random variables with rate n. We use the index t to refer to the t-th time slot $[\mathscr{T}_{t-1}, \mathscr{T}_t)$, $t \geq 0$. At each tick $\mathscr{T}_t$ of the global clock, agent $I_t \in [n]$ is picked uniformly at random. Then, it contacts a neighbor $J_t \in [n]$ with probability $[W]_{I_tJ_t}$, and they update their belief according to (2.5.1). Denoting the communication matrix by $W(t)$, this amounts to $W(t)$ taking the form

$$W(t) = I_n - \frac{(\mathbf{e}_{I_t} - \mathbf{e}_{J_t})(\mathbf{e}_{I_t} - \mathbf{e}_{J_t})^\top}{2}, \tag{2.5.2}$$

with probability $\frac{1}{n}[W]_{I_t J_t}$, where $\mathbf{e}_i$ is the i-th unit vector in the standard basis of $\mathbb{R}^n$. Since the network topology is formed randomly at each time, we need to modify assumption **A3** as follows:

A3*. The network is connected in *expectation* sense, i.e. there exists a path from any agent $i \in [n]$ to any agent $j \neq i$ on graph G, and the second largest eigenvalue of $\mathbb{E}[W(t)]$ is strictly less than one in magnitude.

The assumption, for instance, holds if the underlying structure of the network is connected and nonbipartite. The following theorem shows that agents learn the true state almost surely using the gossip protocol.

Theorem 2.2 (Learning with Gossip Protocol)**.** *Let the bound on log-marginals (assumption* **A1***), global identifiability of the true state (assumption* **A2***), and the connectivity in expectation sense (assumption* **A3****) hold. Then, following the update in* (2.5.1) *using the gossip protocol* (2.5.2)*, all agents learn the truth exponentially fast with an asymptotic rate given by $\mathcal{I}(\theta_1, \theta_2)$, defined in Lemma 2.3.*

The technical analysis is very similar to that of Theorem 2.3 whose proof is provided in Section (2.7).

2.5.2 Information-Based Communication

In many real-world applications, agents do not communicate each and every round. In fact, they only communicate when they need information. An instance of this scenario could be a social network in which individuals aim to decide on a certain product in the market. They do not keep discussing about the best product, whereas they make a decision with a handful of interactions. With no communication (Bayesian update or $W(t) = I$ in (2.5.1)), agents do not distinguish between the true state and its observationally equivalents. On the other hand, a fully non-Bayesian learning ($W(t) = W$ in (2.5.1)) occurs at the cost of all-time communication. Can we stand somewhere between these two extreme cases where agents learn with a low communication cost?

To solve this problem, we propose a switching rule in which agents communicate only when their private signals are not informative. From technical point of view, informativeness is measured with total variation distance between the prior and the posterior of the Bayesian update. That is, given any threshold $\tau > 0$, agent $i \in [n]$ communicates to its neighbors if and only if $\|\mu_{i,t} - \mu_{i,t-1}\|_{TV} < \tau$ given $W(t) = I$. When the condition is satisfied, a bidirectional communication is established, and the matrix $W(t)$ is updated such that $[W(t)]_{ij} = [W(t)]_{ji} = [W]_{ij} = [W]_{ji}$ for all $j \in \mathcal{N}_i$. In summary, the switching protocol works as follows:

Switching Rule

Given $\tau > 0$, for any $i \in [n]$ that satisfies $\|\mu_{i,t} - \mu_{i,t-1}\|_{TV} < \tau$ with $W(t) = I$, the i-th column and row of $W(t)$ take the values of the i-th column and row of the symmetric matrix W. Then, the diagonal elements of $W(t)$ are filled such that the matrix is doubly stochastic.

Before shifting focus to the convergence analysis under the proposed rule, we note that with $\tau = 1$ all signals will be considered uninformative to all agents at every epoch of time; hence, at every time step agents choose to communicate, $W(t) = W$ for all t, and they learn the truth exponentially fast. However, the learning occurs under an all-time communication protocol, which is inefficient when communication is costly. We shall demonstrate that the same learning quality can be achieved through the proposed switching rule, while incurring only a few rounds of communications.

The following lemma concerns the behavior of agents in the Bayesian regime. In particular, it guarantees that with probability one, if the switching condition is satisfied at some time $\mathbf{t}_1$, there exists a $\mathbf{t}_2 > \mathbf{t}_1$ at which the switching condition is satisfied again. Furthermore, the length of interval $\mathbf{t}_2 - \mathbf{t}_1$ is finite almost surely.

Lemma 2.5 (Bayesian Learning)**.** *Let the log-marginals be bounded (assumption* **A1***). Assume that agent $i \in [n]$ is allowed to follow the Bayesian update after some time $\hat{t}$, i.e. $W(t) = I$ in* (2.5.1) *for $t \geq \hat{t}$. We then have*

$$\mu_{i,t}(k) \longrightarrow 0, \quad \forall \theta_k \in \Theta \setminus \bar{\Theta}_i, \tag{2.5.3}$$

almost surely.

Lemma 2.5 simply implies that the Bayesian update does not provide information for agents after a finite (but random) number of iterations. We also state the following proposition (using our notation) from [55] to invoke later in the analysis.

Proposition 2.3. *Consider a sequence of directed graphs $\mathcal{G}_t = ([n], \mathcal{E}_t, A_t)$ for $t \in \mathbb{N}$ where A_t is a stochastic matrix. Assume the existence of real numbers $\delta_{\max} \geq \delta_{\min} > 0$ such that $\delta_{\min} \leq [A_t]_{ij} \leq \delta_{\max}$ for any $(i, j) \in \mathcal{E}_t$. Assume in addition that the graph $\mathcal{G}_t$ is bidirectional for any $t \in \mathbb{N}$. If for all $t_0 \in \mathbb{N}$ there is a node connected to all other nodes across $[t_0, \infty)$, then the left product $A_t A_{t-1} \cdots A_1$ converges to a limit.*

We use the previous technical results to prove that under the proposed switching algorithm, all agents learn the truth, asymptotically and almost surely.

Theorem 2.3 (Learning in Switching Regimes)**.** *Let the bound on log-marginals (assumption* **A1***), global identifiability of the true state (assumption* **A2***), and strong connectivity of the network (assumption* **A3***) hold. Then, following the update in* (2.5.1) *using the switching rule proposed in this section, all agents learn the truth exponentially fast with an asymptotic rate given by $\mathcal{I}(\theta_1, \theta_2)$, defined in Lemma 2.3.*

Theorem 2.3 captures the trade-off between communication and informativeness of private signals. More specifically, private signals do not provide each agent with adequate information to learn the true state. Hence, agents require other signals dispersed throughout the network, which highlights the importance of communication. On the other hand, all-time communication is unnecessary since agents might only need a handful of interactions to augment their imperfect observations with those of their neighbors.

2.6 Example : Binary Signal Detection

In this section, we discuss our numerical experiments. Note that, as mentioned in the footnote of assumption **A3**, in our convergence results the communication matrix need not be diagonalizable, and the assumption is only for convenience. In what follows, we disregard diagonalizability (in the construction of network) for the first section. Therefore, we verify the generality of convergence for arbitrary strongly connected networks.

2.6.1 Convergence of Beliefs

We generate a random network of $n = 50$ agents based on the Erdős-Rényi model. In our example, each link exists with probability 0.3 independent of other links. We verify the strong connectivity of the network before running the experiment. Though generated randomly, the network is fixed throughout the process. Assume that there exist $m = 51$ states in the world and agents are to discover the true state θ_1. At time $t \in [T]$, a signal $s_{i,t} \in \{0, 1\}$ is generated based on the true state such that $\ell_i(\cdot|\theta_1) = \ell_i(\cdot|\theta_{i+1})$. In other words, for agent $i \in [n]$, we have $\bar{\Theta}_i = \{\theta_1, \theta_{i+1}\}$ and θ_{i+1} is observationally equivalent to the true state. Therefore, each agent $i \in [n]$ fails to distinguish θ_1 from θ_{i+1} once relying on the private signals. However, since we have $\bar{\Theta} = \cap_{i=1}^{n} \bar{\Theta}_i = \{\theta_1\}$, the

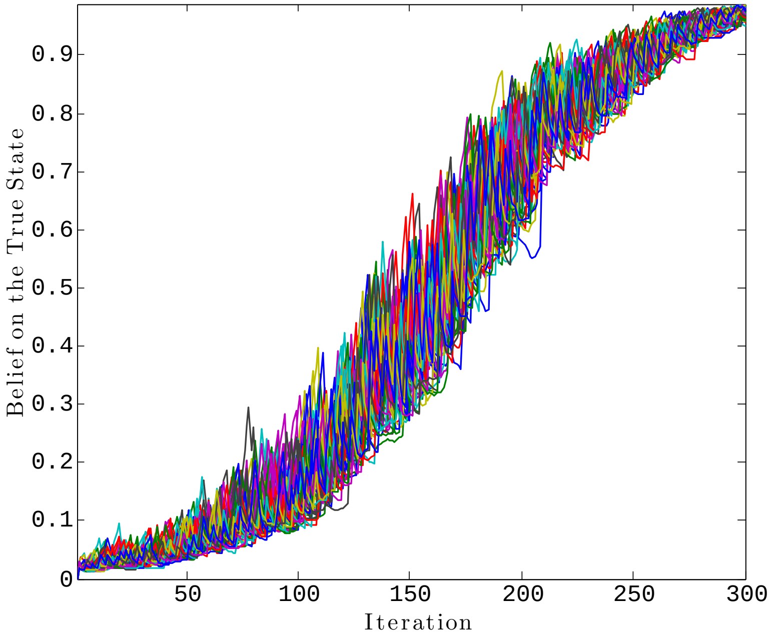

Figure 2.2: The belief evolution for all 50 agents in the network. The global identifiability of the true state and strong connectivity of the network result in learning.

true state is globally identifiable. Consequently, in view of Lemma 2.3, all agents reach a consensus on the true state (Fig. 2.2), and learn the truth exponentially fast.

2.6.2 Optimizing the Spectral Gap

To verify the result of Proposition 2.1, we must construct a communication matrix that is diagonalizable, yet not symmetric. We let

$$W_1 = \begin{bmatrix} 0 & 0.95 & 0.05 \\ 0.95 & 0 & 0.05 \\ 0.05 & 0.95 & 0 \end{bmatrix} \quad \text{and} \quad W_2 = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{bmatrix},$$

and set $W = W_1 \otimes (W_2 \otimes W_2)$. One can verify that W is row stochastic, diagonalizable and asymmetric. Also, $W^t \to \mathbb{1}\pi^\top$ as $t \to \infty$, where π consists of positive elements. The resulting

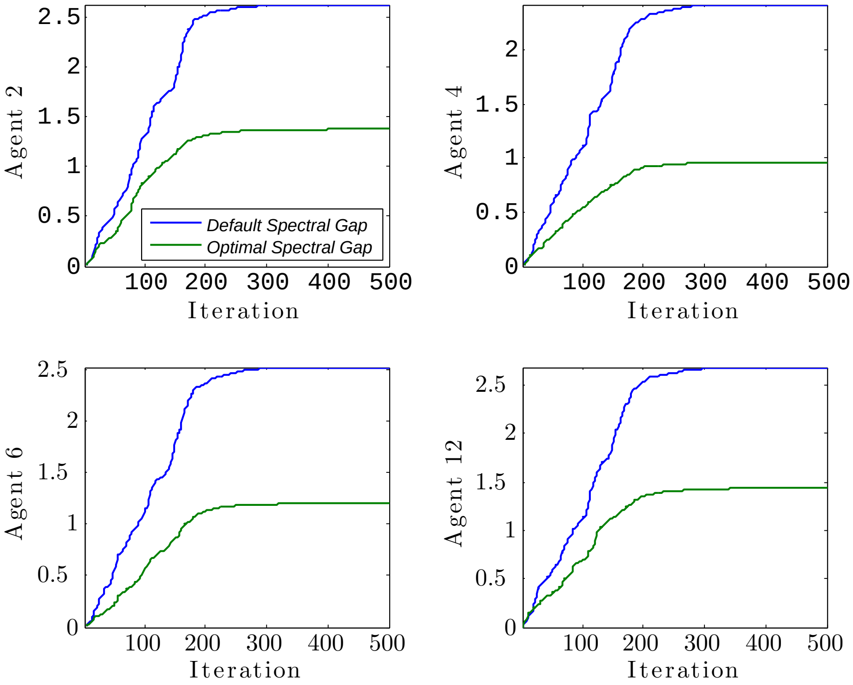

Figure 2.3: The plot of decentralization cost versus time horizon for agents 2, 4, 6 and 12 in the network. The cost in the network with the optimal spectral gap (green) is always less than the network with default weights (blue).

network has a specific structure, but it suits our purposes since it satisfies all the conditions without being symmetric. The signal generating process is precisely the same as the previous section. We now turn to optimizing the spectral gap to speed up learning. We proved in Proposition 2.1 that every default communication matrix can be adjusted to a matrix W' which has the optimal spectral gap when centralities are fixed. Setting the parameter α in (2.4.2) equal to α^* derived in Proposition 2.1, we obtain the optimal network. In this example we have $\gamma(W) = 0.05$, $\alpha^* = 0.7273$ and $\gamma^* = 0.5818$. The dependence of decentralization cost to the spectral gap was theoretically proved in Theorem 2.1. Applying the results of Proposition 2.1 verifies that in the optimal network, agents suffer a lower decentralization cost comparing to the default network (Fig. 2.3). Also, we

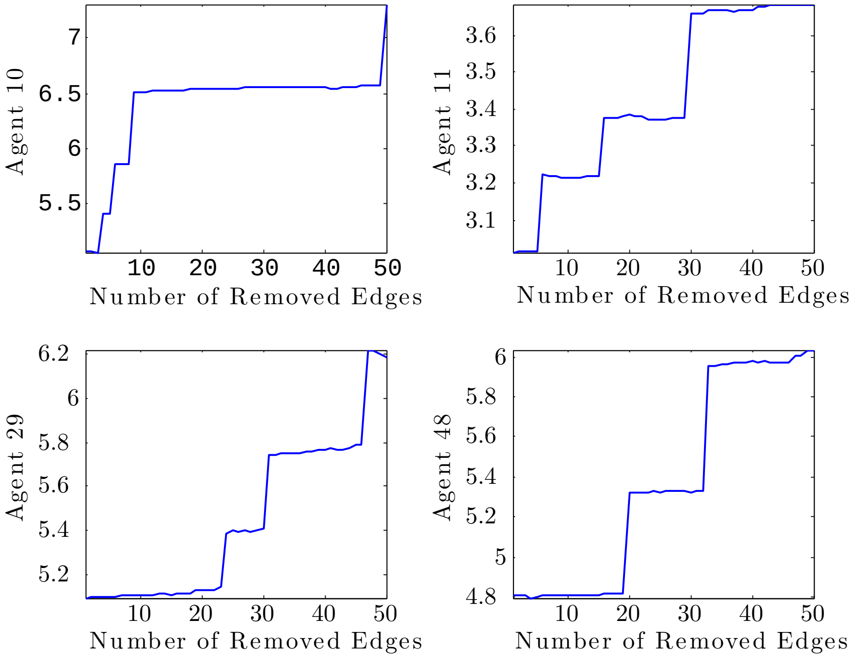

Figure 2.4: The decentralization cost at round $T = 300$ for agents 10, 11, 29 and 48 in the network. Removing the links causes poor communication among agents and increase the decentralization cost.

proved theoretically in Theorem 2.1 that the cost bound is time-independent with high probability. Interestingly, the plot verifies the high probability upper bound on the cost for both cases.

2.6.3 Sensitivity to Link Failure

To evaluate the result of Proposition 2.2, we need a symmetric network. The upper triangle of W is generated using Erdős-Rényi model (similar to the first section), and the matrix is then symmetrized. In this case every agent is equally central, and we have $\pi = \mathbb{1}/n$. To study the impact of link failure, we sequentially select random pairs of agents in the network, and remove their connection. Each time that a link is discarded, we compute the decentralization cost in the new network at iteration $T = 300$, and continue the process until 50 bi-directional edges are eliminated from the network. In view of Proposition 2.2, we expect a monotone decrease in the spectral gap which amounts to a larger decentralization cost. We plot the cost for four agents in the network, and observe that the behavior is almost (not quite) monotonic (Fig. 2.4). The monotone dependence of the upper bound to the spectral gap (Theorem 2.1) does not necessarily guarantee a monotone relationship between cost and the spectral gap. Therefore, we can only roughly expect such behavior.

2.6.4 Efficiency of Information-Based Communication

We now exemplify the efficiency of the switching rule discussed in Section 2.5.2. We set the threshold $\tau > 0$ such that $\log_{10} \tau = -17$, and perform the update (2.5.1) for 1000 iterations. We also run the same update for $\tau = 1$ which corresponds to all-time communication algorithm (2.2.5). Fig. 2.5 represents the belief evolution under both algorithms for a randomly selected agent in the network. We observe that both algorithms converge; however, the switching protocol (our algorithm in this section) outperforms the all-time communication algorithm in terms of efficiency. The selected agent involves in interactions only 41 times in 1000 rounds. Therefore, the communication load simply reduces to 4.1% comparing to the green curve, which proves a significant improvement.

2.7 Omitted Proofs

Proof of Lemma 2.1. The proof is elementary, and it is only given to keep the chapter self-contained. We write the Lagrangian associated to the update (2.2.3) as,

$$L(\mu, \lambda) = -\mu^\top \phi_t + \frac{1}{\eta} \left\langle \mu, \log \frac{\mu}{\mu_0} \right\rangle + \lambda \mu^\top \mathbb{1} - \lambda,$$

where we left the positivity constraint implicit. Differentiating above with respect to μ and λ, and setting the derivatives equal to zero, we get

$$\mu_t(k) = \mu_0(k) \exp\left\{\eta(\phi_t(k) - \lambda) - 1\right\} \quad \text{and} \quad \mu_t^\top \mathbb{1} = 1,$$

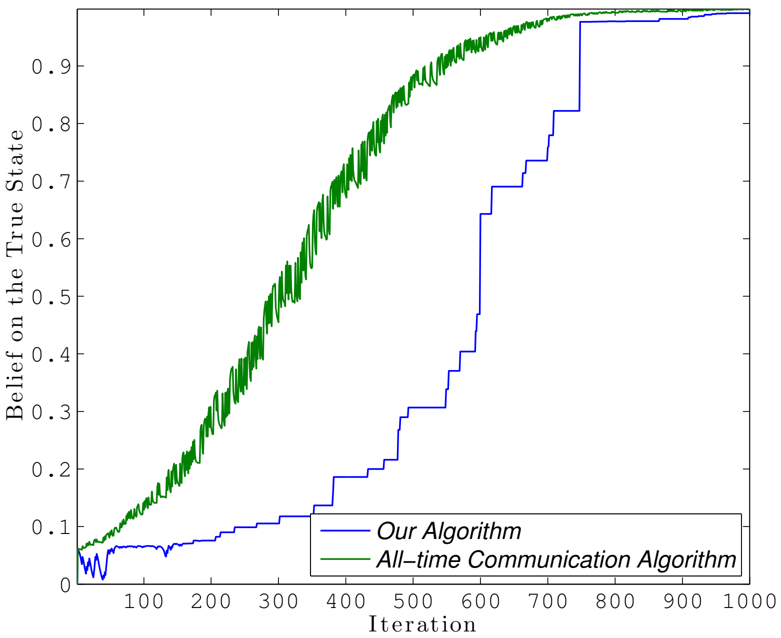

Figure 2.5: The comparison of belief evolution for a randomly selected agent in the network. The blue curve is generated under the switching protocol, while the green one is based on the all-time communication scheme .

respectively, for any $k \in [m]$. Combining the equations above and noting that μ_0 is uniform, we have

$$\frac{1}{m}\exp\{-\eta\lambda - 1\}\sum_{k=1}^{m}\exp\{\eta\phi_t(k)\} = 1,$$

which allows us to solve for λ and calculate the optimal solution μ_t as follows,

$$\mu_t(k) = \frac{\exp\{\eta\phi_t(k)\}}{\sum_{k=1}^{m}\exp\{\eta\phi_t(k)\}}.$$

The proof for $\mu_{i,t}$ follows precisely in the same fashion. To calculate $\phi_{i,t}$, notice that in view of the first update in (2.2.5) we have

$$\begin{bmatrix} \phi_{1,t} \\ \phi_{2,t} \\ \vdots \\ \phi_{n,t} \end{bmatrix} = (W \otimes I_m) \begin{bmatrix} \phi_{1,t-1} \\ \phi_{2,t-1} \\ \vdots \\ \phi_{n,t-1} \end{bmatrix} + \begin{bmatrix} \psi_{1,t} \\ \psi_{2,t} \\ \vdots \\ \psi_{n,t} \end{bmatrix},$$

where $\otimes$ denotes the Kronecker product. The equation above represents a discrete-time linear system. Given the fact that $\phi_{i,0}(k) = 0$ for all $k \in [m]$ and $i \in [n]$, the closed-form solution of the system takes the form

$$\begin{bmatrix} \phi_{1,t} \\ \phi_{2,t} \\ \vdots \\ \phi_{n,t} \end{bmatrix} = \sum_{\tau=1}^{t} (W \otimes I_n)^{t-\tau} \begin{bmatrix} \psi_{1,\tau} \\ \psi_{2,\tau} \\ \vdots \\ \psi_{n,\tau} \end{bmatrix} = \sum_{\tau=1}^{t} \left(W^{t-\tau} \otimes I_n\right) \begin{bmatrix} \psi_{1,\tau} \\ \psi_{2,\tau} \\ \vdots \\ \psi_{n,\tau} \end{bmatrix}.$$

Therefore, extracting $\phi_{i,t}$ for each $i \in [n]$ from the preceding relation completes the proof. ■

Proof of Lemma 2.2. Since the network is strongly connected and the corresponding W is irreducible and aperiodic, by standard properties of stochastic matrices (see e.g. [46]), the diagonalizable matrix W satisfies

$$\left\| \mathbf{e}_i^\top W^t - \pi^\top \right\|_1 \le n\lambda_{\max}(W)^t, \tag{2.7.1}$$

for any $i \in [n]$, where π is the stationary distribution of a Markov chain with transition kernel W. Let us observe the following inequality

$$n\lambda_{\max}(W)^{t-\tau} \le 2 \qquad \text{for} \qquad t - \tau \ge \tilde{t} := \frac{\log\left[\frac{n}{2}\right]}{\log \lambda_{\max}(W)^{-1}},$$

and recall that the inequality $\left\|\mathbf{e}_i^\top W^{t-\tau} - \pi^\top\right\|_1 \leq 2$ always holds since any power of W is stochastic. With that in mind, we use (4.5.1) to break the following sum into two parts to get

$$\begin{aligned}
\sum_{\tau=1}^{t}\sum_{j=1}^{n}\left|\left[W^{t-\tau}\right]_{ij} - \pi(j)\right| &= \sum_{\tau=1}^{t}\left\|\mathbf{e}_i^\top W^{t-\tau} - \pi^\top\right\|_1 \\
&= \sum_{\tau=1}^{t-\tilde{t}}\left\|\mathbf{e}_i^\top W^{t-\tau} - \pi^\top\right\|_1 + \sum_{\tau=t-\tilde{t}+1}^{t}\left\|\mathbf{e}_i^\top W^{t-\tau} - \pi^\top\right\|_1 \\
&\leq \sum_{\tau=1}^{t-\tilde{t}} n\lambda_{\max}(W)^{t-\tau} + 2\tilde{t} \\
&\leq \frac{n\lambda_{\max}(W)^{\tilde{t}}}{1-\lambda_{\max}(W)} + \frac{2\log\frac{n}{2}}{\log\lambda_{\max}(W)^{-1}},
\end{aligned}$$

for any $i \in [n]$. Note that $1-\lambda_{\max}(W) \leq \log\lambda_{\max}(W)^{-1}$ and $2 + 2\log\frac{n}{2} \leq 4\log n$, since $n > 1$. It follows by plugging $\tilde{t}$ into above that

$$\sum_{\tau=1}^{t}\sum_{j=1}^{n}\left|\left[W^{t-\tau}\right]_{ij} - \pi(j)\right| = \sum_{\tau=1}^{t}\left\|\mathbf{e}_i^\top W^{t-\tau} - \pi^\top\right\|_1 \leq \frac{4\log n}{1-\lambda_{\max}(W)},$$

which completes the proof. ■

We use the following inequality in [56] in the proof of Lemma 2.3.

Lemma 2.6. (McDiarmid's Inequality) *Let $X_1, ..., X_N \in \chi$ be independent random variables and consider the mapping $H : \chi^N \mapsto \mathbb{R}$. If for $i \in \{1, ..., N\}$, and every sample $x_1, ..., x_N, x_i' \in \chi$, the function H satisfies*

$$\left|H(x_1, ..., x_{i-1}, x_i, x_{i+1}, ..., x_N) - H(x_1, ..., x_{i-1}, x_i', x_{i+1}, ..., x_N)\right| \leq c_i,$$

then for all $\varepsilon > 0$,

$$\mathbb{P}\Big\{H(x_1, ..., x_N) - \mathbb{E}\left[H(X_1, ..., X_N)\right] \geq \varepsilon\Big\} \leq \exp\left\{\frac{-2\varepsilon^2}{\sum_{i=1}^{N} c_i^2}\right\}.$$

Proof of Lemma 2.3. According to Lemma 2.1, we have

$$\begin{aligned}\mu_{i,t}(1) &= \frac{\exp\{\eta\phi_{i,t}(1)\}}{\sum_{k=1}^{m}\exp\{\eta\phi_{i,t}(k)\}} \\ &= \left(1+\sum_{k=2}^{m}\exp\{\eta\phi_{i,t}(k)-\eta\phi_{i,t}(1)\}\right)^{-1} \\ &\geq 1-\sum_{k=2}^{m}\exp\{\eta\phi_{i,t}(k)-\eta\phi_{i,t}(1)\}, \qquad (2.7.2)\end{aligned}$$

where we used the fact that $(1+x)^{-1} \geq 1-x$ for any $x \geq 0$. Since we know

$$\|\mu_{i,t}-\mathbf{e}_1\|_{\text{TV}} = \frac{1}{2}\left(1-\mu_{i,t}(1)+\sum_{k=2}^{m}\mu_{i,t}(k)\right) = 1-\mu_{i,t}(1),$$

we can combine above with (2.7.2) to obtain

$$\|\mu_{i,t}-\mathbf{e}_1\|_{\text{TV}} \leq \sum_{k=2}^{m}\exp\{\eta\phi_{i,t}(k)-\eta\phi_{i,t}(1)\}. \qquad (2.7.3)$$

For any $k \in [m]$, define

$$\Phi_{i,t}(k) := \sum_{\tau=1}^{t}\sum_{j=1}^{n}\left[W^{t-\tau}\right]_{ij}\log\ell_j(\cdot|\theta_k),$$

and note that $\Phi_{i,t}(k)$ is a function of nt random variables. As required in McDiarmid's inequality in Lemma 2.6, set $H=\Phi_{i,t}(k)$, fix the samples for $nt-1$ random variables, and draw two different samples $s_{j,\tau}$ and $s'_{j,\tau}$ for some $j \in [n]$ and some $\tau \in [t]$. The fixed samples are simply cancelled in the subtraction, and we have

$$\begin{aligned}\left|H(...,s_{j,\tau},...)-H(...,s'_{j,\tau},...)\right| &= \left|\left[W^{t-\tau}\right]_{ij}\left(\log\ell_j(s_{j,t}|\theta_k)-\log\ell_j(s'_{j,t}|\theta_k)\right)\right| \\ &\leq \left[W^{t-\tau}\right]_{ij}2B,\end{aligned}$$

where we used assumption **A1**. Since any power of W is stochastic, summing over $j \in [n]$ and $\tau \in [t]$, we get

$$\sum_{\tau=1}^{t}\sum_{j=1}^{n}\left(\left[W^{t-\tau}\right]_{ij}2B\right)^2 \leq 4B^2t.$$

We now apply McDiarmid's inequality in Lemma 2.6 to obtain

$$\mathbb{P}\Big(\phi_{i,t}(k)-\phi_{i,t}(1)>\mathbb{E}\left[\Phi_{i,t}(k)\right]-\mathbb{E}\left[\Phi_{i,t}(1)\right]+\varepsilon\Big)\le\exp\left\{\frac{-\varepsilon^2}{2B^2t}\right\},$$

for each fixed k. Setting the probability above to δ/m and taking a union bound over all states, the following event holds

$$\phi_{i,t}(k)-\phi_{i,t}(1)\le\mathbb{E}\left[\Phi_{i,t}(k)\right]-\mathbb{E}\left[\Phi_{i,t}(1)\right]+\sqrt{2B^2t\log\frac{m}{\delta}}, \tag{2.7.4}$$

simultaneously for all $k=2,...,m$, with probability at least $1-\delta$. On the other hand, in view of assumption **A1**, we have

$$\begin{aligned}\mathbb{E}\left[\Phi_{i,t}(k)-\Phi_{i,t}(1)\right]&=\sum_{\tau=1}^{t}\sum_{j=1}^{n}\left[W^{t-\tau}\right]_{ij}\mathbb{E}\left[\log\ell_j(\cdot|\theta_k)-\log\ell_j(\cdot|\theta_1)\right]\\&=\sum_{\tau=1}^{t}\sum_{j=1}^{n}\Big(\left[W^{t-\tau}\right]_{ij}-\pi(j)\Big)\mathbb{E}\left[\log\ell_j(\cdot|\theta_k)-\log\ell_j(\cdot|\theta_1)\right]\\&\qquad\qquad+\sum_{\tau=1}^{t}\sum_{j=1}^{n}\pi(j)\mathbb{E}\left[\log\ell_j(\cdot|\theta_k)-\log\ell_j(\cdot|\theta_1)\right]\\&\le 2B\sum_{\tau=1}^{t}\sum_{j=1}^{n}\Big|\left[W^{t-\tau}\right]_{ij}-\pi(j)\Big|-t\sum_{j=1}^{n}\pi(j)D_{KL}\left(\ell_j(\cdot|\theta_1)\|\ell_j(\cdot|\theta_k)\right)\\&=2B\sum_{\tau=1}^{t}\sum_{j=1}^{n}\Big|\left[W^{t-\tau}\right]_{ij}-\pi(j)\Big|-\mathcal{I}(\theta_1,\theta_k)t \qquad (2.7.5)\\&\le\frac{8B\log n}{1-\lambda_{\max}(W)}-\mathcal{I}(\theta_1,\theta_k)t,\end{aligned}$$

where we applied Lemma 2.2 to derive the last step. Using (2.2.4), we simplify above to get

$$\mathbb{E}\left[\Phi_{i,t}(k)-\Phi_{i,t}(1)\right]\le\frac{8B\log n}{1-\lambda_{\max}(W)}-\mathcal{I}(\theta_1,\theta_2)t, \tag{2.7.6}$$

for any $k=2,...,m$. Plugging (2.7.6) into (2.7.4) and combining with (2.7.3), we have

$$\begin{aligned}\|\mu_{i,t}-\mathbf{e}_1\|_{\mathrm{TV}}&\le\sum_{k=2}^{m}\exp\left\{-\eta\mathcal{I}(\theta_1,\theta_2)t+\eta\sqrt{2B^2t\log\frac{m}{\delta}}+\frac{8\eta B\log n}{1-\lambda_{\max}(W)}\right\}\\&\le m\exp\left\{-\eta\mathcal{I}(\theta_1,\theta_2)t+\eta\sqrt{2B^2t\log\frac{m}{\delta}}+\frac{8\eta B\log n}{1-\lambda_{\max}(W)}\right\},\end{aligned}$$

with probability at least $1-\delta$, and thereby completing the proof of the first part. Letting $\delta = 1/t^2$ in above and applying Borel-Cantelli lemma, the almost sure convergence follows immediately. ■

Proof of Lemma 2.4. We recall from the statement of the lemma that $q_{i,t}(k) = \phi_{i,t}(k) - \phi_t(k)$, and calculate the ratio $\mu_{i,t}(k)/\mu_t(k)$ for any $k \in [m]$ as follows,

$$\begin{aligned}\frac{\mu_{i,t}(k)}{\mu_t(k)} &= \exp\{\eta q_{i,t}(k)\} \frac{\mathbb{E}_{\mu_0}[\exp\{\eta\phi_t\}]}{\mathbb{E}_{\mu_0}[\exp\{\eta\phi_{i,t}\}]} \\ &= \exp\{\eta q_{i,t}(k)\} \frac{\mathbb{E}_{\mu_0}[\exp\{\eta\phi_t\}]}{\mathbb{E}_{\mu_0}[\exp\{\eta\phi_t\}\exp\{\eta q_{i,t}\}]} \\ &= \exp\{\eta q_{i,t}(k)\} \frac{1}{\mathbb{E}_{\mu_0}\left[\frac{\exp\{\eta\phi_t\}}{\mathbb{E}_{\mu_0}[\exp\{\eta\phi_t\}]}\exp\{\eta q_{i,t}\}\right]} \\ &= \exp\{\eta q_{i,t}(k)\} \frac{1}{\mathbb{E}_{\mu_0}\left[\frac{\mu_t}{\mu_0}\exp\{\eta q_{i,t}\}\right]} \\ &= \exp\{\eta q_{i,t}(k)\} \frac{1}{\mathbb{E}_{\mu_t}[\exp\{\eta q_{i,t}\}]}.\end{aligned}$$

This entails

$$\frac{1}{\eta}\mathbb{E}_{\mu_{i,t}}\left[\log\frac{\mu_{i,t}}{\mu_t}\right] = \mathbb{E}_{\mu_{i,t}}[q_{i,t}] - \frac{1}{\eta}\log\mathbb{E}_{\mu_t}[\exp\{\eta q_{i,t}\}] \leq \mathbb{E}_{\mu_{i,t}}[q_{i,t}] - \mathbb{E}_{\mu_t}[q_{i,t}],$$

where we used Jensen's inequality on the convex function $-\log(\cdot)$. Setting the expectation measures in the right hand side of above to μ_t, and recalling the ratio $\mu_{i,t}/\mu_t$ from above, we conclude that,

$$\begin{aligned}
\mathbb{E}_{\mu_{i,t}}\left[\log\frac{\mu_{i,t}}{\mu_t}\right] &\leq \mathbb{E}_{\mu_t}\left[\frac{\mu_{i,t}}{\mu_t}\eta q_{i,t}\right] - \mathbb{E}_{\mu_t}\left[\eta q_{i,t}\right]\\
&= \mathbb{E}_{\mu_t}\left[\left(\frac{\exp\{\eta q_{i,t}\}}{\mathbb{E}_{\mu_t}\left[\exp\{\eta q_{i,t}\}\right]} - 1\right)\eta q_{i,t}\right]\\
&= \sum_{k=1}^{m}\mu_t(k)\eta q_{i,t}(k)\left(\frac{\exp\{\eta q_{i,t}(k)\}}{\mathbb{E}_{\mu_t}\left[\exp\{\eta q_{i,t}\}\right]} - 1\right)\\
&= \sum_{k=1}^{m}\mu_t(k)\eta q_{i,t}(k)\frac{\langle \mathbf{e}_k - \mu_t, \exp\{\eta q_{i,t}\}\rangle}{\langle \mu_t, \exp\{\eta q_{i,t}\}\rangle}\\
&\leq \frac{\exp\{\frac{1}{4}\}}{4}\sum_{k=1}^{m}\mu_t(k)\left|\langle \mathbf{e}_k - \mu_t, \exp\{\eta q_{i,t}\}\rangle\right|,
\end{aligned}$$

where we used the condition $\eta\|q_{i,t}\|_\infty \leq 1/4$ to obtain the last line. We now apply Hölder's inequality for primal-dual norm pairs and use $\eta\|q_{i,t}\|_\infty \leq 1/4$ again to simplify above as follows

$$\begin{aligned}
\mathbb{E}_{\mu_{i,t}}\left[\log\frac{\mu_{i,t}}{\mu_t}\right] &\leq \frac{\exp\{\frac{1}{4}\}}{4}\sum_{k=1}^{m}\mu_t(k)\,\|\mathbf{e}_k - \mu_t\|_1\,\|\exp\{\eta q_{i,t}\}\|_\infty\\
&\leq \frac{\exp\{\frac{1}{2}\}}{4}\sum_{k=1}^{m}\mu_t(k)\,\|\mathbf{e}_k - \mu_t\|_1\\
&\leq \frac{\exp\{\frac{1}{2}\}}{4}\,\|\mathbf{e}_1 - \mu_t\|_1 + \frac{\exp\{\frac{1}{2}\}}{2}\sum_{k=2}^{m}\mu_t(k),
\end{aligned} \tag{2.7.7}$$

where the last step follows from the fact that $\|\mathbf{e}_k - \mu_t\|_1 \leq 2$ for any $k \in [m]$. Recalling

$$\begin{aligned}
\frac{1}{2}\|\mathbf{e}_1 - \mu_t\|_1 &= \frac{1}{2}\left(1 - \mu_t(1) + \sum_{k=2}^{m}\mu_t(k)\right)\\
&= \frac{1}{2}\left(\sum_{k=1}^{m}\mu_t(k) - \mu_t(1) + \sum_{k=2}^{m}\mu_t(k)\right) = \sum_{k=2}^{m}\mu_t(k),
\end{aligned}$$

as well as the fact $\|\mathbf{e}_1 - \mu_t\|_{\mathrm{TV}} = \frac{1}{2}\|\mathbf{e}_1 - \mu_t\|_1$, we simplify (2.7.7) to get

$$\mathbb{E}_{\mu_{i,t}}\left[\log\frac{\mu_{i,t}}{\mu_t}\right] \leq \exp\left\{\frac{1}{2}\right\}\|\mathbf{e}_1 - \mu_t\|_{\mathrm{TV}} \leq 2\|\mathbf{e}_1 - \mu_t\|_{\mathrm{TV}}, \tag{2.7.8}$$

and thereby completing the proof. ■

Proof of Theorem 2.1. We recall that $q_{i,t}$ in the statement of Lemma 2.4 satisfies

$$\begin{aligned}\|q_{i,t}\|_\infty &= \left\| \sum_{\tau=1}^{t} \sum_{j=1}^{n} \left(\left[W^{t-\tau}\right]_{ij} - \pi(j) \right) \psi_{j,t} \right\|_\infty \\ &\leq B \sum_{\tau=1}^{t} \sum_{j=1}^{n} \left| \left[W^{t-\tau}\right]_{ij} - \pi(j) \right| \leq \frac{4B\log n}{1-\lambda_{\max}(W)},\end{aligned}$$

due to Lemma 2.2 and assumption **A1**. Therefore, the choice of $\eta = \frac{1-\lambda_{\max}(W)}{16B\log n}$ guarantees that $q_{i,t}$ satisfies $\eta\|q_{i,t}\|_\infty \leq 1/4$ for all $t \in [T]$.

Let us follow exactly the same steps in the proof of Lemma 2.3, and note that the centralized update can be recovered using $W = \mathbb{1}\pi^\top$. It can be verified from (2.7.5) that for any $t \in [T]$, we only remain with

$$\mathbb{E}\left[\Phi_t(k) - \Phi_t(1)\right] \leq -\mathcal{I}(\theta_1,\theta_2)t,$$

which yields

$$\frac{1}{\eta}\log\|\mu_t - \mathbf{e}_1\|_{\text{TV}} \leq -\mathcal{I}(\theta_1,\theta_2)t + \sqrt{2B^2 t \log\frac{m}{\delta_t}} + \frac{\log m}{\eta}, \tag{2.7.9}$$

with probability at least $1-\delta_t$. To have the above work for all $t \in [T]$ (simultaneously) with probability at least $1-\delta$, we need to take a union bound over any $t \in [T]$. Therefore, we have to choose $\{\delta_t\}_{t=1}^T$ such that $\sum_{t=1}^T \delta_t \leq \delta$. Letting $\delta_t := \delta\exp\left\{-t^{1/3}\right\}/6$, we have

$$\sum_{t=1}^{T} \delta_t \leq \frac{\delta}{6}\int_0^\infty \exp\left\{-t^{\frac{1}{3}}\right\} d_t = \frac{\delta}{6}\int_0^\infty 3u^2\exp\left\{-u\right\} d_u = \frac{\delta}{6}3! = \delta. \tag{2.7.10}$$

Let us avoid notational clutter, by defining $a := \mathcal{I}(\theta_1,\theta_2)$, $b := \left(2B^2\log\left[6m/\delta\right]\right)^{1/2}$ and $c :=$

$\sqrt{2}B$, respectively. Then, in view of (2.7.9) and Lemma 2.4, with probability at least $1-\delta_t$ we have

$$\begin{aligned}
D_{KL}(\mu_{i,t}\|\mu_t) &\le 2\|\mathbf{e}_1-\mu_t\|_{\text{TV}}\\
&\le 2m\exp\left\{\eta\left(-at+bt^{\frac{1}{2}}+ct^{\frac{2}{3}}\right)\right\}\\
&\le 2m\exp\left\{-\frac{a}{3}\eta t\right\} \text{ for } t\ge t_1:=\max\left\{\left(\frac{3b}{a}\right)^2,\left(\frac{3c}{a}\right)^3\right\}\\
&\le 2, \text{ for } t\ge t_2:=\frac{3}{a\eta}\log m.
\end{aligned}$$

Let $t_0=\max\{t_1,t_2\}$, note all the inequalities above together, and observe the fact that $\|\mathbf{e}_1-\mu_t\|_{\text{TV}}\le 1$ for any $t\in[T]$. Also, recall the proper choice of δ_t for the union bound (2.7.10) to get

$$\begin{aligned}
\mathbf{Cost}_{i,T}=\sum_{t=1}^{T}D_{KL}(\mu_{i,t}\|\mu_t) &\le 2\sum_{t=1}^{t_0}\|\mathbf{e}_1-\mu_t\|_{\text{TV}}+2\sum_{t=t_0+1}^{T}m\exp\left\{-\frac{a}{3}\eta t\right\}\\
&\le 2t_0+2\sum_{t=t_2+1}^{T}m\exp\left\{-\frac{a}{3}\eta t\right\}\\
&\le 2t_0+2\int_{t_2}^{\infty}m\exp\left\{-\frac{a}{3}\eta t\right\}d_t=2t_0+\frac{6}{a\eta},
\end{aligned}$$

with probability at least $1-\delta$. Plugging our choice of η into above completes the proof. ■

Proof of Lemma 2.5. Given the hypothesis, agent i follows the Bayesian update after $\hat{t}$, and we have

$$\mu_{i,t}(k)=\frac{\mu_{i,t-1}(k)\ell_i(s_{i,t}|\theta_k)}{\sum_{k'\in[m]}\mu_{i,t-1}(k')\ell_i(s_{i,t}|\theta_{k'})},$$

for any $k\in[m]$ and $t\ge\hat{t}$. Recalling that θ_1 denotes the true state, we can write for any $t>\hat{t}$ and $k\neq 1$,

$$\log\frac{\mu_{i,t}(k)}{\mu_{i,t}(1)}=\log\frac{\mu_{i,t-1}(k)}{\mu_{i,t-1}(1)}+\log\frac{\ell_i(s_{i,t}|\theta_k)}{\ell_i(s_{i,t}|\theta_1)}. \tag{2.7.11}$$

Therefore, for any $\theta_k \in \bar{\Theta}_i$, we have

$$\frac{\mu_{i,t}(k)}{\mu_{i,t}(1)} = \frac{\mu_{i,\hat{t}}(k)}{\mu_{i,\hat{t}}(1)},$$

for all $t > \hat{t}$, since in (2.7.11) the likelihood ratio is one, and $\log \frac{\ell_i(s_{i,t}|\theta_k)}{\ell_i(s_{i,t}|\theta_1)} = 0$ by definition of observationally equivalent states. On the other hand, for any $\theta_k \in \Theta \setminus \bar{\Theta}_i$ simplifying (2.7.11) and dividing by t, we obtain for all $t > \hat{t}$

$$\begin{aligned}\frac{1}{t} \log \frac{\mu_{i,t}(k)}{\mu_{i,t}(1)} &= \frac{1}{t} \log \frac{\mu_{i,\hat{t}}(k)}{\mu_{i,\hat{t}}(1)} + \frac{1}{t} \sum_{\tau=\hat{t}+1}^{t} \log \frac{\ell_i(s_{i,\tau}|\theta_k)}{\ell_i(s_{i,\tau}|\theta_1)} \\ &\longrightarrow \mathbb{E}\left[\log \frac{\ell_i(\cdot|\theta_k)}{\ell_i(\cdot|\theta_1)}\right] \\ &= -D_{KL}\left(\ell_i(\cdot|\theta_1) \| \ell_i(\cdot|\theta_k)\right) < 0,\end{aligned}$$

almost surely by the Strong Law of Large Numbers (SLLN). Note that since the signals are i.i.d. over time and the log-marginals are bounded (assumption **A1**), SLLN could be applied. The above entails that $\mu_{i,t}(k) \longrightarrow 0$ for any $\theta_k \in \Theta \setminus \bar{\Theta}_i$, and thereby completing the proof. ■

Proof of Theorem 2.3. Fix any time $t_0 \in \mathbb{N}$. When an agent uses Bayes' rule for $t \geq t_0$, in view of Lemma 2.5, the condition $\|\mu_{i,t} - \mu_{i,t-1}\|_{TV} < \tau$ will be satisfied in a finite (random) time due to almost sure convergence of Bayes' rule. Therefore, all neighboring agents will eventually communicate with each other in the interval $[t_0, \infty)$. On the other hand, the underlying graph $\mathcal{G}$ is strongly connected by assumption **A3**; hence, all the conditions of Proposition 2.3 are satisfied, and the left product $W(t)W(t-1)\cdots W(1)$ has a limit, and since the matrices in the sequence $\{W(t)\}_{t=1}^{\infty}$ are doubly stochastic by the proposed switching rule, we get

$$\prod_{\rho=0}^{t-1} W(t-\rho) \longrightarrow \frac{1}{n}\mathbb{1}\mathbb{1}^{\top}, \tag{2.7.12}$$

almost surely. We recall that Lemma 2.1 provides a closed-form solution of (2.5.1) for when $W(t) = W$. The closed-form of (2.5.1), itself, can be derived in a similar fashion, and we get

$$\begin{aligned}\frac{1}{t}\phi_{i,t}(k) &= \frac{1}{t}\sum_{\tau=0}^{t}\sum_{j=1}^{n}\left[\prod_{\rho=0}^{t-1-\tau} W(t-\rho)\right]_{ij} \log \ell_j(s_{j,\tau}|\theta_k) \\ &= \frac{1}{nt}\sum_{\tau=0}^{t}\sum_{j=1}^{n} \log \ell_j(s_{j,\tau}|\theta_k) + e_{i,t}(k), \qquad (2.7.13)\end{aligned}$$

where

$$e_{i,t}(k) = \frac{1}{t}\sum_{\tau=0}^{t}\sum_{j=1}^{n}\left(\left[\prod_{\rho=0}^{t-1-\tau} W(t-\rho)\right]_{ij} - \frac{1}{n}\right) \log \ell_j(s_{j,\tau}|\theta_k).$$

Since the log-marginals are bounded (assumption **A1**), in view of (2.7.12) we get

$$\begin{aligned}|e_{i,t}(k)| &\le \frac{B}{t}\sum_{\tau=0}^{t}\sum_{j=1}^{n}\left|\left[\prod_{\rho=0}^{t-1-\tau} W(t-\rho)\right]_{ij} - \frac{1}{n}\right| \\ &\longrightarrow 0, \qquad (2.7.14)\end{aligned}$$

as $t \to \infty$, since Cesàro mean preserves the limit. Also, applying SLLN we get

$$\frac{1}{nt}\sum_{\tau=0}^{t}\sum_{j=1}^{n} \log \ell_j(s_{j,\tau}|\theta_k) \longrightarrow \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}\left[\log \ell_j(\cdot|\theta_k)\right],$$

almost surely. Combining above with (2.7.13) and (2.7.14) and recalling the definition of $\mathcal{I}(\theta_1, \theta_k)$ in Lemma 2.3, we derive

$$\frac{1}{t}\phi_{i,t}(k) - \frac{1}{t}\phi_{i,t}(1) \longrightarrow -\mathcal{I}(\theta_1, \theta_k), \qquad (2.7.15)$$

almost surely, which guarantees that

$$e^{\phi_{i,t}(k)-\phi_{i,t}(1)} \longrightarrow 0, \qquad (2.7.16)$$

for any $k \neq 1$, since $\mathcal{I}(\theta_1, \theta_k) > 0$ due to global identifiability of θ_1 (assumption **A2**). Now observe that

$$\mu_{i,t}(1) = \frac{e^{\phi_{i,t}(1)}}{\sum_{k=1}^{m} e^{\phi_{i,t}(k)}} = \frac{1}{1 + \sum_{k=2}^{m} e^{\phi_{i,t}(k)-\phi_{i,t}(1)}}. \qquad (2.7.17)$$

Taking the limit and using (2.7.16), the proof of convergence follows immediately, and per (2.7.15) this convergence is exponentially fast with the asymptotic rate $\mathcal{I}(\theta_1, \theta_2)$ corresponding to the slowest vanishing summand in the denominator of (2.7.17). ■

Chapter 3

Inverse Problem : Network Identification

In the previous chapter we considered an information aggregation procedure over networks. We focused on a setting where the network structure is *given*, and the algorithm outputs beliefs accordingly. However, we now aim to address an inverse-type problem: what would happen if the outputs of an update (say a consensus algorithm) are *given*, and the network structure is *unknown*? Can we reconstruct the network topology if we measure the outputs? We are interested to find the answer to these questions in this chapter.

The reconstruction of networks of dynamical systems is an important task in many realms of science and engineering, including biology, physics and finance [57–61]. Networked dynamical systems have been widely used to study the phenomenon of synchronization [62, 63]. Motivated by this line of research, we propose several algorithms to reconstruct the structure of a directed network of interconnected linear dynamical systems. We begin with an algorithm to find the Boolean structure of the unknown topology. This algorithm is based on the analysis of power spectral properties of the network response when the inputs are wide-sense stationary (WSS) processes of an *unknown* power spectral density (PSD). The measurements are performed via a *node-knockout* procedure inspired by work of Nabi-Abdolyousefi and Mesbahi [64]. Apart from recovering the Boolean structure of the network, we propose another algorithm to recover the exact structure of the network (including edge weights) when an eigenvalue-eigenvector pair of the connectivity matrix is known. This algorithm can be applied, for example, in the case of the connectivity matrix being a Laplacian matrix or the adjacency of a regular graph. Apart from general directed networks, we also propose reconstruction methodologies for directed nonreciprocal networks (networks with no directed edges pointing in opposite directions) and undirected networks. In the latter cases, we propose specialized algorithms able to recover the network structure with less computational cost.

This chapter is organized as follows. In Section 3.1, we introduce some preliminary definitions needed in our exposition and describe the network reconstruction problem under consideration. Section 3.2 provides several theoretical results that are the foundation for our reconstruction techniques. In Section 3.3, we introduce several algorithms to reconstruct the Boolean structure of a directed network, the exact structure of a directed network given an eigenvalue-eigenvector pair, and the structure of undirected and nonreciprocal networks. We also provide an overview of relevant works in Section 3.4. The content of the chapter is mostly from the work of Shahrampour and Preciado [65].

3.1 Preliminaries and Problem Description

I_d	$d \times d$ identity matrix.
$\mathbb{1}_d$	d-dimensional vector of all ones.
$\mathbf{e}_k$	k-th unit vector in the standard basis of $\mathbb{R}^N$.
$\mathbb{E}(\cdot)$	Expectation operator.
$R_{xy}(\tau)$	Cross-correlation function, $\mathbb{E}(x(t)y(t-\tau))$.
$R_x(\tau)$	Auto-correlation function, $\mathbb{E}(x(t)x(t-\tau))$.
$\mathcal{F}\{\cdot\}$	Fourier transform.
$S_{y_iy_j}(\omega)$	Cross-power spectral density (CPSD), $\mathcal{F}\big\{R_{y_iy_j}(\tau)\big\}$.
$S_{y_i}(\omega)$	Power spectral density (PSD), $\mathcal{F}\big\{R_{y_iy_i}(\tau)\big\}$.

Table 3.1: Nomenclature

3.1.1 Graph Theory

A weighted, *directed* graph is defined as the triad $\mathcal{D} := (\mathcal{V}, \mathcal{E}_d, \mathcal{F}_d)$, where $\mathcal{V} := \{v_1, \ldots, v_N\}$ denotes a set of N nodes and $\mathcal{E}_d \subseteq \mathcal{V} \times \mathcal{V}$ denotes a set of m directed edges in $\mathcal{D}$. The function $\mathcal{F}_d : \mathcal{E}_d \to \mathbb{R}_{++}$ associates *positive* real weights to the edges. We define the weighted *in-degree* of node v_i as

$$\deg_{in}(v_i) = \sum_{j:(v_j,v_i)\in\mathcal{E}_d} \mathcal{F}_d\left((v_j, v_i)\right).$$

The *adjacency matrix* of a weighted, directed graph $\mathcal{D}$, denoted by $A_{\mathcal{D}} = [a_{ij}]$, is a $N \times N$ matrix defined entry-wise as $a_{ij} = \mathcal{F}_d((v_j, v_i))$ if edge $(v_j, v_i) \in \mathcal{E}_d$, and $a_{ij} = 0$ otherwise. We define the *Laplacian matrix* $L_{\mathcal{D}}$ as $L_{\mathcal{D}} = \text{diag}(\deg_{in}(v_i)) - A_{\mathcal{D}}$. The Laplacian matrix satisfies $L_{\mathcal{D}}\mathbb{1} = \mathbf{0}$, i.e., the vector $\mathbb{1}/\sqrt{N}$ is an eigenvector of the Laplacian matrix with eigenvalue 0.

3.1.2 Dynamical Network Model

Consider a dynamical network consisting of N linearly coupled identical nodes, with each node being an n-dimensional, LTI, SISO dynamical system. The dynamical network under study can be characterized by

$$\dot{x}_i(t) = Ax_i(t) + b\left(\sum_{j=1}^{N} g_{ij}y_j(t) + w_i\left(t\right)\right), \tag{3.1.1}$$
$$y_i(t) = c^\top x_i(t),$$

where $x_i(t) \in \mathbb{R}^n$ denotes the state vector describing the dynamics of node $v_i \in \mathcal{V}$. $A \in \mathbb{R}^{n\times n}$ and $b, c \in \mathbb{R}^n$ are the given state, input and output matrices corresponding to the state-space representation of each node in isolation. $w_i\left(t\right)$ and $y_i(t) \in \mathbb{R}$ are stochastic processes representing the input noise and the system output, respectively, $g_{ij} \geq 0$ is the coupling strength of a *directed* edge from v_i to v_j, which we shall assume to be *unknown*. It is worth remarking that considering identical nodes allows us to use tensor notation that simplifies our technical analysis. Relaxing this assumption as well as studying coupling strengths of dynamic form are currently under investigation.

Defining the network state vector, the noise vector, and the network output vector as

$$\mathbf{x}(t) := (x_1^\top(t), \ldots, x_N^\top(t))^\top \in \mathbb{R}^{Nn}$$
$$\mathbf{w}(t) := (w_1(t), \ldots, w_N(t))^\top \in \mathbb{R}^{N}$$
$$\mathbf{y}(t) := (y_1(t), \ldots, y_N(t))^\top \in \mathbb{R}^{N},$$

respectively, we can rewrite the network dynamics in (3.1.1), as

$$\dot{\mathbf{x}}(t) = \left(I_N \otimes A + \mathbf{G} \otimes bc^\top\right)\mathbf{x}(t) + \left(I_N \otimes b\right)\mathbf{w}(t), \tag{3.1.2}$$
$$\mathbf{y}(t) = \left(I_N \otimes c^\top\right)\mathbf{x}(t),$$

where $\mathbf{G} = [g_{ij}]$ is the *connectivity* matrix of a (possibly weighted and/or directed) network $\mathcal{D}$. For the networked dynamical system to be stable, we assume the network state matrix $I_N \otimes A + \mathbf{G} \otimes bc^\top$ to be Hurwitz.

Hereafter, we will analyze the following scenario. Consider a collection of N dynamical nodes with a known LTI, SISO dynamics defined by the state-space matrices $(A, b, c^\top, 0)$. The link structure of the network dynamic model, described by the connectivity matrix $\mathbf{G}$, is completely unknown. We assume the input noises, $\{w_i(t)\}_{i=1}^N$, are i.i.d. wide-sense stationary processes of *unknown* but identical power spectral densities, i.e., $S_{w_i}(\omega) = S_w(\omega)$ for all $i = 1, \ldots, N$. We are interested in identifying all the links in the network by exploiting only the information provided by the realizations of the output stochastic processes $y_1(t), \ldots, y_N(t)$. Formally, we can formulate this problem as follows:

Problem 3.1. *Consider the dynamical network model in* (3.1.2)*, whose connectivity matrix* $\mathbf{G}$ *is unknown. Assume that the only available information is a spectral characterization of the output signals* $y_1(t), \ldots, y_N(t)$ *in terms of power and cross-power spectral densities,* $S_{y_i}(\omega)$ *and* $S_{y_i y_j}(\omega)$*, which can be empirically estimated from the output signals*[1]*. Then, find the Boolean structure of the directed network, i.e., the location and direction of each edge.*

It is worth remarking that we assume the input noise to be an exogenous signal of *unknown* power spectral density, $S_w(\omega)$.

[1] One can use, for example, Bartletts averaging method [66] to produce periodogram estimates of power and cross-power spectral densities, $S_{y_i}(\omega)$ and $S_{y_i y_j}(\omega)$.

3.2 The Relationship between Input-Output Power Spectral Densities

We start by stating some assumptions we need in our subsequent developments. The following definition will be useful for determining sufficient conditions for detection of links in a network.

Definition 3.1. *[Excitation Frequency Interval, [67]] The excitation frequency interval of a vector $\mathbf{w}(t)$ of wide-sense stationary processes is defined as an interval $(-\Omega, \Omega)$, with $\Omega > 0$, such that the power spectral densities of the input components $w_i(t)$ satisfy $S_{w_i}(\omega) > 0$ for all $\omega \in (-\Omega, \Omega)$, and all $i \in \{1, 2, ..., N\}$.*

Throughout, we impose the following conditions on the input vector:

A1. The collection of signals $\{w_i(t), i = 1, ..., N\}$ are uncorrelated, zero-mean WSS processes with identical autocorrelation function, i.e., for any $t, \tau \in \mathbb{R}$, $R_{w_i}(\tau) = \mathbb{E}(w_i(t)w_i(t+\tau)) := R_w(\tau)$.

A2. The input noise $\mathbf{w}(t)$ presents a nonempty excitation frequency interval $(-\Omega, \Omega)$.

In our derivation, we will invoke the following variation of the matrix inversion lemma [68]:

Lemma 3.1 (Sherman-Morrison-Woodbury)**.** *Assume that the matrices D and $I + WD^{-1}UE$ are nonsingular. Then, the following identity holds*

$$(D + UEW)^{-1} = D^{-1} - D^{-1}UE\left(I + WD^{-1}UE\right)^{-1} WD^{-1},$$

where E, W, D, and U are matrices of compatible dimensions and I is the identity matrix.

Based on Woodbury's formula, we derive an expression that provides an explicit relationship between the (cross-)power spectral densities of two stochastic outputs, $y_i(t)$ and $y_j(t)$, when we inject a noise $w_k(t)$ into node k with power spectral density $S_w(\omega)$.

Lemma 3.2. *Consider the continuous-time networked dynamical system* (3.1.2)*. Then, under assumptions (A1)-(A2), the following identity holds*

$$\mathbf{S}\left(\omega\right)=S_{w}(\omega)\left(\frac{I_{N}}{\left|h\left(\mathbf{j}\omega\right)\right|^{2}}+\mathbf{G}^{\top}\mathbf{G}-\frac{\mathbf{G}}{h^{*}\left(\mathbf{j}\omega\right)}-\frac{\mathbf{G}^{\top}}{h\left(\mathbf{j}\omega\right)}\right)^{-1}, \tag{3.2.1}$$

where $\mathbf{S}\left(\omega\right):=\left[S_{y_{i}y_{j}}(\omega)\right]$ *is the matrix of output CPSD's, and* $h\left(\mathbf{j}\omega\right):=c^{\top}\left(\mathbf{j}\omega I_{n}-A\right)^{-1}b$ *is the nodal transfer function.*

Proof. The $N\times N$ transfer matrix, $H\left(\mathbf{j}w\right):=\left[H_{ji}\left(\mathbf{j}\omega\right)\right]$, of the state-space model in (3.1.2) is given by

$$\begin{aligned}H\left(\mathbf{j}\omega\right)&=\left(I_{N}\otimes c^{\top}\right)\Big(\mathbf{j}\omega I_{Nn}-I_{N}\otimes A-\mathbf{G}\otimes bc^{\top}\Big)^{-1}\left(I_{N}\otimes b\right)\\&=\left(I_{N}\otimes c^{\top}\right)\Big(I_{N}\otimes\left(\mathbf{j}\omega I_{n}-A\right)-\mathbf{G}\otimes bc^{\top}\Big)^{-1}\left(I_{N}\otimes b\right).\end{aligned} \tag{3.2.2}$$

Assume that we inject a noise signal into the k-th node, i.e., $\mathbf{w}\left(t\right)=w_{k}\left(t\right)\mathbf{e}_{k}$. Hence, the power spectral density measured on the output of node i is equal to

$$S_{y_{i}}(\omega)=H_{ki}(\omega)H_{ki}^{*}(\omega)S_{w_{k}}(\omega).$$

On the other hand, the transfer functions from input $w_{k}\left(t\right)$ to the outputs $y_{i}\left(t\right)$ and $y_{j}\left(t\right)$ are, respectively,

$$\frac{Y_{i}\left(\mathbf{j}\omega\right)}{W_{k}\left(\mathbf{j}\omega\right)}=H_{ki}(\mathbf{j}\omega)\qquad\text{and}\qquad\frac{Y_{j}\left(\mathbf{j}\omega\right)}{W_{k}\left(\mathbf{j}\omega\right)}=H_{kj}(\mathbf{j}\omega),$$

where $Y_{i}\left(\mathbf{j}\omega\right)$ and $W_{k}\left(\mathbf{j}\omega\right)$ are the Fourier transforms of $y_{i}\left(t\right)$ and $w_{k}\left(t\right)$, respectively. Hence,

$$\frac{Y_{j}\left(\mathbf{j}\omega\right)}{Y_{i}\left(\mathbf{j}\omega\right)}=H_{ki}^{-1}(\mathbf{j}\omega)H_{kj}(\mathbf{j}\omega),$$

which implies

$$S_{y_{i}y_{j}}(\omega)=\Big(H_{kj}(\mathbf{j}\omega)H_{ki}^{-1}(\mathbf{j}\omega)\Big)^{*}S_{y_{i}}(\omega).$$

Since $S_{w_k}(\omega) = S_w(\omega)$ for all k, we have that $S_{y_i y_j}(\omega) = H_{ki}(\mathbf{j}\omega) H^*_{kj}(\mathbf{j}\omega) S_w(\omega)$. Assume that we inject noise signals satisfying assumptions (A1)-(A2) into all the nodes in the network, i.e., $\mathbf{w}\left(t\right) = \sum_{k=1}^{N} w_k\left(t\right) \mathbf{e}_k$. Hence, we can apply superposition to obtain

$$\begin{aligned}
\frac{S_{y_i y_j}(\omega)}{S_w(\omega)} &= \sum_{k=1}^{N} H^*_{kj}(\mathbf{j}\omega) H_{ki}(\mathbf{j}\omega) \\
&= \sum_{k=1}^{N} \mathbf{e}_k^\top H^*\left(\mathbf{j}\omega\right) \mathbf{e}_j \mathbf{e}_i^\top H\left(\mathbf{j}\omega\right) \mathbf{e}_k \\
&= \sum_{k=1}^{N} \mathrm{Tr}\Big(H^*\left(\mathbf{j}\omega\right) \mathbf{e}_j \mathbf{e}_i^\top H\left(\mathbf{j}\omega\right) \mathbf{e}_k \mathbf{e}_k^\top \Big) \\
&= \mathrm{Tr}\Big(H^*\left(\mathbf{j}\omega\right) \mathbf{e}_j \mathbf{e}_i^\top H\left(\mathbf{j}\omega\right) \sum_{k=1}^{N} \mathbf{e}_k \mathbf{e}_k^\top \Big) \\
&= \mathbf{e}_i^\top H\left(\mathbf{j}\omega\right) H^*\left(\mathbf{j}\omega\right) \mathbf{e}_j,
\end{aligned} \tag{3.2.3}$$

for any $\omega \in (-\Omega, \Omega)$, where we used the identity $\sum_{k=1}^{N} \mathbf{e}_k \mathbf{e}_k^\top = I_N$ in our derivation.

Let us define the matrices $W := I_N \otimes c^\top$, $U := I_N \otimes b$, $E := -\mathbf{G}$, and $D := I_N \otimes (\mathbf{j}\omega I_n - A)$. Then, we can rewrite the transfer matrix $H\left(\mathbf{j}\omega\right)$ in (3.2.2) as

$$H\left(\mathbf{j}\omega\right) = W(D + UEW)^{-1} U. \tag{3.2.4}$$

Also, we have that $h\left(\mathbf{j}\omega\right) I_N = W D^{-1} U$. Then, applying Lemma 3.1 to (3.2.4), we can rewrite the transfer matrix, as follows

$$\begin{aligned}
H\left(\mathbf{j}\omega\right) &= h\left(\mathbf{j}\omega\right) \left(I_N + \mathbf{G} \big(I_N - h\left(\mathbf{j}\omega\right) \mathbf{G} \big)^{-1} h\left(\mathbf{j}\omega\right) I_N \right) \\
&= h\left(\mathbf{j}\omega\right) \left(I_N + \mathbf{G} \left(\frac{I_N}{h\left(\mathbf{j}\omega\right)} - \mathbf{G} \right)^{-1} \right) \\
&= h\left(\mathbf{j}\omega\right) \left(I_N + \big(\mathbf{G} - \frac{I_N}{h\left(\mathbf{j}\omega\right)} + \frac{I_N}{h\left(\mathbf{j}\omega\right)} \big) \big(\frac{I_N}{h\left(\mathbf{j}\omega\right)} - \mathbf{G} \big)^{-1} \right) \\
&= h\left(\mathbf{j}\omega\right) \left(I_N - I_N + \frac{1}{h\left(\mathbf{j}\omega\right)} \big(\frac{I_N}{h\left(\mathbf{j}\omega\right)} - \mathbf{G} \big)^{-1} \right) \\
&= \left(\frac{I_N}{h\left(\mathbf{j}\omega\right)} - \mathbf{G} \right)^{-1}.
\end{aligned}$$

Substituting above into (3.2.3), we reach the statement of our lemma. □

In the following section, we will use this lemma to reconstruct an unknown network structure $\mathbf{G}$ from the empirical CPSD's of the outputs. We will also show that, assuming that we know one eigenvalue-eigenvector pair of $\mathbf{G}$, we can recover the weighted and directed graph $\mathcal{D}$ (not only its Boolean structure, but also its weights), as well as the PSD of the noise, $S_w(\omega)$. Relevant examples of this scenario are: (*i*) networks of diffusively coupled systems with a Laplacian connectivity matrix [69], i.e., $\mathbf{G} = -L_{\mathcal{D}}$, since Laplacian matrices always satisfy $L_{\mathcal{D}}\mathbb{1}_N = 0$; or (*ii*) k-regular networks [70], i.e., $\mathbf{G} = A_k$, since the adjacency matrix A_k satisfy $A_k\mathbf{1}_N = k$.

As stated in Problem 3.1, the PSD of the input noise $\mathbf{w}(t)$ is not available to us to perform the network reconstruction. The following lemma will allow us reconstruct this PSD when an eigenvalue-eigenvector pair of $\mathbf{G}$ is known *a priori*.

Lemma 3.3. *Consider the continuous-time networked dynamical system* (3.1.2). *Then, under assumptions (A1)-(A2), the input PSD can be computed as*

$$S_w(\omega) = \frac{\lambda^2|h(\mathbf{j}\omega)|^2 - 2\lambda Re\{h(\mathbf{j}\omega)\} + 1}{(\boldsymbol{u}^\top\mathbf{S}^{-1}(\omega)\boldsymbol{u})|h(\mathbf{j}\omega)|^2}, \tag{3.2.5}$$

where $(\lambda, \boldsymbol{u})$ *is an eigenvalue-eigenvector pair of* $\mathbf{G}$*,* $h(\mathbf{j}\omega)$ *is the nodal transfer function, and* $\mathbf{S}(\omega) := \left[S_{y_iy_j}(\omega)\right]$ *is the matrix of CPSD's.*

Proof. From (3.2.1), we have

$$\mathbf{S}^{-1}(\omega)\, S_w(\omega) = \frac{I_N}{|h(\mathbf{j}\omega)|^2} + G^\top G - \frac{G}{h^*(\mathbf{j}\omega)} - \frac{G^\top}{h(\mathbf{j}\omega)}.$$

Pre- and post-multiplying by $\boldsymbol{u}^\top$ and $\boldsymbol{u}$, respectively, we obtain

$$\left(\boldsymbol{u}^\top\mathbf{S}^{-1}(\omega)\,\boldsymbol{u}\right) S_w(\omega) = \frac{1}{|h(\mathbf{j}\omega)|^2} + \lambda^2 - \frac{\lambda}{h(\mathbf{j}\omega)} - \frac{\lambda}{h^*(\mathbf{j}\omega)}.$$

Dividing by $\boldsymbol{u}^\top\mathbf{S}^{-1}(\omega)\,\boldsymbol{u}$, we reach (3.2.5). □

Lemma 3.3 shows that, given the eigenvalue-eigenvector pair $(\lambda, \boldsymbol{u})$, the PSD of the input noise can be reconstructed from the nodal transfer function and the matrix of CPSD's, $\mathbf{S}(\omega)$, which can be numerically approximated from the empirical cross-correlations between output signals.

3.3 Reconstruction Methodologies

Based on the above results, we introduce several methodologies to reconstruct the structure of an unknown network following the dynamics in (3.1.2) when the PSD of the input noise is *unknown*.

Consider Problem 3.1, when $\mathbf{G}$ is an unknown connectivity matrix representing a weighted, directed network $\mathcal{D}$. We propose a reconstruction technique to recover the Boolean structure of $\mathcal{D}$ when the PSD of the input noise is unknown. Note that, in general, the result in Lemma 3.2 is not enough to extract the underlying structure of the network, even if the input noise PSD were known. In what follows, we propose a methodology to reconstruct a directed network of dynamical nodes by *grounding* the dynamics in a series of nodes, similar to the approach proposed in [64] to reconstruct undirected networks following a consensus dynamics.

Definition 3.2 (Grounded Dynamics). *The dynamics of* (3.1.2) *grounded at node* v_j *takes the form*

$$\dot{\widetilde{\mathbf{x}}}(t) = \left(I_{N-1} \otimes A + \widetilde{\mathbf{G}}_j \otimes bc^\top\right) \widetilde{\mathbf{x}}(t) + (I_{N-1} \otimes b)\, \widetilde{\mathbf{w}}(t), \tag{3.3.1}$$

$$\widetilde{\mathbf{y}}(t) = \left(I_{N-1} \otimes c^\top\right) \widetilde{\mathbf{x}}(t),$$

where $\widetilde{\mathbf{w}}(t)$ *is obtained by eliminating the* j*-th entry from the input noise* $\mathbf{w}(t)$*, and* $\widetilde{\mathbf{G}}_j \in \mathbb{R}^{(N-1)\times(N-1)}$ *is obtained by eliminating the* j*-th row and column from* $\mathbf{G}$.

The dynamics in (3.3.1) describes the evolution of (3.1.2) when we ground the state of node v_j to be $x_j(t) \equiv 0$. Applying Lemma 3.2 to the grounded dynamics (3.3.1), one obtains the following expression for the CPSD's:

$$\widetilde{\mathbf{S}}_j(\omega) = S_w(\omega)\left(\frac{I_{N-1}}{\left|h\left(\mathbf{j}\omega\right)\right|^2} + \widetilde{\mathbf{G}}_j^\top \widetilde{\mathbf{G}}_j - \frac{\widetilde{\mathbf{G}}_j}{h^*\left(\mathbf{j}\omega\right)} - \frac{\widetilde{\mathbf{G}}_j^\top}{h\left(\mathbf{j}\omega\right)}\right)^{-1}. \tag{3.3.2}$$

We will use the next Theorem to propose several reconstruction techniques in Subsections 3.3.1 and 3.3.2.

Theorem 3.1. *Consider the networked dynamical system* (3.1.2) *with connectivity matrix* $\mathbf{G} = [g_{ij}]$. *Let us denote by* $S_w(\omega)$ *the PSD of the input noise, by* $\mathbf{S}(\omega) = \left[S_{y_i y_j}(\omega)\right]$ *the* $N \times N$ *matrix of CPSD's for the (ungrounded) dynamics* (3.1.2)*, and by* $\widetilde{\mathbf{S}}_j(\omega) = [\widetilde{S}_{y_i y_k}(\omega)]_{i,k\neq j}$ *the* $N-1 \times N-1$ *matrix of CPSD's for the dynamics in* (3.3.1) *grounded at node* v_j. *Then, under assumptions (A1)-(A2), we have that, for* $i < j$,

$$g_{ji} = \left[S_w\left(\omega_0\right)\left([\mathbf{S}^{-1}\left(\omega_0\right)]_{ii} - [\widetilde{\mathbf{S}}_j^{-1}\left(\omega_0\right)]_{ii}\right)\right]^{1/2}. \tag{3.3.3}$$

For $i > j$

$$g_{ji} = \left[S_w\left(\omega_0\right)\left([\mathbf{S}^{-1}\left(\omega_0\right)]_{ii} - [\widetilde{\mathbf{S}}_j^{-1}\left(\omega_0\right)]_{i-1,i-1}\right)\right]^{1/2}. \tag{3.3.4}$$

Proof. Without loss of generality, we consider the case that $j = N$ (for any other $j \neq N$, we can transform the problem to the case $j = N$ via a simple reordering of rows and columns). Subtracting the diagonal elements of $\mathbf{S}^{-1}(\omega)$ in (3.3.2) from those of $\widetilde{\mathbf{S}}_j^{-1}(\omega)$ in (3.2.1), we obtain

$$[\mathbf{S}^{-1}(\omega)]_{ii} - [\widetilde{\mathbf{S}}_j^{-1}(\omega)]_{ii} = \frac{[\mathbf{G}^\top \mathbf{G}]_{ii} - [\widetilde{\mathbf{G}}_N^\top \widetilde{\mathbf{G}}_N]_{ii}}{S_w(\omega)}.$$

Also, since $[\mathbf{G}^\top \mathbf{G}]_{ii} = \sum_k g_{ki}^2$ and $[\widetilde{\mathbf{G}}_N^\top \widetilde{\mathbf{G}}_N]_{ii} = \sum_{k\neq N} g_{ki}^2$, we have that

$$[\mathbf{G}^\top \mathbf{G}]_{ii} - [\widetilde{\mathbf{G}}_N^\top \widetilde{\mathbf{G}}_N]_{ii} = g_{Ni}^2,$$

for any $i < N$. The same analysis holds for $j \neq N$. Hence, we can recover the entries g_{ji}, for $i < j$, as stated in our theorem. Notice also that, for $j \neq N$ and $i > j$, we must use the entry

$[\widetilde{\mathbf{S}}_j^{-1}(\omega)]_{i-1,i-1}$ in (3.3.4), to take into account that $\widetilde{\mathbf{S}}_j(\omega)$ is an $(N-1)\times(N-1)$ matrix associated to the dynamics grounded at node v_j. □

3.3.1 Boolean Reconstruction of Directed Networks

Theorem 3.1 allows us to reconstruct the Boolean structure of an unknown directed network if we have access to the matrices of CPSD's, $\mathbf{S}(\omega_0)$ and $\widetilde{\mathbf{S}}_j(\omega_0)$, for any ω_0 in the excitation frequency interval $(-\Omega, \Omega)$. In particular, one can verify the existence of a directed edge (i, j) by checking the condition $g_{ji} > 0$, where g_{ji} is computed from Theorem 3.1. In practice, the CPSD's $\mathbf{S}(\omega_0)$ and $\widetilde{\mathbf{S}}_j(\omega_0)$ are empirically computed from the stochastic outputs of the network, $\mathbf{y}(t)$ and $\widetilde{\mathbf{y}}(t)$; therefore, they are subject to numerical errors. Hence, in the implementation, one should relax the condition $g_{ji} > 0$ to $g_{ji} > \tau$, where τ is a small threshold used to account for numerical precision.

Based on Theorem 3.1, we propose Algorithm 1 to find the Boolean representation of $\mathbf{G}$, denoted by $\mathbf{B}(\mathbf{G})$, when a directed dynamical network is excited by an input noise of unknown PSD.

Algorithm 1 incurs the following computational cost:

(i) It computes the cross-correlation functions for all the N^2 pairs of outputs in (3.1.2). For each one of the N grounded dynamics in (3.3.1), the algorithm also computes $(N-1)^2$ pairs of cross-correlation functions, resulting in a total of $\mathcal{O}(N^3)$ computations. To compute these cross-correlations we use time series of length L. Since each each cross-correlation takes $\mathcal{O}\left(L^2\right)$ operations, we have a total of $\mathcal{O}\left(N^3L^2\right)$ operations to compute all the required cross-correlations.

(ii) Algorithm 1 evaluates the DFT of all $(N+1)\,N^2$ cross-correlation functions of length L in *(i)* at a particular frequency $\omega_0 \in (-\Omega, \Omega)$. Since evaluating the DFT at a single frequency takes

Algorithm 1 Boolean reconstruction of directed networks

```
Require: $h(\mathbf{j}\omega)$, $\mathbf{y}(t)$ from (3.1.2), $\widetilde{\mathbf{y}}(t)$ from (3.3.1), and any $\omega_0 \in (-\Omega, \Omega)$;
Compute $\mathbf{S}(\omega_0)$ from $\mathbf{y}(\mathbf{t})$;
for $j = 1 : N$ do
    Compute $\widetilde{\mathbf{S}}_j(\omega_0)$ from $\widetilde{\mathbf{y}}(t)$;
    for $i = 1 : j - 1$ do
        if $[\mathbf{S}^{-1}(\omega_0)]_{ii} - [\widetilde{\mathbf{S}}_j^{-1}(\omega_0)]_{ii} > \tau$ then $b_{ji} = 1$;
        if $[\mathbf{S}^{-1}(\omega_0)]_{ii} - [\widetilde{\mathbf{S}}_j^{-1}(\omega_0)]_{ii} < \tau$ then $b_{ji} = 0$;
    end for
    for $i = j + 1 : N$ do
        if $[\mathbf{S}^{-1}(\omega_0)]_{ii} - [\widetilde{\mathbf{S}}_j^{-1}(\omega_0)]_{i-,1i-1} > \tau$ then $b_{ji} = 1$;
        if $[\mathbf{S}^{-1}(\omega_0)]_{ii} - [\widetilde{\mathbf{S}}_j^{-1}(\omega_0)]_{i-1,i-1} < \tau$ then $b_{ji} = 0$;
    end for
end for
```

$\mathcal{O}(L)$ operations, we have a total of $\mathcal{O}\left(N^3 L\right)$ operations to compute the CPSD's matrices $\mathbf{S}\left(\omega_0\right)$ and $\widetilde{\mathbf{S}}_j\left(\omega_0\right)$, for all $j = 1, \ldots, N$.

(iii) Our algorithm also needs to compute the inverse of $\mathbf{S}\left(\omega\right)$ and $\widetilde{\mathbf{S}}_j\left(\omega\right)$. Since each inversion takes $\mathcal{O}\left(N^3\right)$, we have a total of $\mathcal{O}\left(N^4\right)$ operations to compute the inverses of all the $N+1$ matrices involved in our computations.

Therefore, the total computational cost of our algorithm is $\mathcal{O}\left(N^4 + N^3 L^2\right)$. In the next subsection, we extend Algorithm 1 to reconstruct the exact connectivity matrix $\mathbf{G}$.

3.3.2 Exact Reconstruction of Directed Networks

Apart from a Boolean reconstruction of $\mathbf{G}$, we can also compute the weights of the edges in the network if we know one eigenvalue-eigenvector pair $(\lambda, \mathbf{u})$ of $\mathbf{G}$, as follows. This can be the case of $\mathbf{G}$ being, for example, a Laplacian matrix (since $\mathbf{G}\mathbb{1}_N = 0$, in this case), or the adjacency matrix of a d-regular graph (since $\mathbf{G}\mathbb{1}_N = d\mathbb{1}_N$). In these cases, we use Lemma 3.2.5 to find the value of $S_w\left(\omega_0\right)$ at a particular frequency $\omega_0 \in (-\Omega, \Omega)$. For example, in the case of $\mathbf{G}$ being a Laplacian, we have the following result:

Corollary 3.1. *Consider the networked dynamical system in* (3.1.2)*, when* $\mathbf{G} = -L_{\mathcal{D}}$*, where* $L_{\mathcal{G}}$ *is the Laplacian matrix of a directed graph* $\mathcal{D}$*. Then, under assumptions (A1)-(A2), the PSD of the input noise,* $S_w(\omega)$*, can be computed as*

$$S_w(\omega) = \frac{N}{(\mathbb{1}^\top \mathbf{S}^{-1}\left(\omega\right) \mathbb{1})|h\left(\mathbf{j}\omega\right)|^2}.$$

Proof. This result can be directly obtained from Lemma 3.3 taking into account that the eigenpair $(\lambda, \boldsymbol{u})$ for the Laplacian matrix is $(0, \mathbf{1}_N)$. □

In general, we can reconstruct the weights of directed edges in a dynamical network using Algorithm 2.

Remark 3.2. It is worth remarking that the proposed reconstruction methods do not require the entire power spectra for $\mathbf{S}\left(\omega\right)$ or $S_w\left(\omega\right)$, but only the values of these spectral densities at any frequency $\omega_0 \in (-\Omega, \Omega)$. This dramatically reduces the computational complexity of the reconstruction.

We now turn to two particular types of networks, namely, undirected and nonreciprocal networks, in which the computational cost of reconstruction can be drastically reduced.

3.3.3 Exact Reconstruction of Undirected Networks

Consider Problem 3.1, when the connectivity matrix $\mathbf{G}$ is an unknown (possibly weighted) symmetric matrix. Then, when an eigenpair $(\lambda, \boldsymbol{u})$ is known, we can find the exact structure of the

Algorithm 2 Exact reconstruction of directed networks

```
Require: $h(\mathbf{j}\omega)$, $\mathbf{y}(t)$ from (3.1.2), $\widetilde{\mathbf{y}}(t)$ from (3.3.1), and any $\omega_0 \in (-\Omega, \Omega)$;
Compute $\mathbf{S}(\omega_0)$ from $\mathbf{y}(\mathbf{t})$ and $S_w(\omega_0)$ using (3.2.5);
for $j = 1 : N$ do
    Compute $\widetilde{\mathbf{S}}_j(\omega_0)$ from $\widetilde{\mathbf{y}}(t)$;
    for $i = 1 : j-1$ do
        $g_{ji} = \left[S_w\left(\omega_0\right)\left([\mathbf{S}^{-1}\left(\omega_0\right)]_{ii} - [\widetilde{\mathbf{S}}_j^{-1}\left(\omega_0\right)]_{ii}\right)\right]^{1/2}$;
    end for
    for $i = j+1 : N$ do
        $g_{ji} = \left[S_w\left(\omega_0\right)\left([\mathbf{S}^{-1}\left(\omega_0\right)]_{ii} - [\widetilde{\mathbf{S}}_j^{-1}\left(\omega_0\right)]_{i-1,i-1}\right)\right]^{1/2}$;
    end for
end for
```

network from the matrix of CPSD's, $\mathbf{S}\left(\omega\right) = \left[S_{y_i y_j}\left(\omega\right)\right]_{1\leq i,j\leq N}$, and the nodal transfer function, $h\left(\mathbf{j}\omega\right) = c^\top\left(\mathbf{j}\omega I_n - A\right)^{-1} b$, using the following result:

Theorem 3.3. *Consider the networked dynamical system* (3.1.2)*, when* $\mathbf{G} = \mathbf{G}^\top$*. Then, under assumptions (A1)-(A2), we have that*

$$\mathbf{G} = \left(\mathbf{S}^{-1}\left(\omega_0\right) S_w(\omega_0) - \mathit{Im}^2\left\{h^{-1}\left(\mathbf{j}\omega_0\right)\right\} I_N\right)^{1/2} \\ + \mathit{Re}\left\{h^{-1}\left(\mathbf{j}\omega_0\right)\right\} I_N. \tag{3.3.5}$$

for any $\omega_0 \in (-\Omega, \Omega)$.

Proof. From Lemma 3.2, we obtain the following for $\mathbf{G}^\top = \mathbf{G}$:

$$\begin{aligned}\mathbf{S}^{-1}\left(\omega\right) S_w(\omega) &= \frac{I_N}{\left|h\left(\mathbf{j}\omega\right)\right|^2} + \mathbf{G}^2 - \frac{\mathbf{G}}{h^*\left(\mathbf{j}\omega\right)} - \frac{\mathbf{G}}{h\left(\mathbf{j}\omega\right)} \\ &= \mathbf{G}^2 - 2\mathrm{Re}\{h^{-1}\left(\mathbf{j}\omega\right)\}\mathbf{G} \\ &+ I_N\left(\mathrm{Im}^2\{h^{-1}\left(\mathbf{j}\omega\right)\} + \mathrm{Re}^2\{h^{-1}\left(\mathbf{j}\omega\right)\}\right) \\ &= \left(\mathbf{G} - \mathrm{Re}\{h^{-1}\left(\mathbf{j}\omega\right)\} I_N\right)^2 + \mathrm{Im}^2\{h^{-1}\left(\mathbf{j}\omega\right)\} I_N,\end{aligned}$$

thereby completing the proof. □

Based on Theorem 3.3, we can reconstruct the connectivity matrix $\mathbf{G} = \mathbf{G}^\top$ when we know an eigenpair of $\mathbf{G}$. The input PSD in (3.3.5) can be computed using Lemma 3.3. Notice that this algorithm does not require grounding the dynamics of the network, resulting in a reduced computational cost. In particular, the computational cost is dominated by the computation of $\mathbf{S}\left(\omega_0\right)$, which requires $\mathcal{O}\left(N^2L^2\right)$ operations, and its inversion, which requires $\mathcal{O}\left(N^3\right)$, resulting in a total cost of $\mathcal{O}\left(N^2L^2 + N^3\right)$.

3.3.4 Reconstruction of Non-Reciprocal Networks

Another particular network structure that does not require grounding in the reconstruction method is the so-called nonreciprocal directed networks. In a *nonreciprocal network*, having an edge $(v_j, v_i) \in \mathcal{E}_d$ implies that $(v_i, v_j) \notin \mathcal{E}_d$. In other words, the connectivity matrix of a purely unidirectional network satisfies Tr$(\mathbf{G}^2) = \sum_i \sum_j g_{ij}g_{ji} = 0$, since, if $g_{ij} \neq 0$, then $g_{ij} = 0$ (and assuming there are no self-loops in the network).

The following theorem allows the Boolean reconstructing of a nonreciprocal network. Moreover, if we have access to an eigenpair of $\mathbf{G}$, this theorem could be used to perform an exact reconstruction without grounding the dynamics of the network.

Theorem 3.4. *Consider the networked dynamical system* (3.1.2)*, with a connectivity matrix satisfying* $\mathbf{G} \geq 0$ *(nonnegativity) and Tr*$(\mathbf{G}^2) = 0$ *(nonreciprocity). Then, under assumptions (A1)-(A2), we have that*

$$g_{ij} = \max\left\{ S_w(\omega)\left(\frac{[Im\{\mathbf{S}^{-1}(\omega)\}]_{ij}}{Im\{h^{-1}(\mathbf{j}\omega)\}}\right), 0\right\}, \tag{3.3.6}$$

for $1 \leq i \neq j \leq N$.

Proof. Under purview of Lemma 3.2, we obtain

$$\mathbf{S}^{-1}(\omega) S_w(\omega) = \frac{I_N}{|h(\mathbf{j}\omega)|^2} + \mathbf{G}^\top\mathbf{G} - \frac{\mathbf{G}}{h^*(\mathbf{j}\omega)} - \frac{\mathbf{G}^\top}{h(\mathbf{j}\omega)}.$$

Taking the imaginary parts, we obtain

$$\begin{aligned}\mathrm{Im}\{\mathbf{S}^{-1}(\omega) S_w(\omega)\} &= \mathrm{Im}\{-\frac{\mathbf{G}}{h^*(\mathbf{j}\omega)} - \frac{\mathbf{G}^\top}{h(\mathbf{j}\omega)}\} \\ &= \mathrm{Im}\{h^{-1}(\mathbf{j}\omega)\}(\mathbf{G} - \mathbf{G}^\top),\end{aligned}$$

which entails

$$\mathbf{G} - \mathbf{G}^\top = \frac{S_w(\omega)}{\mathrm{Im}\{h^{-1}(\mathbf{j}\omega)\}}\mathrm{Im}\{\mathbf{S}^{-1}(\omega)\}.$$

Given that $\mathbf{G} \geq 0$ and the network is nonreciprocal, if $\left[\mathbf{G} - \mathbf{G}^\top\right]_{ij} > 0$, then $g_{ij} > 0$ and $g_{ji} = 0$. If $\left[\mathbf{G} - \mathbf{G}^\top\right]_{ij} < 0$, then $g_{ij} = 0$ and $g_{ji} > 0$. Finally, if $\left[\mathbf{G} - \mathbf{G}^\top\right]_{ij} = 0$, then no directed edge between v_i and v_j exists. These three conditional statements can be condensed into (3.3.6). □

Using this theorem, we can find the the Boolean representation of $\mathbf{G}$, $\mathbf{B}(\mathbf{G}) = [b_{ij}]$, as follows,

$$b_{ij} = \begin{cases} 1, & \text{if } \frac{[\mathrm{Im}\{\mathbf{S}^{-1}(\omega_0)\}]_{ij}}{\mathrm{Im}\{h^{-1}(\mathbf{j}\omega_0)\}} > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $\omega_0 \in (-\Omega, \Omega)$. Moreover, if an eigenvalue eigenvector pair of $\mathbf{G}$ is known, we can recover $S_w(\omega_0)$ using Lemma 3.3, which allows us to recover the value of g_{ij} directly from 3.3.6. Following the analysis of previous algorithms, the computational cost of the reconstruction of a nonreciprocal directed network is $\mathcal{O}\left(N^2L^2 + N^3\right)$.

3.4 Related Literature

In the literature, we find a wide collection of approaches aiming to solve the network reconstruction problem. In the physics literature, we find in [59] a method to identify a network of dynamical systems which assumes that the input of each node can be individually manipulated. In [71], an approach based on Granger's causality [72] and the theory of reproducing kernel Hilbert spaces is proposed. In the statistics community, Bach and Jordan [73] used the Bayesian information criterion (BIC) to estimate sparse graphs from stationary time series. The optimization community has recently proposed a collection of papers aiming to find the sparsest network given a priori structural information [58, 60]. Although the assumption of sparsity is well justified in some applications, this assumptions might lead to unsuccessful topology inference, as illustrated in [74, 75]. Gonçalves et al. [74] investigate the necessary and sufficient conditions for reconstruction of LTI networks. Their work has been recently extended to reconstruction in the presence of intrinsic noise in [76]. On the other hand, for tree networks, several techniques for reconstruction are proposed in [61] and [77]. More recently, in a seminal work by Materassi and Salapaka [67], the authors propose a methodology for reconstruction of directed networks using locality properties of the Wiener filters. Although being applicable to many networks, this methodology is not exact when two nonadjacent nodes point towards a common node. In [64, 78, 79], several techniques are proposed to extract structural information of an undirected network running consensus dynamics. In particular, Nabi-Abdolyousefi et al. proposed in [64] a reconstruction technique based on a node-knockout procedure, where nodes are sequentially forced to broadcast a zero state (without being removed from the network).

Fazlyab and Preciado [80] propose an identification+control method over networks. In their approach, the unknown network is recovered using the combination of Lyapunov based adaptive feedback input and sliding mode control. In [81, 82], authors provide a sufficient condition that guarantees identifiability for a class of linear network dynamic systems exhibiting continuous-time weighted consensus protocols. Another interesting approach to network identification problem is distributed reconstruction addressed recently in [83, 84]. Finally, identification of subspaces has been studied in [85] for directed acyclic graphs.

Part II

Individual and Collaborative Online Learning

Chapter 4

Multi-Armed Bandits in Multi-Agent Networks

Online prediction, learning and decision making is a main topic of research in the theory of machine learning. A popular model for studying sequential decision problems is the multi-armed bandit (MAB) problem. Early studies on the problem dates back to 1933 when W. R. Thompson proposed the celebrated *Thompson Sampling* method. The problem has been extensively studied ever since, and many variants of it have been investigated in the literature [14–18]

Capturing the exploration-exploitation dilemma for a *learner*, MAB is defined by a set of *arms* or *actions*. At each time step, the learner chooses an arm and receives its corresponding *payoff* or *reward*. The objective is to maximize the total payoff obtained from sequentially selecting the arms. Equivalently, the learner aims to minimize regret when competing with the best single arm in hindsight. The reward model could be stochastic or non-stochastic, and optimal algorithms are proposed for both cases [15, 16]. While early studies on MAB dates back to nine decades ago, the problem has received considerable attention due to its modern applications. MAB could be an instance of sequential decision making for ad placement, website optimization or packet routing [18].

In this chapter we depart from the classical setting, and address the stochastic MAB in a multi-player network. Consider a group of sensors (players) that measure the location of a finite number of targets (arms). Each sensor contacts one target per time step, and can only measure a specified coordinate of its position. The target reveals a noisy version of the coordinate to the sensor, and the noise characteristics are different among sensors. They aim to track the closest target to the origin, and with one coordinate at hand, sensors must communicate with each other to supplement their imperfect observations. The problem is even harder when some targets are not responsive all the time. Motivated by this example, we propose two algorithms in Section 4.3, and apply them to the problem in Section 4.4.

4.1 Outline of the Problem and Results

The multi-player MAB is an instance of many problems where a group of *players* or *agents* collaborate to achieve a team task, say maximizing a *global* payoff. Players intend to reach consensus on an arm which best fits the network, i.e., the arm that maximizes the global reward. Naturally, each arm may reveal different rewards when chosen by distinct players. The goal is to compete with the arm that has the highest average reward among players. Alternatively, one can also think of the following scenario. Each arm has a *true* global payoff that can be written as an average of *individual* payoffs. Once a player pulls an arm, the adversary filters out the corresponding individual reward, and unveils a noisy version of that to the player. Agents are not able to compete with the best global arm unless they benefit from side observations gained from local communication. The model has a flavor of distributed algorithms where the parameter of interest is not fully observable to an individual learner [9, 30, 32]. However, it is in a bandit setup where the player only receives the payoff of a chosen action.

Pulling an arm, a player incurs an individual regret which is the difference between the payoff of the action and the best global arm. The network regret is then the average of individual regrets. We propose an algorithm named Distributed Upper Estimated Reward (d-UER) to minimize the network regret. The algorithm exploits a confidence bound that relies on the network topology and connectedness. We further extend the setting to *sleeping* MAB where some actions might be unavailable to players at each round. In this environment, the natural benchmark to compete with is the sequence of best available arms per round [86]. We develop the Distributed Awake Upper Estimated Reward (d-AUER) algorithm for sleeping bandit problem. Our algorithms are optimal in the sense that in a complete network they scale down the regret of their single-player counterpart by network size. We finally apply our methods to distributed detection of targets in sensor networks, and provide numerical experiments for our theoretical findings [87].

4.1.1 Related Literature

In recent years, many variants of MAB have been a major focus of research in several communities. In [88] a decentralized MAB has been formulated with applications in cognitive radio networks and multi-channel communication systems. In this model, simultaneous selection of one arm by a few players results in *zero* or *shared* reward. The authors in [89] propose a decentralized method for allocating multiple users to a set of wireless channels. Similarly, when multiple players use the same channel, the channel quality reduces due to interference. The work of [90] is also in the same spirit in which any collaboration among players is prohibited and adds to regret. The authors study the stochastic and rested Markovian reward model, and build on a distributed bipartite matching to introduce a new decentralized policy.

There is also an extensive literature focused on decentralized MAB problems with application in advertising systems. In the setting proposed in [91], the interaction between users in a social network provides information for an external decision maker. The decision maker benefits from the side observation to choose a content for each user. In [92] only a single *major* agent in the network has access to its reward sequence, while other agents are aware of the sampling pattern of the major agent. Comparing to the classical MAB, the asymptotic lower bound on regret scales down by the number of agents when the network is connected. On the other hand, the network model in [93, 94] encodes the connection between arms. That is, sampling an arm reveals *side information* on the reward of neighboring arms. The authors in [95] propose an algorithmic approach to networked contextual bandits, where the learner leverages side observations provided as a result of social relationships. Outside of network context, structured bandits is addressed in [96] where the reward of arms may depend on each other through a parameter.

Of particular relevance to the sleeping bandits is the work in [97] where a *graphical* MAB is introduced. In this setting the subset of available arms at any round is a function of the arm chosen in the previous round. The authors develop a block allocation algorithm for the problem that achieves a logarithmic regret. In [98] a combinatorial MAB problem is formulated where multiple arms can be selected once they respect a given constraint. The learner is rewarded with a linear combination of chosen arms, and the objective is to compete with the best linear combination. Finally, our work lies on the spectrum that covers a wide range from the classical MAB to distributed detection and learning. The works of [16–18, 86] form one side of the spectrum which is known as *one-player* MAB. On the other side, we can place distributed detection algorithms under *full information* setting [32, 40]. In these models, the world is governed by a fixed true *state* (arm), aimed to be recovered by a network of agents. Despite the local access to data, agents receive information about all states per round.

4.2 Notation and Problem Formulation

$[n]$	The set $\{1, 2, ..., n\}$ for any integer n
$x^\top$	Transpose of the vector x
$x(k)$	The k-th element of vector x
$\mathbf{1}\{\cdot\}$	The indicator function
$\mathbb{1}$	Vector of all ones
$\sigma_i(W)$	The i-th largest singular value of matrix W

Table 4.1: Notation

Consider a multi-agent network where N *players* or *agents* sequentially select *arms* or *actions*. The set of arms is of size K which is a common knowledge among players. Pulling arm $k \in [K]$ at time $t \in [T]$ yields a *reward* $X_{i,t}(k) \in [0, 1]$ for player $i \in [N]$. We study a stochastic model of rewards where $\mu_i = \mathbb{E}[X_{i,t}] \in \mathbb{R}^K$ is a fixed vector over time horizon. Also, the average reward of each arm might be different among players, i.e., for any $k \in [K]$ and $i \neq j$, $\mu_i(k)$ is not necessarily equal to $\mu_j(k)$. Therefore, a "good" arm for a player might be a "bad" arm for another one. The rewards are independent and identically distributed over time, while they are also independent across players and arms. The random variable $I_{i,t}$ represents the action of player i at time t, and the player only observes the corresponding reward $X_{i,t}(I_{i,t})$ at that period.

4.2.1 Standard Setting

In the classical framework, agents want to maximize an average *global* welfare. That is, the players' objective is to identify the most rewarding arm k^*,

$$k^* := \text{argmax}_{k\in[K]} \left\{ \mu(k) := \frac{1}{N}\sum_{i=1}^{N} \mu_i(k) \right\},$$

which best suits the whole network. Without loss of generality, we assume the following order

$$\mu(1) \geq \mu(2) \geq \cdots \geq \mu(K), \tag{4.2.1}$$

to simplify the exposition of our results. For any pair $k \leq m$, we define

$$\Delta_{k,m} := \mu(k) - \mu(m) = \frac{1}{N}\sum_{i=1}^{N} \mu_i(k) - \mu_i(m),$$

to capture the suboptimality of arm m comparing to arm k. Let $n_{i,t}(k)$ denote the number of times that arm k has been chosen by player i until time t. Then, players aim to minimize the regret in the following sense,

$$\mathbf{R}_T := T\mu(1) - \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E}\left[\mu(I_{i,t})\right] = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=2}^{K} \Delta_{1,k}\mathbb{E}[n_{i,T}(k)], \tag{4.2.2}$$

where the expectation is taken over the randomness in the choice of arms.

4.2.2 Sleeping Bandits

In the *sleeping* MAB problem, not every arm is awake all the time. At time $t \in [T]$, there exists a specific set of arms $A_{i,t} \subseteq [K]$ available to player i, and the player cannot choose some action $k \notin A_{i,t}$. The dependence of $A_{i,t}$ to i reiterates that at any time $t \in [T]$, an available arm to a player might be unaccessible to another player. In this scenario, it is reasonable to compete with the sequence of best *available* arms. Let $k^*_{i,t}$:=argmax$_{k\in A_{i,t}}\{\mu(k)\}$ be the best available arm to player

i at round t, and for any $m > k$, $n_{i,t}\left(m \mid k\right)$ denote the number of times that agent i has played the suboptimal arm m until time t given that some better arm in the set $[k]$ has been available. We now define the regret with respect to the described benchmark as follows

$$\begin{aligned}\overline{\mathbf{R}}_T &:= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu(k_{i,t}^{*}) - \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left[\mu(I_{i,t})\right]\\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{m=2}^{K}\sum_{k=1}^{m-1}\Delta_{k,m}\mathbb{E}\left[n_{i,T}\left(m \mid k\right) - n_{i,T}\left(m \mid k-1\right)\right]\\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{m=2}^{K}\sum_{k=1}^{m-1}\left(\Delta_{k,m} - \Delta_{k+1,m}\right)\mathbb{E}\left[n_{i,T}\left(m \mid k\right)\right], \qquad (4.2.3)\end{aligned}$$

where the last step follows from rearranging the terms and the convention that $\Delta_{m,m} = 0$ and $n_{i,T}\left(m \mid 0\right) = 0$. Given the order of arms (4.2.1), a player regrets over pulling arm m only if some arm $k < m$ is awake when making the decision.

4.2.3 Network Structure

A player cannot track the best arms in isolation as the best "global" arm might be "suboptimal" for the player. Therefore, players need to exchange information with each other at every round. We let the symmetric and doubly stochastic matrix W encode the interaction structure among agents. The matrix has positive diagonals, and any positive entry $[W]_{ij} > 0$ implies that player $i \in [N]$ assigns a weight $[W]_{ij} = [W]_{ji}$ to observations of player $j \in [N]$. Of course, when $[W]_{ij} = 0$, agents i and j never communicate with each other directly. Therefore,

$$\begin{aligned}\text{for all } i \in [N] : \sum_{j \in \mathcal{N}_i}[W]_{ij} &= \sum_{j=1}^{N}[W]_{ij} = 1\\ \text{for all } j \in [N] : \sum_{i \in \mathcal{N}_j}^{N}[W]_{ij} &= \sum_{i=1}^{N}[W]_{ij} = 1,\end{aligned}$$

where $\mathcal{N}_i := \{j \in [N] : [W]_{ij} > 0\}$ is the *local* neighborhood of agent i. We assume that the underlying network is *connected*, i.e., there exists a *path* from any player $i \in [N]$ to any player

$j \in [N]$. Intuitively, the assumption guarantees the information flow over the network.

We now state a few properties of the depicted network model, and refer the interested reader to [99] for a complete survey on stochastic matrices. It follows from doubly stochasticity of W that the largest singular value is $\sigma_1(W) = 1$. Furthermore, since W has positive diagonal and the topology is connected, the Markov chain W is irreducible and aperiodic. As a consequence, the largest singular value is unique, and it holds that $\sigma_2(W) < 1$. Also, the stationary distribution of the chain is unique and $W^t \to \frac{1}{N}\mathbb{1}\mathbb{1}^\top$, as $t \to \infty$. Finally, without loss of generality, we assume that N is large enough $(N > 8)$ to simplify our regret bounds.

4.3 Algorithms

We now present our technical results and their consequences. We first describe the d-UER algorithm for the case of all-awake arms, and then we propose d-AUER to deal with sleeping bandits setting. Our algorithms are optimal up to constant factors in that removing network error recovers the result for one player MAB in both settings. Omitted proofs are included in the supplementary material.

4.3.1 The d-UER Algorithm

In this section, we delineate the d-UER algorithm to examine the case where every arm is available at any time. The algorithm can be cast as a distributed variant of the celebrated UCB1 [16]. As we discussed in the Preliminaries 4.2, the feedback setup does not allow a single player to compete solely with the best arm. Therefore, players need to communicate to collectively explore the arms. While taking into account an upper confidence bound, each player aggregates observations in her local neighborhood to make decision as follows:

Unlike the UCB1 algorithm, d-UER exploits a confidence bound that depends on parameter d.

Algorithm 3 `Distributed Upper Estimated Reward`

```
Input : The parameters $d$ and $N$.
Initialization :
  Each action is played once, and the rewards are stored in vector $X_{i,0}$ for all $i \in [N]$.
  For each $i \in [N]$ and $k \in [K]$, let $\phi_{i,0}(k) = 0$, $n_{i,0}(k) = 1$ and $\psi_{i,1}(k) = X_{i,0}(k)$.
for $t = 1$ to $T$ do
  for $i = 1$ to $N$ do
    Calculate the vector $\phi_{i,t} = \sum_{j=1}^{N} [W]_{ij} \phi_{j,t-1} + \psi_{i,t}$.
    Select $I_{i,t} = \text{argmax}_{k \in [K]} \left\{ \frac{1}{n_{i,t-1}(k)} \phi_{i,t}(k) + \sqrt{2 \log t \left( \frac{1}{N n_{i,t-1}(k)} + \frac{2d}{n_{i,t-1}^2(k)} \right)} \right\}$.
    Update the counter as $n_{i,t}(k) = n_{i,t-1}(k) + \mathbf{1}\{k = I_{i,t}\}$ for any $k \in [K]$.
    Score $\mu(I_{i,t})$, observe $X_{i,t}(I_{i,t})$ and let $\psi_{i,t+1}(k) = X_{i,t}(k) \mathbf{1}\{k = I_{i,t}\}$ for any $k \in [K]$.
  end for
end for
```

We shall see that this parameter must be tuned as an upper bound on a quantity that depends on network characteristics. We state the following lemma which provides a closed-from solution for $\{\phi_{i,t}\}_{t=1}^{T}$, and sheds light on the mixture behavior of Markov chain W.

Lemma 4.1. *Any update of the form $\phi_{i,t} = \sum_{j=1}^{N} [W]_{ij} \phi_{j,t-1} + \psi_{i,t}$ can be expressed as,*

$$\phi_{i,t} = \sum_{\tau=1}^{t} \sum_{j=1}^{n} \left[W^{t-\tau}\right]_{ij} \psi_{j,\tau},$$

whenever the update is initialized at $\phi_{i,0}(k) = 0$, for any $i \in [N]$ and $k \in [K]$. Also, given strong connectivity of the network, the doubly stochastic matrix W with positive diagonal satisfies

$$\sum_{\tau=1}^{t} \sum_{j=1}^{N} \left| \left[W^{t-\tau}\right]_{ij} - \frac{1}{N} \right| \leq \frac{2}{1 - \sigma_2(W)} + \frac{\log N}{\log\left[\sigma_2(W)^{-1}\right]},$$

for any $i \in [N]$, where $\sigma_2(W) < 1$ is the second largest singular value of W.

The lemma suggests that the update simultaneously admits new information and averages out the past. The connectivity of the network plays an important role in decision making, since it allows

$W^t \to \frac{1}{N}\mathbb{1}\mathbb{1}^\top$ as $t \to \infty$. Indeed, when the underlying topology is disconnected, information cannot propagate through the whole network. On the other hand, the lemma suggests that the regret relies on how fast the Markov chain W mixes. This is captured by dependence of the RHS of above to $\sigma_2(W)$.

Theorem 4.1. *The regret of* `d-UER` *algorithm, defined in* (4.2.2)*, satisfies the following bound*

$$\mathbf{R}_T \leq \sum_{k=2}^{K} \left\{ 4 \max\left\{ \frac{12 \log T}{N \Delta_{1,k}}, Nd \right\} + 2.5 \left(1 + \log\left[\frac{4}{\Delta_{1,k}}\right]\right) \boldsymbol{dE}_1 \boldsymbol{dE}_2 + \frac{2\pi^2}{3}\Delta_{1,k} \right\},$$

whenever $d \geq \boldsymbol{dE}_1$, where

$$\boldsymbol{dE}_1 := \frac{2}{1-\sigma_2(W)} + \frac{\log N}{\log\left[\sigma_2(W)^{-1}\right]}, \quad \text{and} \quad \boldsymbol{dE}_2 := \frac{\log N}{\log\left[\sigma_2(W)^{-1}\right]}.$$

Theorem 4.1 indicates that the regret depends on the network size and second largest singular value of W. The local feedback does not provide each player with adequate information, yielding a delay in proper decision making. For instance, in cycle and path networks where the diameter is $\mathcal{O}(N)$ the incurred penalty $\boldsymbol{dE}_1 = \tilde{\mathcal{O}}(N^2)$ is large, whereas in a complete network $W = \frac{1}{N}\mathbb{1}\mathbb{1}^\top$ the Markov chain is mixed from the outset. The scenario can be seen as N copies of a single-player MAB where $\sigma_2(W) = 0$. In this case, the network errors become $\boldsymbol{dE}_1 = 2$ and $\boldsymbol{dE}_2 = 0$, and the well-known result of [16] for `UCB1` algorithm is recovered (scaled down by a factor of N). This advantage is gained through reducing the variance of samples by distributing N samples among N individuals. Recall that `UCB1` algorithm is optimal in the sense that a lower bound is available under mild assumptions on reward distributions (see e.g. [14, 18]).

4.3.2 The `d-AUER` Algorithm

We now extend the results to sleeping bandits where some arms might be unavailable at every round. The single-player version of the problem has been addressed in [86]. Here, agents also suffer from an insufficient feedback which provides only local information. Naturally, players compete with the sequence of best awake arms as in Algorithm 4. Again note that unlike the single-player version in [86], the estimator and confidence bound rely on network structure.

Algorithm 4 `Distributed Awake Upper Estimated Reward`

```
Input : The parameters $d$ and $N$.
Initialization :
  For each $i \in [N]$ and $k \in [K]$, let $\phi_{i,0}(k) = 0$, $n_{i,0}(k) = 0$ and $\psi_{i,1}(k) = 0$.
for $t = 1$ to $T$ do
  for $i = 1$ to $N$ do
    Calculate the vector $\phi_{i,t} = \sum_{j=1}^N [W]_{ij}\phi_{j,t-1} + \psi_{i,t}$.
    if $\exists k \in A_{i,t}$ such that $n_{i,t-1}(k) = 0$ then
      Choose the action $I_{i,t} = k$.
    else
      Select $I_{i,t} = \text{argmax}_{k \in A_{i,t}} \left\{ \frac{1}{n_{i,t-1}(k)}\phi_{i,t}(k) + \sqrt{2\log t \left( \frac{1}{N n_{i,t-1}(k)} + \frac{2d}{n_{i,t-1}^2(k)} \right)} \right\}$.
    end if
    Update the counter as $n_{i,t}(k) = n_{i,t-1}(k) + \mathbf{1}\{k = I_{i,t}\}$ for any $k \in [K]$.
    Score $\mu(I_{i,t})$, observe $X_{i,t}(I_{i,t})$ and let $\psi_{i,t+1}(k) = X_{i,t}(k)\mathbf{1}\{k = I_{i,t}\}$ for any $k \in [K]$.
  end for
end for
```

Lemma 4.2. *For any sequence of nonnegative real numbers $\{a_k\}_{k=1}^m$, we have*

$$\sum_{k=1}^{m-1} \frac{a_k}{\left(\sum_{s=k}^m a_s\right)^2} \leq \frac{1}{a_m}, \tag{4.3.1}$$

as long as $a_m > 0$. In particular, letting $a_k = \Delta_{k,k+1}$, we have

$$\sum_{k=1}^{m-1} \frac{\Delta_{k,k+1}}{\Delta_{k,m}^2} \leq \frac{2}{\Delta_{m-1,m}}. \tag{4.3.2}$$

The inequality (4.3.2) plays a key role in bounding the regret of sleeping bandit problem. We remark that in [86] the authors derive the same inequality as a corollary of a lemma which involves a complicated proof. While the result of [86] is also valuable for the case that the difference between arms is small, we provided an easy alternative to derive (4.3.2). Let us now present the main result of this section.

Theorem 4.2. *The regret of* d-AUER *algorithm, defined in* (4.2.3)*, satisfies the following bound*

$$\overline{\mathbf{R}}_T \leq \sum_{k=2}^{K} \left\{ \frac{\frac{96}{N}\log T + 8Nd + 30k\, \boldsymbol{dE}_1\, \boldsymbol{dE}_2}{\Delta_{k-1,k}} + \frac{2k\pi^2}{3}\Delta_{1,k} \right\},$$

so long as $\Delta_{k-1,k} > 0$ *and* $d \geq \boldsymbol{dE}_1$*, for* $k > 1$.

Theorem 4.2 articulates the relation of regret and network errors in the sleeping bandit model. Interestingly, we observe that $W = \frac{1}{N}\mathbb{1}\mathbb{1}^\top$ (which results in $\boldsymbol{dE}_1 = 2$ and $\boldsymbol{dE}_2 = 0$) recovers the regret bound of AUER algorithm [86] for single-player case (scaled down by a factor of N). Similar to Theorem 4.1 the result interpolates between well-connected and poorly connected networks using $\boldsymbol{dE}_1$ and $\boldsymbol{dE}_2$. Notice that one can simply relax the condition $\Delta_{k-1,k} > 0$ in Theorem 4.2 as follows. For an arbitrary choice of $\varepsilon \geq 0$, separate out any arm $k > 1$ such that $\Delta_{k-1,k} \leq \varepsilon$. Then, in view of (4.2.3), the regret bound in the theorem can be modified to

$$\overline{\mathbf{R}}_T \leq \mathcal{O}(\varepsilon T) + \sum_{k=2}^{K} \left\{ \frac{\frac{96}{N}\log T + 8Nd + 30k\, \boldsymbol{dE}_1\, \boldsymbol{dE}_2}{\Delta_{k-1,k}} \mathbf{1}\left\{\Delta_{k-1,k} > \varepsilon\right\} + \frac{2k\pi^2}{3}\Delta_{1,k} \right\}.$$

4.4 Application : Detection of the Closest Target in Sensor Networks

4.4.1 Sensing Model

We now present the application of our methods to distributed detection in sensor networks [100]. Consider a strongly connected network of N sensors that respects a fixed topology. The sensors (players) sequentially measure the location of K targets (arms) that live in $\mathbb{R}^2$ space. At any time $t \in [T]$, each sensor can contact one target to query the location, and the target discloses a noisy version of its position. The sensors broadcast the noisy data over the network to detect the farthest (or equivalently the closest) target to the origin.

We partition the set of sensors to two sets $\mathcal{X}$ and $\mathcal{Y}$ of the same size. Sensors in $\mathcal{X}$ measure the x-coordinate, while the other half in $\mathcal{Y}$ measure the y-coordinate. For any sensor $i \in [N]$ at time $t \in [T]$, let random variables $\theta_{i,t}$ and $r_{i,t}$ be drawn independently from uniform distribution with supports $[-\pi, \pi]$ and $[0, \overline{r}_i]$, respectively. Then, the location of target $k \in [K]$ from the standpoint of sensor $i \in \mathcal{X}$ takes the following form

$$x_{i,t}(k) = p(k) + r_{i,t}\cos(\theta_{i,t}) \qquad\qquad y_{i,t}(k) = r_{i,t}\sin(\theta_{i,t}), \tag{4.4.1}$$

whereas sensor $j \in \mathcal{Y}$ observes

$$x_{j,t}(k) = r_{j,t}\cos(\theta_{j,t}) \qquad\qquad y_{j,t}(k) = q(k) + r_{j,t}\sin(\theta_{j,t}). \tag{4.4.2}$$

Therefore, any sensor $i \in [N]$ measures a (wrong) distance of target k as

$$\boldsymbol{dist}^2_{i,t}(k) = x^2_{i,t}(k) + y^2_{i,t}(k), \tag{4.4.3}$$

and report it to other sensors in its local neighborhood. Calculating the expected squared-distance, we obtain

$$\mathbb{E}\left[\boldsymbol{dist}^2_{i,t}(k)\right] = p^2(k)\mathbf{1}\{i \in \mathcal{X}\} + q^2(k)\mathbf{1}\{i \in \mathcal{Y}\} + \frac{\overline{r}_i^2}{3}.$$

One can observe that each sensor has a different perception about the expected distance of target k from the origin. Therefore, they cannot identify the farthest target on their own. However, for any $k \in [K]$ it holds that

$$d^2(k) := \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\boldsymbol{dist}^2_{i,t}(k)\right] = \frac{1}{2}p^2(k) + \frac{1}{2}q^2(k) + \frac{1}{N}\sum_{i=1}^{N}\frac{\overline{r}_i^2}{3},$$

which allows sensors to correctly distinguish the farthest target, since the maximizers

$$\text{argmax}_{k\in[K]}\left\{d^2(k)\right\} = \text{argmax}_{k\in[K]}\left\{p^2(k) + q^2(k)\right\},$$

coincide. In this example, sleeping bandits corresponds to when some targets are not responsive, and sensors cannot obtain measurements from them. Therefore, we have the following corollary of Theorems 4.1 and 4.2.

Corollary 4.1. *Assume the sensing model given in* (4.4.1) *and* (4.4.2) *in the sensor network, and let sensors observe the feedback* (4.4.3) *at time* $t \in [T]$. *Then, the* d-UER *algorithm enjoys the regret bound*

$$\mathbf{R}_T \leq \mathcal{O}\left\{\sum_{k=2}^{K}\left\{\max\left\{\frac{24\log T}{Nu_{1,k}}, Nd\right\} + \log\left[\frac{8}{u_{1,k}}\right]\boldsymbol{dE}_1\boldsymbol{dE}_2 + u_{1,k}\right\}\right\},$$

where $u_{k,m} := p^2(k) + q^2(k) - p^2(m) - q^2(m)$ *for any* $k < m$. *Moreover, the* d-AUER *algorithm satisfies the regret bound*

$$\overline{\mathbf{R}}_T \leq \mathcal{O}\left\{\sum_{k=2}^{K}\frac{\frac{48}{N}\log T + 4Nd + 15k\,\boldsymbol{dE}_1\,\boldsymbol{dE}_2}{u_{k-1,k}} + ku_{1,k}\right\}.$$

4.4.2 Numerical Experiments

We now illustrate our approach via simulation of the described sensor network. Let $N = 30$ and $K = 4$ be the number of sensors and targets, respectively. For any target $k \in [K]$, the true coordinates $p(k)$ and $q(k)$ are drawn independently from a uniform distribution on the unit interval. Also, we let $\overline{r}_\ell = 0.1 + 0.02\ell$ for any $\ell \in [N]$ to discriminate between sensors with respect to noise radius. In our experiment, the minimum gap is $\min_{k\in[K]}\{\Delta_{1,k}\} \approx 0.2$.

We would like to evaluate the performance of d-UER algorithm in three networks : complete, cycle and 4-regular (all with self-loops). Using the sensing model in the previous section for each network, we average out 50 experiment runs to plot Fig. 4.1. As verified in theoretical results, the regret bound scales inversely with $1 - \sigma_2(W)$, called the spectral gap. We can observe the impact in Fig. 4.1 where the networks are sorted correctly with respect to this metric. The complete network (largest spectral gap) has the best performance, while the 4-regular outperforms the cycle (due to its larger spectral gap). We can see that the spectral gap is roughly an indicator of the network connectivity.

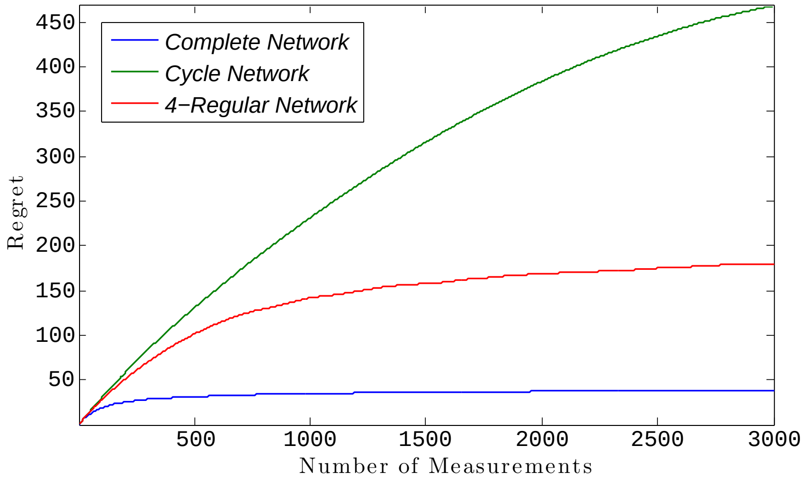

Figure 4.1: Performance of d-UER in complete, cycle and 4-regular networks.

We next turn to focus on importance of communication in detection. Each sensor shall not be able to find the closest target based on its own observations. In other words, agents might contact a wrong target in the order of measurement numbers, resulting in a linear regret. We investigate the phenomenon using the same procedure with $N = 8$ sensors. To this end, we compare a disconnected network versus a complete network in Fig. 4.2. In the disconnected network sensors are not able to distinguish the closest target, and the regret grows linearly in time.

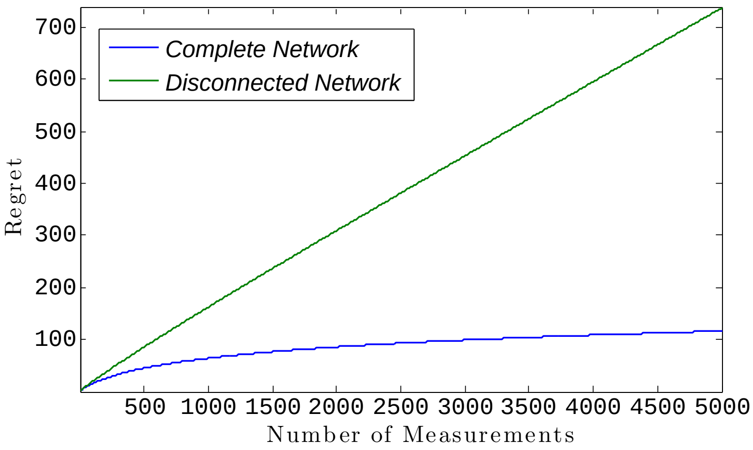

Figure 4.2: Sensors fail to detect the right target in a disconnected network, yielding a linear regret.

4.5 Proofs

Note : As mentioned in Section 4.2.3, for the proofs we sometimes assume that $N > 8$ to simplify the bounds. This assumption is made with no loss of generality, and only avoids notational clutter.

Proof of Lemma 4.1. The proof of the first part is standard (see e.g. Lemma 1 in [32]). For the second part, we follow the lines in the proof of Lemma 2 in [32]. Let $\mathbf{e}_i$ be the i-th unit vector in the standard basis of $\mathbb{R}^N$. The Markov chain W is irreducible and aperiodic, so by standard properties of stochastic matrices (see e.g. [46]), we have

$$\left\| \mathbf{e}_i^\top W^t - \frac{1}{N}\mathbb{1}^\top \right\|_1 \le \sqrt{N}\sigma_2(W)^t, \tag{4.5.1}$$

for any $i \in [N]$, as $\frac{1}{N}\mathbb{1}^\top$ is the stationary distribution of the transition kernel W. Hence,

$$\sqrt{N}\sigma_2(W)^{t-\tau} \le 2 \qquad \text{for} \qquad t - \tau \ge \tilde{t} := \frac{\log\left[\frac{\sqrt{N}}{2}\right]}{\log\left[\sigma_2(W)^{-1}\right]},$$

and recall that the inequality $\left\|\mathbf{e}_i^\top W^{t-\tau} - \frac{1}{N}\mathbb{1}^\top\right\|_1 \le 2$ always holds since any power of W is doubly stochastic. With that in mind, we use (4.5.1) to break the following sum into two parts to get

$$
\begin{aligned}
\sum_{\tau=1}^{t}\sum_{j=1}^{N}\left|\left[W^{t-\tau}\right]_{ij}-\frac{1}{N}\right| &= \sum_{\tau=1}^{t}\left\|\mathbf{e}_i^\top W^{t-\tau}-\frac{1}{N}\mathbb{1}^\top\right\|_1 \\
&= \sum_{\tau=1}^{t-\tilde{t}}\left\|\mathbf{e}_i^\top W^{t-\tau}-\frac{1}{N}\mathbb{1}^\top\right\|_1 + \sum_{\tau=t-\tilde{t}+1}^{t}\left\|\mathbf{e}_i^\top W^{t-\tau}-\frac{1}{N}\mathbb{1}^\top\right\|_1 \\
&\le \sum_{\tau=1}^{t-\tilde{t}}\sqrt{N}\sigma_2(W)^{t-\tau}+2\tilde{t} \\
&\le \frac{\sqrt{N}\sigma_2(W)^{\tilde{t}}}{1-\sigma_2(W)}+2\tilde{t} \le \frac{2}{1-\sigma_2(W)}+\frac{\log N}{\log\left[\sigma_2(W)^{-1}\right]},
\end{aligned}
$$

for any $i \in [N]$. □

Proof of Theorem 4.1. We provide the proof in several steps:

Step 1 : Preliminaries

Recall the definition of $\boldsymbol{dE}_1$ in the statement of the theorem. Throughout the proof we refer to the following quantities

$$
\begin{aligned}
\ell &:= \max\left\{\frac{48\log T}{N\Delta_{1,k}^2}, \frac{4Nd}{\Delta_{1,k}}\right\} & c_{t,s} &:= \sqrt{2\log t\left(\frac{1}{Ns}+\frac{2d}{s^2}\right)} \\
\ell' &:= \frac{4\,\boldsymbol{dE}_1}{\Delta_{1,k}} & \hat{t} &:= \frac{5\log\left[\frac{4\sqrt{N}}{\Delta_{1,k}}\right]}{4\log\left[\sigma_2^{-1}(W)\right]}, \qquad (4.5.2)
\end{aligned}
$$

listed here for reader's convenience. To bound the regret (4.2.2), we need to bound the expected number of times that suboptimal arms are played during the entire game. For any $\ell, \ell' > 0$ (and in particular for the choice of ℓ and ℓ' given above), we have

$$
n_{i,T}(k) = 1+\sum_{t=1}^{T}\mathbf{1}\left\{I_{i,t}=k\right\} \le \ell+\sum_{t=1}^{T}\mathbf{1}\left\{I_{i,t}=k, n_{i,t-1}(k)\ge \ell\right\} = \ell + P_T + Q_T, \quad (4.5.3)
$$

where

$$
\begin{aligned}
P_T &:= \sum_{t=1}^{T}\mathbf{1}\left\{I_{i,t}=k, n_{i,t-1}(k)\ge\ell, n_{i,t-1}(1)>\ell'\right\} \\
Q_T &:= \sum_{t=1}^{T}\mathbf{1}\left\{I_{i,t}=k, n_{i,t-1}(k)\ge\ell, n_{i,t-1}(1)\le\ell'\right\}.
\end{aligned}
$$

Though P_T and Q_T depend on i and k, we suppress the dependence to avoid notational clutter. We need to bound P_T and Q_T to complete the proof.

Step 2 : Bounding P_T

For any $k > 1$, we have

$$\begin{aligned}
P_T &= \sum_{t=1}^{T} \mathbf{1}\left\{I_{i,t} = k, n_{i,t-1}(k) \geq \ell, n_{i,t-1}(1) > \ell'\right\} \\
&\leq \sum_{t=1}^{T} \sum_{s_k \geq \ell}^{t} \sum_{s_1 > \ell'}^{t} \mathbf{1}\left\{I_{i,t} = k, n_{i,t-1}(k) = s_k, n_{i,t-1}(1) = s_1\right\} && (4.5.4) \\
&\leq \sum_{t=1}^{T} \sum_{s_k \geq \ell}^{t} \sum_{s_1 > \ell'}^{t} \mathbf{1}\left\{\frac{\phi_{i,t}(k)}{s_k} + c_{t,s_k} \geq \frac{\phi_{i,t}(1)}{s_1} + c_{t,s_1}, n_{i,t-1}(k) = s_k, n_{i,t-1}(1) = s_1\right\}, \\
& && (4.5.5)
\end{aligned}$$

where we recall the definition of $c_{t,s}$ from (4.5.2). Let

$$\mathcal{S}_{k,t} := \left\{\tau \in [t] : I_{i,\tau-1} = k\right\}, \tag{4.5.6}$$

notice the explicit form of $\phi_{i,t}$ given in Lemma 4.1, and recall that $\psi_{i,t}(k) = X_{i,t-1}(k)\mathbf{1}\{k = I_{i,t-1}\}$ for any $k \in [K]$. Then, the indicator (4.5.5) implies that at least one of the following statements must hold

$$\frac{1}{s_1} \sum_{\tau \in \mathcal{S}_{1,t}} \sum_{j=1}^{N} \left[W^{t-\tau}\right]_{ij} \left(X_{j,\tau-1}(1) - \mu_j(1)\right) \leq -c_{t,s_1} \tag{4.5.7}$$

$$\frac{1}{s_k} \sum_{\tau \in \mathcal{S}_{k,t}} \sum_{j=1}^{N} \left[W^{t-\tau}\right]_{ij} \left(X_{j,\tau-1}(k) - \mu_j(k)\right) \geq c_{t,s_k} \tag{4.5.8}$$

$$\frac{1}{s_1} \sum_{\tau \in \mathcal{S}_{1,t}} \sum_{j=1}^{N} \left[W^{t-\tau}\right]_{ij} \mu_j(1) - \frac{1}{s_k} \sum_{\tau \in \mathcal{S}_{k,t}} \sum_{j=1}^{N} \left[W^{t-\tau}\right]_{ij} \mu_j(k) < 2c_{t,s_k}. \tag{4.5.9}$$

We can write

$$\begin{aligned}\text{LHS of (4.5.9)} &= \frac{1}{s_1} \sum_{\tau \in \mathcal{S}_{1,t}} \sum_{j=1}^{N} \left(\left[W^{t-\tau}\right]_{ij} - \frac{1}{N} \right) \mu_j(1) \\ &\quad - \frac{1}{s_k} \sum_{\tau \in \mathcal{S}_{k,t}} \sum_{j=1}^{N} \left(\left[W^{t-\tau}\right]_{ij} - \frac{1}{N} \right) \mu_j(k) + \Delta_{1,k} \\ &\geq -\boldsymbol{dE}_1 \left(\frac{1}{s_1} + \frac{1}{s_k} \right) + \Delta_{1,k} \geq \frac{\Delta_{1,k}}{2}, \qquad (4.5.10)\end{aligned}$$

using the second part of Lemma 4.1 to bound the sums, and noting that $s_k \geq \ell$ and $s_1 > \ell'$ where ℓ and ℓ' are defined in (4.5.2). On the other hand, we have

$$\text{RHS of (4.5.9)} \leq 2c_{T,s_k} \leq \frac{\Delta_{1,k}}{2}, \qquad \forall s_k \geq \ell, \qquad (4.5.11)$$

since by the definition of ℓ in (4.5.2) we have

$$\begin{aligned}4c_{T,s_k}^2 = \frac{8 \log T}{s_k} \left(\frac{1}{N} + \frac{2d}{s_k} \right) &\leq \frac{N\Delta_{1,k}^2}{6} \left(\frac{1}{N} + \frac{2d}{s_k} \right) \\ &\leq \frac{N\Delta_{1,k}^2}{6} \left(\frac{1}{N} + \frac{2\Delta_{1,k} d}{4Nd} \right) \leq \frac{\Delta_{1,k}^2}{4}.\end{aligned}$$

Combining (4.5.9), (4.5.10) and (4.5.11), we get

$$\text{RHS of (4.5.9)} \leq \frac{\Delta_{1,k}}{2} \leq \text{LHS of (4.5.9)} < \text{RHS of (4.5.9)},$$

which results in a contradiction, and implies (4.5.9) never holds for $s_k \geq \ell$ and $s_1 > \ell'$. To study (4.5.7) we use McDiarmid's inequality. When sequences $\{X_{j,\tau-1}(1)\}_{j,\tau}$ and $\{X'_{j,\tau-1}(1)\}_{j,\tau}$ are equal but for the fixed sample (τ', j'), the difference of the sum is bounded as

$$\left| \frac{1}{s_1} \sum_{\tau \in \mathcal{S}_{1,t}} \sum_{j=1}^{N} \left[W^{t-\tau}\right]_{ij} \left(X_{j,\tau-1}(1) - X'_{j,\tau-1}(1) \right) \right| \leq \frac{\left[W^{t-\tau'}\right]_{ij'}}{s_1},$$

and we can compute,

$$
\begin{aligned}
\frac{1}{s_1^2}\sum_{\tau'\in\mathcal{S}_{1,t}}\sum_{j'=1}^{N}\left[W^{t-\tau'}\right]_{ij'}^2 &= \frac{1}{Ns_1}+\frac{1}{s_1^2}\sum_{\tau'\in\mathcal{S}_{1,t}}\sum_{j'=1}^{N}\left(\left[W^{t-\tau'}\right]_{ij'}^2-\frac{1}{N^2}\right)\\
&\le \frac{1}{Ns_1}+\frac{2}{s_1^2}\sum_{\tau'\in\mathcal{S}_{1,t}}\sum_{j'=1}^{N}\left(\left[W^{t-\tau'}\right]_{ij'}-\frac{1}{N}\right)\\
&\le \frac{1}{Ns_1}+\frac{2\boldsymbol{dE}_1}{s_1^2}\le\frac{1}{Ns_1}+\frac{2d}{s_1^2},
\end{aligned}
$$

where the last line is due to the second part of Lemma 4.1. Therefore, we have the right confidence bound to use for McDiarmid's inequality (given that $d \ge \boldsymbol{dE}_1$), and we get

$$
\mathbb{P}\left\{\text{Eq. (4.5.7) holds}\right\}\le\exp\left\{-\log\left(t^4\right)\right\}=\frac{1}{t^4}. \tag{4.5.12}
$$

A similar statement holds for (4.5.8), and combining with (4.5.5) we conclude

$$
\begin{aligned}
\mathbb{E}[P_T] &\le \sum_{t=1}^{T}\sum_{s_k\ge\ell}^{t}\sum_{s_1>\ell'}^{t}\Big(\mathbb{P}\left\{\text{Eq. (4.5.7) holds}\right\}+\mathbb{P}\left\{\text{Eq. (4.5.8) holds}\right\}\Big)\\
&\le \sum_{t=1}^{T}\sum_{s_k\ge\ell}^{t}\sum_{s_1>\ell'}^{t}\frac{2}{t^4}\le\sum_{t=1}^{\infty}\frac{2}{t^2}=\frac{\pi^2}{3}.
\end{aligned} \tag{4.5.13}
$$

Step 3 : Bounding $\boldsymbol{Q_T}$

We now return to bound Q_T as follows. First, note that

$$
\begin{aligned}
Q_T &= \sum_{t=1}^{T}\mathbf{1}\left\{I_{i,t}=k, n_{i,t-1}(k)\ge\ell, n_{i,t-1}(1)\le\ell'\right\}\\
&\le \sum_{s_1=1}^{\ell'}\sum_{t=1}^{T}\mathbf{1}\left\{I_{i,t}=k, n_{i,t-1}(k)\ge\ell, n_{i,t-1}(1)=s_1\right\}.
\end{aligned}
$$

Let us for each $s_1 \in [1, \ell']$ denote by t_{s_1} the first time that the indicator holds for the particular value of s_1. Fixing any $\hat{t} > 0$, we have

$$
\begin{aligned}
Q_T &\leq \sum_{s_1=1}^{\ell'} \sum_{t=t_{s_1}}^{t_{s_1+1}-1} \mathbf{1}\left\{I_{i,t} = k, n_{i,t-1}(k) \geq \ell, n_{i,t-1}(1) = s_1\right\} \\
&\leq \ell'\hat{t} + \sum_{s_1=1}^{\ell'} \sum_{t=t_{s_1}+\hat{t}}^{t_{s_1+1}-1} \mathbf{1}\left\{I_{i,t} = k, n_{i,t-1}(k) \geq \ell, n_{i,t-1}(1) = s_1\right\} \\
&\leq \ell'\hat{t} + \sum_{s_1=1}^{\ell'} \sum_{t=t_{s_1}+\hat{t}}^{t_{s_1+1}-1} \sum_{s_k=\ell}^{t} \mathbf{1}\left\{I_{i,t} = k, n_{i,t-1}(k) = s_k, n_{i,t-1}(1) = s_1\right\}, \qquad (4.5.14)
\end{aligned}
$$

where the last sum is similar to (4.5.4) with different indices. Hence, to satisfy the indicator, at least one of the statements (4.5.7), (4.5.8) and (4.5.9) must hold (for new indices). Since $s_k \geq \ell$ the analysis of RHS of (4.5.9) given in (4.5.11) is still valid. Observe that by standard properties of irreducible and aperiodic Markov chains we have [46],

$$
\sum_{j=1}^{N} \left| \left[W^t\right]_{ij} - \frac{1}{N} \right| \leq \sqrt{N}\sigma_2^t(W) < \frac{\Delta_{1,k}}{4}, \qquad \forall t > \frac{\log\left[\frac{4\sqrt{N}}{\Delta_{1,k}}\right]}{\log\left[\sigma_2^{-1}(W)\right]}. \qquad (4.5.15)
$$

To analyze the LHS, let $\hat{t}$ be defined as in (4.5.2) and recall (4.5.6). Then, for any $s_1 \in [1, \ell']$ and $t \in [t_{s_1} + \hat{t}, t_{s_1+1} - 1]$ we have $\mathcal{S}_{1,t} = \mathcal{S}_{1,t_{s_1}}$ by definition of t_{s_1}. Hence, we modify the expression in (4.5.10) as

$$
\begin{aligned}
\text{LHS of (4.5.9)} &= \frac{1}{s_1} \sum_{\tau \in \mathcal{S}_{1,t_{s_1}}} \sum_{j=1}^{N} \left(\left[W^{t-\tau}\right]_{ij} - \frac{1}{N} \right) \mu_j(1) \\
&\quad - \frac{1}{s_k} \sum_{\tau \in \mathcal{S}_{k,t}} \sum_{j=1}^{N} \left(\left[W^{t-\tau}\right]_{ij} - \frac{1}{N} \right) \mu_j(k) + \Delta_{1,k} \\
&\geq -\sqrt{N}\sigma_2^{\hat{t}}(W) - \frac{\boldsymbol{dE}_1}{s_k} + \Delta_{1,k} \geq \frac{\Delta_{1,k}}{2},
\end{aligned}
$$

where in the last line we used Lemma 4.1, equation (4.5.15) and the fact that $s_k \geq \ell$. Combining above with (4.5.11) implies that (4.5.9) never holds. Notice that our argument about the probability of events (4.5.7) and (4.5.8) holds for any $s_1, s_k, t > 0$, and therefore, the tail bound (4.5.12) holds true again. Employing these facts and returning to (4.5.14) we get

$$\mathbb{E}[Q_T] \leq \ell' \hat{t} + \frac{\pi^2}{3}. \tag{4.5.16}$$

Step 4 : Finishing the Proof

Substituting (4.5.13) and (4.5.16) into (4.5.3) gives us the bound

$$\mathbb{E}[n_{i,T}(k)] \leq \ell + \mathbb{E}[P_T] + \mathbb{E}[Q_T] \leq \ell + \ell' \hat{t} + \frac{2\pi^2}{3}. \tag{4.5.17}$$

Recall the definition of $\boldsymbol{dE}_1$ and $\boldsymbol{dE}_2$ from the statement of the theorem and the fact that N is large enough. Then, plugging the above into (4.2.2) using quantities defined in (4.5.2) concludes the proof. □

Proof of Lemma 4.2. Noting the contiguous intervals $\mathcal{I}_k := (\sum_{s=0}^{k-1} a_{m-s}, \sum_{s=0}^{k} a_{m-s}]$ for any $k \in [m-1]$, the sum in the LHS of (4.3.1) is an under approximation of the area under the curve x^{-2} on the interval $x \in \bigcup_{k=1}^{m-1} \mathcal{I}_k = [a_m, \sum_{k=1}^m a_k]$, and therefore,

$$\sum_{k=1}^{m-1} \frac{a_k}{\left(\sum_{s=k}^m a_s\right)^2} \leq \int_{a_m}^{\infty} \frac{1}{x^2} d_x = \frac{1}{a_m}.$$

Now let $a_k = \Delta_{k,k+1}$, and observe that

$$\sum_{k=1}^{m-1} \frac{\Delta_{k,k+1}}{\Delta_{k,m}^2} = \frac{1}{\Delta_{m-1,m}} + \sum_{k=1}^{m-2} \frac{\Delta_{k,k+1}}{\left(\sum_{s=k}^{m-1} \Delta_{s,s+1}\right)^2} \leq \frac{2}{\Delta_{m-1,m}}.$$

□

Proof of Theorem 4.2. We slightly change the notation introduced in (4.5.2) as follows,

$$\ell_{km} := \max\left\{\frac{48 \log T}{N\Delta_{k,m}^2}, \frac{4Nd}{\Delta_{k,m}}\right\} \qquad c_{t,s} := \sqrt{2\log t\left(\frac{1}{Ns} + \frac{2d}{s^2}\right)}$$

$$\ell'_{km} := \frac{4\,\boldsymbol{dE}_1}{\Delta_{k,m}} \qquad \hat{t}_{km} := \frac{5\log\left[\frac{4\sqrt{N}}{\Delta_{k,m}}\right]}{4\log\left[\sigma_2^{-1}(W)\right]}. \tag{4.5.18}$$

Let us now proceed with bounding $n_{i,T}\left(m \mid k\right)$ in (4.2.3) which represents the number of times that m was played by agent i given that at least one arm in the set $[k]$ was awake $(k < m)$. Recalling that $A_{i,t}$ represents the set of awake arms at time $t \in [T]$ for player $i \in [N]$, we have

$$\begin{aligned}
n_{i,T}\left(m \mid k\right) &= \sum_{t=1}^{T} \mathbf{1}\left\{I_{i,t} = m, A_{i,t} \cap [k] \neq \emptyset\right\} \\
&\leq \ell_{km} + \sum_{t=1}^{T} \mathbf{1}\left\{I_{i,t} = m, A_{i,t} \cap [k] \neq \emptyset, n_{i,t}(m) \geq \ell_{km}\right\} \\
&\leq \ell_{km} + \sum_{k'=1}^{k}\sum_{t=1}^{T} \mathbf{1}\left\{I_{i,t} = m, k' \in A_{i,t}, n_{i,t}(m) \geq \ell_{km}\right\} \\
&\leq \ell_{km} + \sum_{k'=1}^{k}\sum_{t=1}^{T} \mathbf{1}\left\{I_{i,t} = m, k' \in A_{i,t}, n_{i,t}(m) \geq \ell_{km}, n_{i,t}(k') > \ell'_{km}\right\} \\
&\quad\; + \sum_{k'=1}^{k}\sum_{t=1}^{T} \mathbf{1}\left\{I_{i,t} = m, k' \in A_{i,t}, n_{i,t}(m) \geq \ell_{km}, n_{i,t}(k') \leq \ell'_{km}\right\},
\end{aligned}$$

where we arrive to a similar equation to (4.5.3). Therefore, following exactly the lines in the proof of Theorem 4.1, the final bound resembles the one in (4.5.17), and we obtain

$$E\left[n_{i,T}\left(m \mid k\right)\right] \leq \ell_{km} + k\ell'_{km}\hat{t}_{km} + \frac{2k\pi^2}{3}. \tag{4.5.19}$$

Note that since we used a new notation (4.5.18) in the proof of this theorem, in above we replaced the variables in (4.5.17) with their corresponding quantities defined in (4.5.18). Also, the extra factor of k is an artifact of the outer summation over $k' \in [k]$. Since $\log x \leq x$ for $x > 0$, we can bound

$$\begin{aligned}
\ell'_{km}\hat{t}_{km} &= \frac{4\,\boldsymbol{dE}_1}{\Delta_{k,m}}\left(\frac{5\log\left[\frac{4\sqrt{N}}{\Delta_{k,m}}\right]}{4\log\left[\sigma_2^{-1}(W)\right]}\right) \\
&\leq \frac{5\,\boldsymbol{dE}_1}{\Delta_{k,m}}\left(\frac{0.5\log N + 4\Delta_{k,m}^{-1}}{\log\left[\sigma_2^{-1}(W)\right]}\right) \\
&\leq \frac{5\,\boldsymbol{dE}_1}{\Delta_{k,m}^2}\left(\frac{0.5\log N + 4}{\log\left[\sigma_2^{-1}(W)\right]}\right) \leq \frac{15\log N}{\log\left[\sigma_2^{-1}(W)\right]}\Delta_{k,m}^{-2}\,\boldsymbol{dE}_1,
\end{aligned}$$

since $\Delta_{k,m} \leq 1$ and N is large enough. Recalling (4.5.18) as well as the definition of $\boldsymbol{dE}_2$ from Theorem 4.1, we can simplify (4.5.19) using above as follows,

$$E\left[n_{i,T}\left(m \mid k\right)\right] \leq \frac{\frac{48}{N}\log T + 4Nd + 15k \;\; \boldsymbol{dE}_1\, \boldsymbol{dE}_2}{\Delta_{k,m}^2} + \frac{2k\pi^2}{3}.$$

Substituting above into (4.2.3) and noting that the bound is independent of i, we obtain

$$\begin{aligned}
\overline{\mathbf{R}}_T &\leq \sum_{m=2}^{K}\sum_{k=1}^{m-1} \Delta_{k,k+1}\left(\frac{\frac{48}{N}\log T + 4Nd + 15k \;\; \boldsymbol{dE}_1\, \boldsymbol{dE}_2}{\Delta_{k,m}^2} + \frac{2k\pi^2}{3}\right) \\
&\leq \sum_{m=2}^{K}\left\{\left(\frac{48}{N}\log T + 4Nd + 15m\, \boldsymbol{dE}_1\, \boldsymbol{dE}_2\right)\sum_{k=1}^{m-1}\frac{\Delta_{k,k+1}}{\Delta_{k,m}^2}\right\} + \sum_{m=2}^{K}\frac{2m\pi^2}{3}\Delta_{1,m} \\
&\leq \sum_{m=2}^{K}\left\{\frac{\frac{96}{N}\log T + 8Nd + 30m\, \boldsymbol{dE}_1\, \boldsymbol{dE}_2}{\Delta_{m-1,m}} + \frac{2m\pi^2}{3}\Delta_{1,m}\right\},
\end{aligned}$$

where in the last step we applied (4.3.2). □

Chapter 5

Online Optimization in Dynamic Environments

Multi-armed bandit is recognized as a special case (partial feedback version) of the well-known expert advice problem [101–103]. The expert advice problem, itself, can be categorized in the class of online linear optimization problems. More generally, online convex optimization has been well-studied in the literature, and there are numerous algorithms solving the problem in *static* regime. In this chapter, we revisit the topic using non-static performance metric to shed light on the behavior of online algorithms in *dynamic* environments. The content of this chapter is mostly relevant to the work of [104].

In an online optimization problem, a *learner* plays against an *adversary* or *nature*. At each round $t \in \{1, \ldots, T\}$, the learner chooses an action x_t from some convex feasible set $\mathcal{X} \subseteq \mathbb{R}^d$. Then, nature reveals a convex function $f_t \in \mathcal{F}$ to the learner. As a result, the learner incurs the corresponding *loss* $f_t(x_t)$. A learner aims to minimize his *regret*, a comparison to a single best action in hindsight:

$$\mathbf{Reg}_T^s := \sum_{t=1}^{T} f_t(x_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^{T} f_t(x). \tag{5.0.1}$$

Let us refer to this as *static* regret in the sense that the comparator is *time-invariant*. In the literature, there are numerous algorithms that guarantee a static regret rate of $\mathcal{O}(\sqrt{T})$ (see e.g. [19–21]). Moreover, when the loss functions are *strongly* convex, a rate of $\mathcal{O}(\log T)$ could be achieved [105]. Furthermore, minimax optimality of algorithms with respect to the worst-case adversary has been established (see e.g. [106]).

There are two major directions in which the above-mentioned results can be strengthened: (1) by exhibiting algorithms that compete with non-static comparator sequences (that is, making the benchmark harder), and (2) by proving regret guarantees that take advantage of *niceness* of nature's sequence (that is, exploiting some non-adversarial quality of nature's moves). Both of these distinct directions are important avenues of investigation. In the present chapter, we attempt to address these two aspects by developing a single, adaptive algorithm with a regret bound that shows the interplay between the difficulty of the comparison sequence and niceness of the sequence of nature's moves.

With respect to the first aspect, a more stringent benchmark is a *time-varying* comparator, a notion that can be termed *dynamic* regret [21, 107–109]:

$$\mathbf{Reg}_T^d := \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(x_t^*), \tag{5.0.2}$$

where $x_t^* := \text{argmin}_{x \in \mathcal{X}} f_t(x)$. More generally, dynamic regret against a comparator sequence $\{u_t\}_{t=1}^T$ is

$$\mathbf{Reg}_T^d(u_1, \dots, u_T) := \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(u_t).$$

It is well-known that in the worst case, obtaining a bound on dynamic regret is not possible. However, it is possible to achieve worst-case bounds in terms of

$$C_T(u_1, \dots, u_T) := \sum_{t=1}^{T} \big\|u_t - u_{t-1}\big\|, \tag{5.0.3}$$

i.e., the *regularity* of the comparator sequence, interpolating between the static and dynamic regret notions. Furthermore, the authors in [110] introduce an algorithm which proposes a variant of C_T involving a dynamical model.

In terms of the second direction, there are several ways of incorporating potential regularity of nature's sequence. The authors in [111, 112] bring forward the idea of predictable sequences, a generic way to incorporate some external knowledge about the gradients of the loss functions. Let $\{M_t\}_{t=1}^{T}$ be a *predictable* sequence computable by the learner at the beginning of round t. This sequence can then be used by an algorithm in order to achieve regret in terms of

$$D_T := \sum_{t=1}^{T} \big\|\nabla f_t(x_t) - M_t\big\|_*^2. \tag{5.0.4}$$

The framework of predictable sequences captures *variation* and *path-length* type regret bounds (see e.g. [113–115]). Yet another way in which niceness of the adversarial sequence can be captured is through a notion of *temporal variability* studied in [116]:

$$V_T := \sum_{t=1}^{T} \sup_{x \in \mathcal{X}} \big|f_t(x) - f_{t-1}(x)\big|. \tag{5.0.5}$$

What is interesting—and intuitive— is that dynamic regret against the optimal sequence $\{x_t^*\}_{t=1}^{T}$ becomes a feasible objective when V_T is *small*. When only noisy versions of gradients are revealed to the algorithm, Besbes et al. in [116] show that using a restarted Online Gradient Descent (`OGD`) [21] algorithm, one can get a bound of form $T^{2/3}(V_T + 1)^{1/3}$ on the expected regret. However, the regret bounds attained in [116] are only valid when an upper bound on V_T is known to the learner before the game begins. For the full information online convex optimization setting, when one receives exact gradients instead of noisy gradients, a bound of order V_T is trivially obtained by simply playing (at each round) the minimum of the previous round.

The three quantities we just introduced — C_T, D_T, V_T — measure distinct aspects of the online optimization problem, and their interplay is an interesting object of study. Our main contribution, presented in Section 5.2, is to develop a fully adaptive method (without prior knowledge of these quantities) whose dynamic regret is given in terms of these three complexity measures. This is done for the full information online convex optimization setting, and augments the existing regret bounds in the literature which focus on only one of the three notions — C_T, D_T, V_T — (and not all the three together). To establish a sub-linear bound on the dynamic regret, we utilize a variant of the Optimistic Mirror Descent (`OMD`) algorithm [111].

When noiseless gradients are available and we can calculate variations at each round, we not only establish a regret bound in terms of V_T and T (without a priori knowledge of a bound on V_T), but also show how the bound can in fact be improved when deviation D_T is $o(T)$. We further also show how the bound can automatically adapt to C_T the length of sequence of comparators. Importantly, this avoids suboptimal bounds derived only in terms of one of the quantities — C_T, V_T — in an environment where the other one is small.

The second contribution of this work is the technical analysis of the algorithm. The bound on the dynamic regret is derived by applying the *doubling trick* to a non-monotone quantity which results in a non-monotone step size sequence (which has not been investigated to the best of authors' knowledge).

As an instance of learning in dynamic environments, we provide uncoupled strategies for two players playing a sequence of drifting zero sum games (Section 5.3). We show how when the two players play the provided strategies, their payoffs converge to the average minimax value of the sequence of games (provided the games drift slowly). In this case, both players simultaneously enjoy no regret guarantees against best sequences of actions in hindsight that vary slowly. This is a generalization of the results by Daskalakis *et al.* [117], and Rakhlin *et al.* [112], both of which are for fixed games played repeatedly.

5.1 Preliminaries

5.1.1 Notation

Throughout, we assume that for any action $x \in \mathcal{X} \subset \mathbb{R}^d$ at any time t, it holds that

$$|f_t(x)| \leq G. \tag{5.1.1}$$

We denote by $\|\cdot\|_*$ the dual norm of $\|\cdot\|$, by $[T]$ the set of natural numbers $\{1, \ldots, T\}$, and by $f_{1:t}$ the shorthand of $f_1, ..., f_t$, respectively. Whenever C_T is written without arguments, it will refer to regularity $C_T(x_1^*, \ldots, x_T^*)$ of the sequence of minimizers of the loss functions. We point out that our initial statements hold for the regularity of any sequence of comparators. However, for upper bounds involving $\sqrt{C_T}$, one needs to choose a computable quantity to tune the step size, and hence our main results are stated for $C_T(x_1^*, \ldots, x_T^*)$.

The quantity D_T is defined with respect to an arbitrary predictable sequence $\{M_t\}_{t=1}^T$, but this dependence is omitted for brevity.

5.1.2 Existing Regret Bounds in the Dynamic Setting

We state and discuss relevant results from the literature on online learning in dynamic environments. For any comparator sequence $\{u_t\}_{t=1}^T$ and the specific minima sequence $\{x_t^*\}_{t=1}^T$ the following results are established in the literature:

Reference	Regret Notion
	Regret Rate
[21] [110]	$\sum_{t=1}^{T} f_t(x_t) - f_t(u_t)$
	$\mathcal{O}\left(\sqrt{T}(1 + C_T(u_1, \ldots, u_T))\right)$
[116]	$\sum_{t=1}^{T} \mathbb{E}\left[f_t(x_t)\right] - f_t(x_t^*)$
	$\mathcal{O}\left(T^{2/3}(1 + V_T)^{1/3}\right)$
[112]	$\sum_{t=1}^{T} f_t(x_t) - f_t(u)$
	$\mathcal{O}\left(\sqrt{D_T}\right)$
Our work	$\sum_{t=1}^{T} f_t(x_t) - f_t(x_t^*)$
	$\tilde{\mathcal{O}}\left(\sqrt{D_T + 1} + \min\left\{\sqrt{(D_T + 1)C_T}, (D_T + 1)^{1/3}T^{1/3}V_T^{1/3}\right\}\right)$

Table 5.1: Comparison of the results

where $\tilde{\mathcal{O}}(\cdot)$ hides the $\log T$ factor. In our initial result, Lemma 5.1 below also yields a rate of $\mathcal{O}\left(\sqrt{D_T + 1}(1 + C_T(u_1, \ldots, u_T))\right)$ for any comparator sequence $\{u_t\}_{t=1}^{T}$. A detailed explanation of the bounds will be done after Theorem 5.1.

We remark that the authors in [116] consider a setting in which a *variation budget* (an upper bound on V_T) is known to the learner, but he/she only has noisy gradients available. Then, the restarted OGD guarantees the mentioned rate for convex functions; the rate is modified to $\sqrt{(V_T + 1)T}$ for strongly convex functions.

For the case of *noiseless* gradients, we first aim to show that our algorithm is adaptive in the sense that the learner needs not know an upper bound on V_T in advance when he/she can calculate variations observed so far. Furthermore, we shall establish that our method recovers the known bounds for stationary settings (as well as cases where V_T does not change gradually along the time horizon)

5.1.3 Comparison of Regularity and Variability

We now show that V_T and C_T are not comparable in general. To this end, we consider the classical problem of prediction with expert advice. In this setting, the learner deals with the linear loss $f_t(x) = \langle f_t, x\rangle$ on the d-dimensional probability simplex. Assume that for any $t \geq 1$, we have the vector sequence

$$f_t = \begin{cases} (-\frac{1}{T}, 0, 0, \dots, 0) & \text{, if} \quad t \text{ even} \\ (0, -\frac{1}{T}, 0, \dots, 0) & \text{, if} \quad t \text{ odd} \end{cases} .$$

Setting u_t, the comparator of round t, to be the minimizer of f_t, i.e. $u_t = x_t^*$, we have

$$C_T = \sum_{t=1}^{T} \|x_t^* - x_{t-1}^*\|_1 = \Theta(T) \qquad V_T = \sum_{t=1}^{T} \|f_t - f_{t-1}\|_\infty = \mathcal{O}\left(1\right),$$

according to (5.0.3) and (5.0.5), respectively. We see that V_T is considerably smaller than C_T in this scenario. On the other hand, consider prediction with expert advice with two experts. Let $f_t = (-1/2, 0)$ on even rounds and $f_t = (0, 1/2)$ on odd rounds. Expert 1 remains to be the best throughout the game, and thus $C_T = \mathcal{O}(1)$, while variation $V_T = \Theta(T)$. Therefore, one can see that taking into account only one measure might lead us to suboptimal regret bounds. We show that both measures play a key role in our regret bound. Finally, we note that if $M_t = \nabla f_{t-1}(x_{t-1})$, the notion of D_T can be related to V_T in certain cases, yet we keep the predictable sequence arbitrary and thus as playing a role separate from V_T and C_T.

5.2 Adaptive Optimistic Mirror Descent

5.2.1 Optimistic Mirror Descent and Relation to Regularity

We now outline the `OMD` algorithm previously proposed in [111]. Let $\mathcal{R}$ be a 1-strongly convex function with respect to a norm $\|\cdot\|$, and $\mathcal{D}_{\mathcal{R}}(\cdot,\cdot)$ represent the Bregman divergence with respect to

$\mathcal{R}$. Also, let $\mathcal{H}_t$ be the set containing all available information to the learner at the beginning of time t. Then, the learner can compute the vector $M_t : \mathcal{H}_t \to \mathbb{R}^d$, which we call the predictable process. Supposing that the learner has access to the side information $M_t \in \mathbb{R}^d$ from the outset of round t, the `OMD` algorithm is characterized via the following interleaved sequence,

$$x_t = \text{argmin}_{x \in \mathcal{X}} \Big\{ \eta_t \langle x, M_t \rangle + \mathcal{D}_{\mathcal{R}}(x, \hat{x}_{t-1}) \Big\} \tag{5.2.1}$$

$$\hat{x}_t = \text{argmin}_{x \in \mathcal{X}} \Big\{ \eta_t \langle x, \nabla_t \rangle + \mathcal{D}_{\mathcal{R}}(x, \hat{x}_{t-1}) \Big\}, \tag{5.2.2}$$

where $\nabla_t := \nabla f_t(x_t)$, and η_t is the *step size* that can be chosen adaptively to attain low regret. One could observe that for $M_t = 0$, the `OMD` algorithm amounts to the well-known *Mirror Descent* algorithm [118, 119]. On the other hand, the special case of $M_t = \nabla_{t-1}$ recovers the scheme proposed in [114]. It is shown in [111] that the static regret satisfies

$$\mathbf{Reg}_T^s \leq 4 R_{\max} \left(\sqrt{D_T} + 1 \right),$$

using the step size

$$\eta_t = R_{\max} \min \left\{ \left(\sqrt{D_{t-1}} + \sqrt{D_{t-2}} \right)^{-1}, 1 \right\},$$

where $R_{\max}^2 := \sup_{x,y \in \mathcal{X}} \mathcal{D}_{\mathcal{R}}(x, y)$. The following lemma extends the result to arbitrary sequence of comparators $\{u_t\}_{t=1}^T$. Throughout, we assume that $\|\nabla_0 - M_0\|_*^2 = 1$ by convention.

Lemma 5.1. *Let $\mathcal{X}$ be a convex set in a Banach space $\mathcal{B}$. Let $\mathcal{R} : \mathcal{B} \mapsto \mathbb{R}$ be a 1-strongly convex function on $\mathcal{X}$ with respect to a norm $\|\cdot\|$, and let $\|\cdot\|_*$ denote the dual norm. For any $L > 0$, employing the time-varying step size*

$$\eta_t = \frac{L}{\sqrt{\sum_{s=0}^{t-1} \|\nabla_s - M_s\|_*^2} + \sqrt{\sum_{s=0}^{t-2} \|\nabla_s - M_s\|_*^2}},$$

and running the Optimistic Mirror Descent algorithm for any comparator sequence $\{u_t\}_{t=1}^T$*, yields*

$$\mathbf{Reg}_T^d(u_1, \ldots, u_T) \leq 2\sqrt{1 + D_T}L + 2\sqrt{1 + D_T}\frac{\gamma C_T(u_1, \ldots, u_T) + 4R_{\max}^2}{L},$$

so long as $\mathcal{D}_\mathcal{R}(x, z) - \mathcal{D}_\mathcal{R}(y, z) \leq \gamma\|x - y\|, \forall x, y, z \in \mathcal{X}$*.*

Lemma 5.1 underscores the fact that one can get a tighter bound for regret once the learner advances a sequence of conjectures $\{M_t\}_{t=1}^T$ well-aligned with the gradients. Moreover, if the learner has prior knowledge of C_T (or an upper bound on it), then the regret bound would be $\mathcal{O}\left(\sqrt{(D_T + 1)C_T}\right)$ by tuning L.

Note that when the function $\mathcal{R}$ is Lipschitz on $\mathcal{X}$, the Lipschitz condition on the Bregman divergence is automatically satisfied. For the particular case of KL divergence this can be achieved via mixing a uniform distribution to stay away from boundaries (see e.g. Section 5.3.2 in this regard). In this case, the constant γ is of $\mathcal{O}(\log T)$.

5.2.2 The Adaptive Optimistic Mirror Descent Algorithm

The main objective of the chapter is to develop the *Adaptive Optimistic Mirror Descent (*`AOMD`*)* algorithm. The `AOMD` algorithm incorporates all notions of variation D_T, C_T and V_T to derive a comprehensive regret bound. The proposed method builds on the `OMD` algorithm with adaptive step size, combined with a *doubling trick* applied to a threshold growing non-monotonically (see e.g. [19, 111] for application of doubling trick on monotone quantities). The scheme is adaptive in the sense that no prior knowledge of D_T, C_T or V_T is necessary.

Observe that the prior knowledge of a variation budget (an upper bound on V_T) does not tell us how the changes between cost functions are distributed throughout the game. For instance, the variation can increase gradually along the time horizon, while it can also take place in the form of discrete switches. The learner does not have any information about the variation pattern. Therefore, she must adopt a flexible strategy that achieves low regret in the benign case of finite switches or shocks, while it is simultaneously able to compete with the worst-case of gradual change. Before describing the algorithm, let us first use Lemma 5.1 to bound the general dynamic regret in terms of D_T, C_T and V_T.

Lemma 5.2. *Let $\mathcal{X}$ be a convex set in a Banach space $\mathcal{B}$. Let $\mathcal{R} : \mathcal{B} \mapsto \mathbb{R}$ be a 1-strongly convex function on $\mathcal{X}$ with respect to a norm $\|\cdot\|$. Run the Optimistic Mirror Descent algorithm with the step size given in the statement of Lemma 5.1. Letting the comparator sequence be $\{u_t\}_{t=1}^T$, for any $L > 2R_{\max}$ we have*

$$\mathbf{Reg}_T^d(u_1, \ldots, u_T) \leq 4\sqrt{1 + D_T}L + \mathbf{1}\left\{\gamma C_T(u_1, \ldots, u_T) > L^2 - 4R_{\max}^2\right\} \frac{4\gamma R_{\max} T V_T}{L^2 - 4R_{\max}^2},$$

so long as $\mathcal{D}_{\mathcal{R}}(x, z) - \mathcal{D}_{\mathcal{R}}(y, z) \leq \gamma\|x - y\|, \forall x, y, z \in \mathcal{X}$.

We now describe AOMD algorithm shown in table 5, and prove that it automatically adapts to V_T, D_T and C_T. The algorithm can be cast as a repeated OMD using different step sizes. The learner sets the parameter $L = 3R_{\max}$ in Lemma 5.1, and runs the OMD algorithm. Along the process, the learner collects deviation, variation and regularity observed so far, and checks the doubling condition in table 5 after each round. Once the condition is satisfied, the learner doubles L, discards the accumulated deviation, variation and regularity, and runs a new OMD algorithm. Note importantly that the doubling condition results in a non-monotone sequence of step size during the learning process.

Algorithm 5 Adaptive Optimistic Mirror Descent Algorithm

```
Parameter : $R_{\max}$, some arbitrary $x_0 \in \mathcal{X}$
Initialize $N = 1$, $C_{(1)} = V_{(1)} = 0$, $D_{(1)} = 1$, $x_1 = x_0$, $L_1 = 3R_{\max}$, $\Delta_1 = 0$ and $k_1 = 1$.
for $t = 1$ to $T$ do
  % check doubling condition
  if $L_N^2 < \gamma \min\left\{C_{(N)}\ ,\ V_{(N)}^{2/3}\Delta_N^{2/3}D_{(N)}^{-1/3}\right\} + 4R_{\max}^2$ then
    % increment $N$ and double $L_N$
    $N = N + 1$
    $L_N = 3R_{\max}2^{N-1}$, $C_{(N)} = V_{(N)} = 0$, $D_{(N)} = 1$ and $\Delta_N = 0$
    $k_N = t$
  end if
  Play $x_t$ and suffer loss $f_t(x_t)$
  Calculate $M_{t+1}$ (predictable sequence) and gradient $\nabla_t = \nabla f_t(x_t)$
  % update $D_{(N)}, C_{(N)}, V_{(N)}$ and $\Delta_N$
  $D_{(N)} = D_{(N)} + \|\nabla_t - M_t\|_*^2$
  $C_{(N)} = C_{(N)} + \|x_t^* - x_{t-1}^*\|$
  $V_{(N)} = V_{(N)} + \sup_{x \in \mathcal{X}} |f_t(x) - f_{t-1}(x)|$
  $\Delta_N = \Delta_N + 1$
  % set step-size and perform optimistic mirror descent update
  $\eta_{t+1} = L_N \left(\sqrt{D_{(N)}} + \sqrt{D_{(N)} - \|\nabla_t - M_t\|_*^2}\right)^{-1}$
  $\hat{x}_t = \operatorname*{argmin}_{x \in \mathcal{X}} \left\{\eta_t \langle x, \nabla_t \rangle + \mathcal{D}_{\mathcal{R}}(x, \hat{x}_{t-1})\right\}$
  $x_{t+1} = \operatorname*{argmin}_{x \in \mathcal{X}} \left\{\eta_{t+1} \langle x, M_{t+1} \rangle + \mathcal{D}_{\mathcal{R}}(x, \hat{x}_t)\right\}$
end for
```

Notice that once we have completed running the algorithm, N is the number of doubling epochs, Δ_i is the number of instances in epoch i, k_i and $k_{i+1} - 1$ are the start and end points of epoch i, $\sum_{i=1} \Delta_i = T$, $\sum_{i=1}^N C_{(i)} = C_T$, $\sum_{i=1}^N D_{(i)} = D_T + N$ and $\sum_{i=1}^N V_{(i)} = V_T$. Also, there is a technical reason for initialization choice of L which shall become clear in the proof of Lemma 5.2.

Theorem 5.1 shows the bound enjoyed by the proposed `AOMD` algorithm.

Theorem 5.1. *Assume that $\mathcal{D}_{\mathcal{R}}(x,z) - \mathcal{D}_{\mathcal{R}}(y,z) \le \gamma\|x-y\|, \forall x,y,z \in \mathcal{X}$, and let $C_T = \sum_{t=1}^{T} \left\|x_t^* - x_{t-1}^*\right\|$. The* `AOMD` *algorithm enjoys the following bound on dynamic regret :*

$$\mathbf{Reg}_T^d \le \tilde{\mathcal{O}}\left(\sqrt{D_T+1}\right) + \tilde{\mathcal{O}}\left(\min\left\{\sqrt{(D_T+1)C_T}, (D_T+1)^{1/3}T^{1/3}V_T^{1/3}\right\}\right),$$

where $\tilde{\mathcal{O}}(\cdot)$ hides a $\log T$ factor.

Based on Theorem 5.1 we can obtain the following table that summarizes bounds on $\mathbf{Reg}_T^d$ for various cases (disregarding the first term $\tilde{\mathcal{O}}\left(\sqrt{D_T+1}\right)$ in the bound above):

Regime	Rate
$C_T \le T^{2/3}(D_T+1)^{-1/3}V_T^{2/3}$	$\tilde{\mathcal{O}}\left(\sqrt{C_T(D_T+1)}\right)$
$V_T \le D_T+1$	$\tilde{\mathcal{O}}\big((D_T+1)^{2/3}T^{1/3}\big)$
$D_T \le V_T-1$	$\tilde{\mathcal{O}}\big(V_T^{2/3}T^{1/3}\big)$
$D_T = \mathcal{O}(T)$	$\tilde{\mathcal{O}}\big(T^{2/3}V_T^{1/3}\big)$

Table 5.2: Regret bound in different regimes

The following remarks are in order :

- In all cases, given the condition $V_T = o(T)$, the regret is sub-linear. When the gradients are bounded, the regime $D_T = \mathcal{O}(T)$ always holds, guaranteeing the worst-case bound of $\tilde{\mathcal{O}}\big(T^{2/3}V_T^{1/3}\big)$.
- Theorem 5.1 allows us to recover $\tilde{\mathcal{O}}(1)$ regret for certain cases where $V_T = \mathcal{O}(1)$. Let nature divide the horizon into B batches, and play a smooth convex function $f_i(x)$ on each batch $i \in [B]$, that is for some $H_i > 0$ it holds that

$$\|\nabla f_i(x) - \nabla f_i(y)\|_* \le H_i\|x-y\|, \tag{5.2.3}$$

$\forall i \in [B]$ and $\forall x, y \in \mathcal{X}$. Set $M_t = \nabla f_i(\hat{x}_{t-1})$ and note that the gradients are Lipschitz continuous. In this case, the `OMD` corresponding to each batch can be recognized as the *Mirror Prox* method [120], which results in $\tilde{\mathcal{O}}(1)$ regret during each period. Also, since $C_T = \mathcal{O}(1)$ the bound in Theorem 5.1 is of $\mathcal{O}(\log T)$.

5.3 Applications

5.3.1 Competing with Strategies

So far, we mainly considered dynamic regret $\mathbf{Reg}_T^d$ defined in Equation 5.0.2. However, in many scenarios one might want to consider regret against a more specific set of strategies, defined as follows :

$$\mathbf{Reg}_T^\Pi := \sum_{t=1}^T f_t(x_t) - \inf_{\pi \in \Pi} \sum_{t=1}^T f_t(\pi_t(f_{1:t-1})),$$

where each $\pi \in \Pi$ is a sequence of mappings $\pi = (\pi_1, \ldots, \pi_T)$ and $\pi_t : \mathcal{F}^{t-1} \to \mathcal{X}$. Notice that if Π is the set of all mappings then $\mathbf{Reg}_T^\Pi$ corresponds to dynamic regret $\mathbf{Reg}_T^d$ and if Π corresponds to set of constant history independent mappings, that is, each $\pi \in \Pi$ is indexed by some $x \in \mathcal{X}$ and $\pi_1^x(\cdot) = \ldots = \pi_T^x(\cdot) = x$, then $\mathbf{Reg}_T^\Pi$ corresponds to the static regret $\mathbf{Reg}_T^s$. We now define

$$C_T^\Pi = \sum_{t=1}^T \left\| \pi_t^*(f_{1:t-1}) - \pi_{t-1}^*(f_{1:t-2}) \right\|,$$

where $\pi_t^* = \mathrm{arginf}_{\pi \in \Pi} \sum_{s=1}^t f_s(\pi_s(f_{1:s-1}))$. Assume that there exists sequence of mappings $\tilde{C}_1, \ldots, \tilde{C}_T$ where $\tilde{C}_t$ maps any $f_1, \ldots, f_t$ to reals and is such that for any t and any $f_1, \ldots, f_{t-1}$,

$$\tilde{C}_{t-1}(f_{1:t-1}) \leq \tilde{C}_t(f_{1:t}),$$

and further, for any T and any $f_1, \ldots, f_T$,

$$\sum_{t=1}^T \left\| \pi_t^*(f_{1:t-1}) - \pi_{t-1}^*(f_{1:t-2}) \right\| \leq \tilde{C}_T(f_{1:T}).$$

In this case a simple modification of `AOMD` algorithm where $C_{(N)}$'s are replaced by $\tilde{C}_{\Delta_N}(f_{k_N:k_{N+1}-1})$ leads to the following corollary of Theorem 5.1.

Corollary 5.1. *Assume that $\mathcal{D}_{\mathcal{R}}(x,z) - \mathcal{D}_{\mathcal{R}}(y,z) \leq \gamma\|x-y\|, \forall x, y, z \in \mathcal{X}$. The `AOMD` algorithm with the modification mentioned above achieves the following bound on regret*

$$\mathbf{Reg}_T^{\Pi} \leq \tilde{\mathcal{O}}\left(\sqrt{D_T+1}\right) + \tilde{\mathcal{O}}\left(\min\left\{\sqrt{(D_T+1)\tilde{C}_T(f_{1:T})}, (D_T+1)^{1/3}T^{1/3}V_T^{1/3}\right\}\right).$$

The corollary naturally interpolates between the static and dynamic regret. In other words, letting $\tilde{C}_T(f_{1:T}) = 0$ (which holds for constant mappings), we recover the result of [112] (up to logarithmic factors), whereas $\tilde{C}_T(f_{1:T}) = C_T$ simply recovers the regret bound in Theorem 5.1 corresponding to dynamic regret. The extra log factor is the cost of adaptivity of the algorithm as we assume no prior knowledge about the environment.

5.3.2 Switching Zero-sum Games with Uncoupled Dynamics

Consider two players playing T zero sum games defined by matrices $A_t \in [-1,1]^{m\times n}$ for each $t \in [T]$. We would like to provide strategies for the two players such that, if both players honestly follow the prescribed strategies, the average payoffs of the players approach the average minimax value for the sequence of games at some fast rate. Furthermore, we would also like to guarantee that if one of the players (say the second) deviates from the prescribed strategy, then the first player still has small regret against sequence of actions that do not change drastically. To this end, one can use a simple modification of the `AOMD` algorithm for both players that uses KL divergence as $\mathcal{D}_{\mathcal{R}}$, and mixes in a bit of uniform distribution on each round, producing an algorithm similar to the one in [112] for unchanging uncoupled dynamic games. The following theorem provides bounds for when both players follow the strategy and bound on regret for player I when player II deviates from the strategy.

```
On round $t$, Player I performs

        Play $x_t$ and observe $f_t^\top A_t$

        Update

            $\hat{x}_t(i) \propto \hat{x}'_{t-1}(i) \exp\{-\eta_t [f_t^\top A_t]_i\}$

            $\hat{x}'_t = (1-\beta)\,\hat{x}_t + (\beta/n)\,\mathbf{1}_n$

            $x_{t+1}(i) \propto \hat{x}'_t(i) \exp\{-\eta_{t+1} [f_t^\top A_t]_i\}$

and simultaneously Player II performs

        Play $f_t$ and observe $A_t x_t$

        Update

            $\hat{f}_t(i) \propto \hat{f}'_{t-1}(i) \exp\{-\eta'_t [A_t x_t]_i\}$

            $\hat{f}'_t = (1-\beta)\,\hat{f}_t + (\beta/m)\,\mathbf{1}_m$

            $f_{t+1}(i) \propto \hat{f}'_t(i) \exp\{-\eta'_{t+1} [A_t x_t]_i\}$
```

Note that in the description of the algorithm as well as the following proposition and its proof, any letter with the prime symbol refers to Player II, and it is used to differentiate the letter from its counterpart for player I.

Proposition 5.1. *Define* $\mathscr{F}_t := \sum_{i=1}^{t} \left\| f_i^\top A_i - f_{i-1}^\top A_{i-1} \right\|_\infty^2$*, and let*

$$\eta_t = \min\left\{ \log(T^2 n) \frac{L}{\sqrt{\mathscr{F}_{t-1}} + \sqrt{\mathscr{F}_{t-2}}}, \frac{1}{32L} \right\}.$$

Also define $\mathscr{A}_t := \sum_{i=1}^{t} \left\| A_i x_i - A_{i-1} x_{i-1} \right\|_\infty^2$*, and let*

$$\eta'_t = \min\left\{ \log(T^2 m) \frac{L}{\sqrt{\mathscr{A}_{t-1}} + \sqrt{\mathscr{A}_{t-2}}}, \frac{1}{32L} \right\}.$$

Let $\beta = 1/T^2$*,* $M_t = f_{t-1}^\top A_{t-1}$*, and* $M'_t = A_{t-1} x_{t-1}$*. When Player I uses the prescribed strategy, irrespective of the actions of player II, the regret of Player I w.r.t. any sequence of actions* $u_1, \ldots, u_T$

is bounded as :

$$\sum_{t=1}^{T}\left(f_t^\top A_t x_t - f_t^\top A_t u_t\right) \le 2\log(T^2 n)\left(C_T(u_1,\ldots,u_T)+2\right)\left(32L + \frac{2\sqrt{\mathscr{F}_T}}{\log(T^2 n)L}\right) + \log(T^2 n)\frac{L}{2}\sqrt{\mathscr{F}_T}.$$

Further if both players follow the prescribed strategies then, as long as

$$2L^2 > \max\left\{C_T, C'_T\right\} + 3, \tag{5.3.1}$$

we get,

$$\sum_{t=1}^{T}\sup_{f_t\in\Delta_m} f_t^\top A_t x_t \le \sum_{t=1}^{T}\inf_{x_t\in\Delta_n}\sup_{f_t\in\Delta_m} f_t^\top A_t x_t + \frac{256L}{T} + \frac{1}{2L} + 4\sum_{t=1}^{T}\|A_{t-1}-A_t\|_\infty + 32L\left(\log(T^2 n)C_T + \log(T^2 m)C'_T + 2\log(T^4 nm)\right) + \left(C_T + C'_T + 4\right)\frac{20 + 4\sqrt{\sum_{t=1}^{T}\|A_{t-1}-A_t\|_\infty^2}}{L}.$$

A simple consequence of the above proposition is that if for instance the game matrix A_t changes at most K times over the T rounds, and we knew this fact a priori, then by letting $L = \frac{1}{\sqrt{\log(T^2 n)}}$, we get that regret for Player I w.r.t. any sequence of actions that switches at most K times even when Player II deviates from the prescribed strategy is $\mathcal{O}\left((K+2)\sqrt{\log(T^2 n)T}\right)$. At the same time if both players follow the strategy, then average payoffs of the players converge to the average minimax equilibrium at the rate of $\mathcal{O}\left(L\left(K+2\right)\log(T^4 nm)\right)$ under the condition on L given in (5.3.1). This shows that if the game matrix only changes/switches a constant number of times, then players get $\sqrt{\log(T)T}$ regret bound against arbitrary sequences and comparator actions that switch at most K times while simultaneously get a convergence rate of $\mathcal{O}\left(\log(T)\right)$ to average equilibrium when both players are honest. Also, when we let $K = 0$ and set L to some constant, the proposition recovers the rate in static setting [112] where the matrix sequence is time-invariant.

5.4 Proofs

Proof of Lemma 5.1. For any $u_t \in \mathcal{X}$, it holds that

$$\langle x_t - u_t, \nabla_t \rangle = \langle x_t - \hat{x}_t, \nabla_t - M_t \rangle + \langle x_t - \hat{x}_t, M_t \rangle + \langle \hat{x}_t - u_t, \nabla_t \rangle . \tag{5.4.1}$$

First, observe that for any primal-dual norm pair we have

$$\langle x_t - \hat{x}_t, \nabla_t - M_t \rangle \leq \|x_t - \hat{x}_t\| \, \|\nabla_t - M_t\|_* \ .$$

Any update of the form $a^* = \arg\min_{a \in \mathcal{X}} \langle a, x \rangle + \mathcal{D}_{\mathcal{R}}(a, c)$ satisfies for any $d \in \mathcal{X}$,

$$\langle a^* - d, x \rangle \leq \mathcal{D}_{\mathcal{R}}(d, c) - \mathcal{D}_{\mathcal{R}}(d, a^*) - \mathcal{D}_{\mathcal{R}}(a^*, c) \ .$$

This entails

$$\langle x_t - \hat{x}_t, M_t \rangle \leq \frac{1}{\eta_t} \left\{ \mathcal{D}_{\mathcal{R}}(\hat{x}_t, \hat{x}_{t-1}) - \mathcal{D}_{\mathcal{R}}(\hat{x}_t, x_t) - \mathcal{D}_{\mathcal{R}}(x_t, \hat{x}_{t-1}) \right\}$$

and

$$\langle \hat{x}_t - u_t, \nabla_t \rangle \leq \frac{1}{\eta_t} \left\{ \mathcal{D}_{\mathcal{R}}(u_t, \hat{x}_{t-1}) - \mathcal{D}_{\mathcal{R}}(u_t, \hat{x}_t) - \mathcal{D}_{\mathcal{R}}(\hat{x}_t, \hat{x}_{t-1}) \right\} .$$

Combining the preceding relations and returning to (5.4.1), we obtain

$$\begin{aligned} \langle x_t - u_t, \nabla_t \rangle &\leq \frac{1}{\eta_t} \left\{ \mathcal{D}_{\mathcal{R}}(u_t, \hat{x}_{t-1}) - \mathcal{D}_{\mathcal{R}}(u_t, \hat{x}_t) - \mathcal{D}_{\mathcal{R}}(\hat{x}_t, x_t) - \mathcal{D}_{\mathcal{R}}(x_t, \hat{x}_{t-1}) \right\} \\ &\qquad + \|\nabla_t - M_t\|_* \, \|x_t - \hat{x}_t\| \\ &\leq \frac{1}{\eta_t} \left\{ \mathcal{D}_{\mathcal{R}}(u_t, \hat{x}_{t-1}) - \mathcal{D}_{\mathcal{R}}(u_t, \hat{x}_t) - \frac{1}{2} \|\hat{x}_t - x_t\|^2 - \frac{1}{2} \|\hat{x}_{t-1} - x_t\|^2 \right\} \\ &\qquad + \|\nabla_t - M_t\|_* \, \|x_t - \hat{x}_t\| , \end{aligned} \tag{5.4.2}$$

where in the last step we appealed to strong convexity: $\mathcal{D}_{\mathcal{R}}(x, y) \geq \frac{1}{2}\|x - y\|^2$ for any $x, y \in \mathcal{X}$.

Using the simple inequality $ab \leq \frac{\rho a^2}{2} + \frac{b^2}{2\rho}$ for any $\rho > 0$ to split the product term, we get

$$\begin{aligned}\langle x_t - u_t, \nabla_t \rangle \leq \frac{1}{\eta_t}\Big\{ \mathcal{D}_{\mathcal{R}}(u_t, \hat{x}_{t-1}) - \mathcal{D}_{\mathcal{R}}(u_t, \hat{x}_t) - \frac{1}{2}\|\hat{x}_t - x_t\|^2 - \frac{1}{2}\|\hat{x}_{t-1} - x_t\|^2 \Big\} \\ + \frac{\eta_{t+1}}{2}\|\nabla_t - M_t\|_*^2 + \frac{1}{2\eta_{t+1}}\|x_t - \hat{x}_t\|^2,\end{aligned}$$

Applying the bound

$$\frac{1}{2\eta_{t+1}}\|x_t - \hat{x}_t\|^2 - \frac{1}{2\eta_t}\|x_t - \hat{x}_t\|^2 \leq R_{\max}^2\left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right),$$

and summing over $t \in [T]$ yields ,

$$\begin{aligned}\sum_{t=1}^T \langle x_t - u_t, \nabla_t \rangle &\leq \sum_{t=1}^T \frac{\eta_{t+1}}{2}\|\nabla_t - M_t\|_*^2 + \sum_{t=1}^T \frac{1}{\eta_t}\Big\{ \mathcal{D}_{\mathcal{R}}(u_t, \hat{x}_{t-1}) - \mathcal{D}_{\mathcal{R}}(u_t, \hat{x}_t) \Big\} + \frac{R_{\max}^2}{\eta_{T+1}} \\ &\leq \sum_{t=1}^T \frac{\eta_{t+1}}{2}\|\nabla_t - M_t\|_*^2 + R_{\max}^2\left(\frac{1}{\eta_1} + \frac{1}{\eta_{T+1}}\right) \\ &\qquad + \sum_{t=2}^T \left\{ \frac{\mathcal{D}_{\mathcal{R}}(u_t, \hat{x}_{t-1})}{\eta_t} - \frac{\mathcal{D}_{\mathcal{R}}(u_{t-1}, \hat{x}_{t-1})}{\eta_{t-1}} \right\} \\ &\leq \sum_{t=2}^T \left\{ \frac{\mathcal{D}_{\mathcal{R}}(u_t, \hat{x}_{t-1})}{\eta_t} - \frac{\mathcal{D}_{\mathcal{R}}(u_{t-1}, \hat{x}_{t-1})}{\eta_t} \right\} \\ &\qquad + \sum_{t=2}^T \left\{ \frac{\mathcal{D}_{\mathcal{R}}(u_{t-1}, \hat{x}_{t-1})}{\eta_t} - \frac{\mathcal{D}_{\mathcal{R}}(u_{t-1}, \hat{x}_{t-1})}{\eta_{t-1}} \right\} \\ &\qquad + \sum_{t=1}^T \frac{\eta_{t+1}}{2}\|\nabla_t - M_t\|_*^2 + \frac{2R_{\max}^2}{\eta_{T+1}} \\ &\leq \sum_{t=1}^T \frac{\eta_{t+1}}{2}\|\nabla_t - M_t\|_*^2 + \gamma\sum_{t=2}^T \frac{\|u_t - u_{t-1}\|}{\eta_t} \\ &\qquad + \sum_{t=2}^T \mathcal{D}_{\mathcal{R}}(u_{t-1}, \hat{x}_{t-1})\left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}\right) + \frac{2R_{\max}^2}{\eta_{T+1}} \\ &\leq \sum_{t=1}^T \frac{\eta_{t+1}}{2}\|\nabla_t - M_t\|_*^2 + \gamma\sum_{t=2}^T \frac{\|u_t - u_{t-1}\|}{\eta_t} + \frac{4R_{\max}^2}{\eta_{T+1}},\end{aligned}$$

where we used the Lipschitz continuity of $\mathcal{D}_{\mathcal{R}}$ in the penultimate step. Now let us set

$$\begin{aligned}\eta_t &= \frac{L}{\sqrt{\sum_{s=0}^{t-1}\|\nabla_s - M_s\|_*^2} + \sqrt{\sum_{s=0}^{t-2}\|\nabla_s - M_s\|_*^2}}\\ &= \frac{L\left(\sqrt{\sum_{s=0}^{t-1}\|\nabla_s - M_s\|_*^2} - \sqrt{\sum_{s=0}^{t-2}\|\nabla_s - M_s\|_*^2}\right)}{\|\nabla_{t-1} - M_{t-1}\|_*^2},\end{aligned}$$

and $\|\nabla_0 - M_0\|_*^2 = 1$ to have

$$\begin{aligned}\sum_{t=1}^{T}\langle x_t - u_t, \nabla_t\rangle &\leq \frac{L}{2}\sum_{t=1}^{T}\left\{\sqrt{\sum_{s=0}^{t}\|\nabla_s - M_s\|_*^2} - \sqrt{\sum_{s=0}^{t-1}\|\nabla_s - M_s\|_*^2}\right\}\\ &\quad + \frac{2\gamma\sqrt{1+\sum_{t=1}^{T}\|\nabla_t - M_t\|_*^2}}{L}\sum_{t=2}^{T}\|u_t - u_{t-1}\|\\ &\quad + \frac{8R_{\max}^2\sqrt{1+\sum_{t=1}^{T}\|\nabla_t - M_t\|_*^2}}{L}\\ &\leq 2\sqrt{1+\sum_{t=1}^{T}\|\nabla_t - M_t\|_*^2}\left(L + \frac{\gamma\sum_{t=1}^{T}\|u_t - u_{t-1}\| + 4R_{\max}^2}{L}\right).\end{aligned}$$

Appealing to convexity of $\{f_t\}_{t=1}^T$, and replacing C_T (5.0.3) and D_T (5.0.4) in above, completes the proof . ■

Proof of Lemma 5.2. We define

$$U_T := \left\{u_1, ..., u_T \in \mathcal{X} : \gamma\sum_{t=1}^{T}\|u_t - u_{t-1}\| \leq L^2 - 4R_{\max}^2\right\}, \tag{5.4.3}$$

and

$$(u_1^*, ..., u_T^*) := \text{argmin}_{u_1,...,u_T \in U_T}\sum_{t=1}^{T} f_t(u_t).$$

Our choice of $L > 2R_{\max}$ guarantees that any sequence of fixed comparators $u_t = u$ for $t \in [T]$ belongs to U_T, and hence, $(u_1^*, ..., u_T^*)$ exists. Noting that $(u_1^*, ..., u_T^*)$ is an element of U_T, we have

$\gamma \sum_{t=1}^{T} \left\|u_t^* - u_{t-1}^*\right\| + 4R_{\max}^2 \leq L^2$. We now apply Lemma 5.1 to $\{u_t^*\}_{t=1}^T$ to bound the dynamic regret for arbitrary comparator sequence $\{u_t\}_{t=1}^T$ as follows,

$$\begin{aligned}
\mathbf{Reg}_T^d(u_1, ..., u_T) &= \sum_{t=1}^{T} \left\{ f_t(x_t) - f_t(u_t^*) \right\} + \sum_{t=1}^{T} \left\{ f_t(u_t^*) - f_t(u_t) \right\} \\
&\leq 4\sqrt{1 + D_T} L + \sum_{t=1}^{T} \left\{ f_t(u_t^*) - f_t(u_t) \right\} \\
&\leq 4\sqrt{1 + D_T} L \\
&+ \mathbf{1}\left\{ \gamma \sum_{t=1}^{T} \|u_t - u_{t-1}\| > L^2 - 4R_{\max}^2 \right\} \left(\sum_{t=1}^{T} \left\{ f_t(u_t^*) - f_t(u_t) \right\} \right),
\end{aligned} \tag{5.4.4}$$

where the last step follows from the fact that

$$\sum_{t=1}^{T} f_t(u_t^*) - \sum_{t=1}^{T} f_t(u_t) \leq 0 \quad \text{if} \quad (u_1, ..., u_T) \in U_T.$$

Given the definition of $R_{\max}^2$, by strong convexity of $\mathcal{D}_{\mathcal{R}}(x, y)$, we get that $\|x - y\| \leq \sqrt{2} R_{\max}$, for any $x, y \in \mathcal{X}$. This entails that once we divide the horizon into B number of batches and use a single, fixed point as a comparator along each batch, we have

$$\sum_{t=1}^{T} \|u_t - u_{t-1}\| \leq B\sqrt{2} R_{\max}, \tag{5.4.5}$$

since there are at most B number of changes in the comparator sequence along the horizon. Now let $B = \frac{L^2 - 4R_{\max}^2}{\gamma\sqrt{2}R_{\max}}$, and for ease of notation, assume that T is divisible by B. Noting that $f_t(x_t^*) \leq f_t(u_t)$, we use an argument similar to that of [116] to get for any fixed $t_i \in [(i-1)(T/B) +$

$1, i(T/B)]$,

$$\begin{aligned}\sum_{t=1}^{T}\Big\{f_t(u_t^*) - f_t(u_t)\Big\} &\le \sum_{t=1}^{T}\Big\{f_t(u_t^*) - f_t(x_t^*)\Big\} \tag{5.4.6}\\ &= \sum_{i=1}^{B}\sum_{t=(i-1)(T/B)+1}^{i(T/B)}\Big\{f_t(u_t^*) - f_t(x_t^*)\Big\} \notag\\ &\le \sum_{i=1}^{B}\sum_{t=(i-1)(T/B)+1}^{i(T/B)}\Big\{f_t(x_{t_i}^*) - f_t(x_t^*)\Big\} \tag{5.4.7}\\ &\le \left(\frac{T}{B}\right)\sum_{i=1}^{B}\max_{t\in[(i-1)(T/B)+1, i(T/B)]}\Big\{f_t(x_{t_i}^*) - f_t(x_t^*)\Big\}. \tag{5.4.8}\end{aligned}$$

Note that $x_{t_i}^*$ is fixed for each batch i. Substituting our choice of $B = \frac{L^2 - 4R_{\max}^2}{\gamma\sqrt{2}R_{\max}}$ in (5.4.5) implies that the comparator sequence $u_t = x_{t_i}^* \mathbf{1}\left\{\frac{(i-1)T}{B} + 1 \le t \le \frac{iT}{B}\right\}$ belongs to U_T, and (5.4.7) follows by optimality of $(u_1^*, ..., u_T^*)$. We now claim that for any $t \in [(i-1)(T/B)+1, i(T/B)]$, we have,

$$f_t(x_{t_i}^*) - f_t(x_t^*) \le 2\sum_{s=(i-1)(T/B)+1}^{i(T/B)}\sup_{x\in\mathcal{X}}|f_s(x) - f_{s-1}(x)|. \tag{5.4.9}$$

Assuming otherwise, there must exist a $\hat{t}_i \in [(i-1)(T/B)+1, i(T/B)]$ such that

$$f_{\hat{t}_i}(x_{t_i}^*) - f_{\hat{t}_i}(x_{\hat{t}_i}^*) > 2\sum_{t=(i-1)(T/B)+1}^{i(T/B)}\sup_{x\in\mathcal{X}}|f_t(x) - f_{t-1}(x)|,$$

which results in

$$\begin{aligned}f_t(x_{\hat{t}_i}^*) &\le f_{\hat{t}_i}(x_{\hat{t}_i}^*) + \sum_{t=(i-1)(T/B)+1}^{i(T/B)}\sup_{x\in\mathcal{X}}|f_t(x) - f_{t-1}(x)|\\ &< f_{\hat{t}_i}(x_{t_i}^*) - \sum_{t=(i-1)(T/B)+1}^{i(T/B)}\sup_{x\in\mathcal{X}}|f_t(x) - f_{t-1}(x)| \le f_t(x_{t_i}^*),\end{aligned}$$

The preceding relation for $t = t_i$ violates the optimality of $x_{t_i}^*$, which is a contradiction. Therefore, Equation (5.4.9) holds for any $t \in [(i-1)(T/B)+1, i(T/B)]$ Combining (5.4.6), (5.4.8) and

(5.4.9) we have

$$\sum_{t=1}^{T}\left\{f_t(u_t^*)-f_t(u_t)\right\} \leq \frac{2T}{B}\sum_{i=1}^{B}\sum_{t=(i-1)(T/B)+1}^{i(T/B)}\sup_{x\in\mathcal{X}}|f_t(x)-f_{t-1}(x)|$$

$$= \frac{2TV_T}{B} = \frac{2\gamma\sqrt{2}R_{\max}TV_T}{L^2-4R_{\max}^2}. \tag{5.4.10}$$

Using the above in Equation (5.4.4) we conclude the following upper bound

$$\begin{aligned}\mathbf{Reg}_T^d(u_1,...,u_T) \leq{}& 4\sqrt{1+D_T}L \\ &+\mathbf{1}\left\{\gamma\sum_{t=1}^{T}\|u_t-u_{t-1}\| > L^2-4R_{\max}^2\right\}\frac{4\gamma R_{\max}TV_T}{L^2-4R_{\max}^2},\end{aligned}$$

thereby completing the proof. ■

Proof of Theorem 5.1. For the sake of clarity in presentation, we stick to the following notation for the proof

$$\underline{D}_{(i)} := D_{(i)} - \|\nabla_{k_{i+1}-1} - M_{k_{i+1}-1}\|_*^2$$

$$\underline{C}_{(i)} := C_{(i)} - \|x^*_{k_{i+1}-1} - x^*_{k_{i+1}-2}\|$$

$$\underline{V}_{(i)} := V_{(i)} - \sup_{x\in\mathcal{X}}\left|f_{k_{i+1}-1}(x) - f_{k_{i+1}-2}(x)\right|$$

$$\underline{\Delta}_{(i)} := \Delta_i - 1,$$

for any doubling epoch $i = 1, ..., N$, where we recall that $k_{i+1}-1$ is the last instance of epoch i. Therefore, any symbol with lower bar refers to its corresponding quantity removing only the value of the last instance of that interval.

Let the **AOMD** algorithm run with the step size given by Lemma 5.1 in the following form

$$\eta_t = \frac{L_i}{\sqrt{\sum_{s=0}^{t-1}\|\nabla_s - M_s\|_*^2} + \sqrt{\sum_{s=0}^{t-2}\|\nabla_s - M_s\|_*^2}},$$

and let L_i be tuned with a doubling condition explained in the algorithm. Once the condition stated in the algorithm fails, the following pair of identities must hold

$$\begin{aligned}
&\gamma \min\{\underline{C}_{(i)}\ ,\ \underline{\Delta}_i^{2/3} \underline{V}_{(i)}^{2/3} \underline{D}_{(i)}^{-1/3}\} + 4R_{\max}^2 \leq L_i^2 \\
&\gamma \min\{C_{(i)}\ ,\ \Delta_i^{2/3} V_{(i)}^{2/3} D_{(i)}^{-1/3}\} + 4R_{\max}^2 > L_i^2. \qquad (5.4.11)
\end{aligned}$$

Observe that the algorithm doubles L_i only after the condition fails, so at violation points we suffer at most $2G$ by boundedness (5.1.1). Then, under purview of Lemma 5.2, it holds that

$$\begin{aligned}
\mathbf{Reg}_T^d &\leq \sum_{i=1}^N \left\{ 4\sqrt{\underline{D}_{(i)}} L_i + \mathbf{1}\left\{\gamma \underline{C}_{(i)} > L_i^2 - 4R_{\max}^2\right\} \frac{4\gamma R_{\max} \underline{\Delta}_i \underline{V}_{(i)}}{L_i^2 - 4R_{\max}^2} \right\} + 2NG \\
&\leq \sum_{i=1}^N \left\{ 4\sqrt{D_{(i)}} L_i + \mathbf{1}\left\{\underline{C}_{(i)} > \underline{\Delta}_i^{2/3} \underline{V}_{(i)}^{2/3} \underline{D}_{(i)}^{-1/3}\right\} \frac{4\gamma R_{\max} \underline{\Delta}_i \underline{V}_{(i)}}{L_i^2 - 4R_{\max}^2} \right\} + 2NG, \quad (5.4.12)
\end{aligned}$$

where the last step follows directly from (5.4.11) and the fact that $\underline{D}_{(i)} \leq D_{(i)}$. Bounding $\sqrt{D_{(i)}} L_i$ in above, using the second inequality in (5.4.11), we get

$$\begin{aligned}
\sqrt{D_{(i)}} L_i &\leq \sqrt{\gamma \min\left\{D_{(i)} C_{(i)}\ ,\ \Delta_i^{2/3} V_{(i)}^{2/3} D_{(i)}^{2/3}\right\} + 4R_{\max}^2 D_{(i)}} \\
&\leq 2R_{\max}\sqrt{D_{(i)}} + \sqrt{\gamma} \min\left\{\sqrt{D_{(i)} C_{(i)}}\ ,\ \Delta_i^{1/3} V_{(i)}^{1/3} D_{(i)}^{1/3}\right\},
\end{aligned}$$

by the simple inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$. Plugging the bound above into (5.4.12) and noting that

$$\sum_{i=1}^N \sqrt{D_{(i)}} = N \sum_{i=1}^N \frac{1}{N} \sqrt{D_{(i)}} \leq N \sqrt{\frac{1}{N} \sum_{i=1}^N D_{(i)}} = \sqrt{N D_T + N},$$

by Jensen's inequality, we obtain

$$\begin{aligned}
\mathbf{Reg}_T^d \leq 2NG + 8R_{\max}\sqrt{N D_T + N} + 4\sqrt{\gamma} \sum_{i=1}^N \min\left\{\sqrt{D_{(i)} C_{(i)}}\ ,\ D_{(i)}^{1/3} \Delta_i^{1/3} V_{(i)}^{1/3}\right\} \\
+ \sum_{i=1}^N \mathbf{1}\left\{\underline{C}_{(i)} > \underline{\Delta}_i^{2/3} \underline{V}_{(i)}^{2/3} \underline{D}_{(i)}^{-1/3}\right\} \frac{4R_{\max} \underline{\Delta}_i \underline{V}_{(i)}}{\min\left\{\underline{C}_{(i)}, \underline{\Delta}_i^{2/3} \underline{V}_{(i)}^{2/3} \underline{D}_{(i)}^{-1/3}\right\}},
\end{aligned}$$

where we used the first inequality in (5.4.11) to bound the last term. Given the condition in the indicator function $\mathbf{1}\{\cdot\}$, we can simplify above to derive,

$$\begin{aligned}
\mathbf{Reg}_T^d &\leq 2NG + 8R_{\max}\sqrt{ND_T + N} + 4\sqrt{\gamma}\sum_{i=1}^{N} \min\left\{\sqrt{D_{(i)}C_{(i)}}\ ,\ D_{(i)}^{1/3}\Delta_i^{1/3}V_{(i)}^{1/3}\right\} \\
&\qquad + 4R_{\max}\sum_{i=1}^{N} \mathbf{1}\left\{\underline{C}_{(i)} > \underline{\Delta}_i^{2/3}\,\underline{V}_{(i)}^{2/3}\,\underline{D}_{(i)}^{-1/3}\right\}\underline{D}_{(i)}^{1/3}\,\underline{V}_{(i)}^{1/3}\,\underline{\Delta}_i^{1/3} \\
&= 2NG + 8R_{\max}\sqrt{ND_T + N} + 4\sqrt{\gamma}\sum_{i=1}^{N} \min\left\{\sqrt{D_{(i)}C_{(i)}}\ ,\ D_{(i)}^{1/3}\Delta_i^{1/3}V_{(i)}^{1/3}\right\} \\
&\qquad + 4R_{\max}\sum_{i=1}^{N} \mathbf{1}\left\{\sqrt{\underline{D}_{(i)}\underline{C}_{(i)}} > \underline{\Delta}_i^{1/3}\,\underline{V}_{(i)}^{1/3}\,\underline{D}_{(i)}^{1/3}\right\}\underline{D}_{(i)}^{1/3}\,\underline{V}_{(i)}^{1/3}\,\underline{\Delta}_i^{1/3} \\
&\leq 2NG + 8R_{\max}\sqrt{ND_T + N} + 4\sqrt{\gamma}\sum_{i=1}^{N} \min\left\{\sqrt{D_{(i)}C_{(i)}}\ ,\ D_{(i)}^{1/3}\Delta_i^{1/3}V_{(i)}^{1/3}\right\} \\
&\qquad + 4R_{\max}\sum_{i=1}^{N} \min\left\{\sqrt{\underline{D}_{(i)}\underline{C}_{(i)}}, \underline{D}_{(i)}^{1/3}\,\underline{V}_{(i)}^{1/3}\,\underline{\Delta}_i^{1/3}\right\}. \qquad (5.4.13)
\end{aligned}$$

Let $\ell := 2\sqrt{\gamma} + 2R_{\max}$. Given the fact that

$$\underline{C}_{(i)} \leq C_{(i)} \qquad \underline{D}_{(i)} \leq D_{(i)} \qquad \underline{V}_{(i)} \leq V_{(i)} \qquad \underline{\Delta}_i \leq \Delta_i,$$

we return to (5.4.13) to derive

$$\begin{aligned}
\mathbf{Reg}_T^d &\leq 2NG + 8R_{\max}\sqrt{ND_T + N} + 2\ell\sum_{i=1}^{N} \min\left\{\sqrt{D_{(i)}C_{(i)}}\ ,\ D_{(i)}^{1/3}\Delta_i^{1/3}V_{(i)}^{1/3}\right\} \\
&\leq 2NG + 8R_{\max}\sqrt{ND_T + N} + 2\ell\min\left\{\sum_{i=1}^{N}\sqrt{D_{(i)}C_{(i)}}, \sum_{i=1}^{N} D_{(i)}^{1/3}\Delta_i^{1/3}V_{(i)}^{1/3}\right\} \\
&\leq 2N\left(G + 4R_{\max}\sqrt{D_T + 1} + \ell\min\left\{\sqrt{(D_T+1)C_T}, (D_T+1)^{1/3}T^{1/3}V_T^{1/3}\right\}\right). \qquad (5.4.14)
\end{aligned}$$

where we bounded the sums using the following fact about the summands

$$C_{(i)} \leq C_T \qquad D_{(i)} \leq D_T + 1 \qquad V_{(i)} \leq V_T \qquad \Delta_i \leq T.$$

To bound the number of batches N, we recall that $L_i = 3R_{\max}2^{i-1}$, and use the second inequality in (5.4.11) to bound L_{N-1} as follows

$$\begin{aligned}
N &= 2 + \log_2(2^{N-2}) = 2 + \log_2(L_{N-1}) - \log_2(3R_{\max}) \\
&\leq 2 + \frac{1}{2}\log_2\left(\gamma \min\left\{C_{(N-1)}, \Delta_{N-1}^{2/3} V_{(N-1)}^{2/3} D_{(N-1)}^{-1/3}\right\} + 4R_{\max}^2\right) - \log_2(3R_{\max}) \\
&\leq 2 + \frac{1}{2}\log_2\left(\gamma C_{(N-1)} + 4R_{\max}^2\right) - \log_2(3R_{\max}) \\
&\leq 2 + \frac{1}{2}\log_2\left(2\gamma R_{\max} T + 4R_{\max}^2\right) - \log_2(3R_{\max}).
\end{aligned}$$

In view of the preceding relation and (5.4.14), we have

$$\mathbf{Reg}_T^d \leq \kappa\Bigg(G + 4R_{\max}\sqrt{D_T + 1} + \ell \min\left\{\sqrt{(D_T+1)C_T}, (D_T+1)^{1/3} T^{1/3} V_T^{1/3}\right\}\Bigg),$$

where $\kappa := 4 + \log_2\left(2\gamma R_{\max} T + 4R_{\max}^2\right) - 2\log_2(3R_{\max})$, thereby completing the proof. ■

Proof of Proposition 5.1. Assume that the player I uses the prescribed strategy. This corresponds to using the optimistic mirror descent update with $\mathcal{R}(x) = \sum_{i=1}^n x_i \log(x_i)$ as the function that is strongly convex w.r.t. $\|\cdot\|_1$. Correspondingly, $\nabla_t = f_t^\top A_t$ and $M_t = f_{t-1}^\top A_{t-1}$. Following the line of proof in Lemma 5.1, in particular, using Equation 5.4.2 for the specific case with $\mathcal{D}_{\mathcal{R}}$ as KL divergence, we get that for any t and any $u_t \in \Delta_n$,

$$\begin{aligned}
f_t^\top A_t x_t - f_t^\top A_t u_t &\leq \frac{1}{\eta_t}\Bigg\{\sum_{i=1}^n u_t[i]\log\left(\frac{\hat{x}_t[i]}{\hat{x}'_{t-1}[i]}\right) - \frac{1}{2}\left\|\hat{x}_t - x_t\right\|_1^2 - \frac{1}{2}\left\|\hat{x}'_{t-1} - x_t\right\|_1^2\Bigg\} \\
&\qquad + \left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty \left\|x_t - \hat{x}_t\right\|_1 \\
&\leq \frac{1}{\eta_t}\Bigg\{\sum_{i=1}^n u_t[i]\log\left(\frac{\hat{x}'_t[i]}{\hat{x}'_{t-1}[i]}\right) - \frac{1}{2}\left\|\hat{x}_t - x_t\right\|_1^2 - \frac{1}{2}\left\|\hat{x}'_{t-1} - x_t\right\|_1^2\Bigg\} \\
&\qquad + \left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty \left\|x_t - \hat{x}_t\right\|_1 + \frac{1}{\eta_t}\max_{i\in[n]}\log\left(\frac{\hat{x}_t[i]}{\hat{x}'_t[i]}\right).
\end{aligned}$$

Now let us bound for some i the term, $\log\left(\frac{\hat{x}_t[i]}{\hat{x}'_t[i]}\right)$. Notice that if $\hat{x}_t[i] \leq \hat{x}'_t[i]$ then the term is anyway bounded by 0. Now assume $\hat{x}_t[i] > \hat{x}'_t[i]$. Letting $\beta = 1/T^2$, since $\hat{x}'_t[i] = (1 - T^{-2})\hat{x}_t[i] +$

$1/(nT^2)$, we can have $\hat{x}_t[i] > \hat{x}'_t[i]$ only when $\hat{x}_t[i] > 1/n$. Hence,

$$\log\left(\frac{\hat{x}_t[i]}{\hat{x}'_t[i]}\right) = \log\left(\frac{\hat{x}_t[i]}{(1-T^{-2})\hat{x}_t[i] + 1/(nT^2)}\right) \le \frac{2}{T^2}.$$

Using this we can conclude that :

$$f_t^\top A_t x_t - f_t^\top A_t u_t \le \frac{1}{\eta_t}\left\{\sum_{i=1}^n u_t[i]\log\left(\frac{\hat{x}'_t[i]}{\hat{x}'_{t-1}[i]}\right) - \frac{1}{2}\left\|\hat{x}_t - x_t\right\|_1^2 - \frac{1}{2}\left\|\hat{x}'_{t-1} - x_t\right\|_1^2\right\}$$
$$+ \left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty \left\|x_t - \hat{x}_t\right\|_1 + \frac{2}{T^2}\frac{1}{\eta_t}.$$

Summing over $t \in [T]$ we obtain that :

$$\sum_{t=1}^T \left(f_t^\top A_t x_t - f_t^\top A_t u_t\right) \le \sum_{t=1}^T \frac{1}{\eta_t}\left\{\sum_{i=1}^n u_t[i]\log\left(\frac{\hat{x}'_t[i]}{\hat{x}'_{t-1}[i]}\right) - \frac{1}{2}\left\|\hat{x}_t - x_t\right\|_1^2 - \frac{1}{2}\left\|\hat{x}'_{t-1} - x_t\right\|_1^2\right\}$$
$$+ \sum_{t=1}^T \left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty \left\|x_t - \hat{x}_t\right\|_1 + \frac{2}{T^2}\sum_{t=1}^T \frac{1}{\eta_t}.$$

Note that $\frac{1}{\eta_t} \le \mathcal{O}\left(\sqrt{T}\right)$ and so assuming T is large enough, $\frac{1}{T^2}\sum_{t=1}^T \frac{1}{\eta_t} \le 1$ and so,

$$\sum_{t=1}^T \left(f_t^\top A_t x_t - f_t^\top A_t u_t\right) \le \sum_{t=1}^T \frac{1}{\eta_t}\left\{\sum_{i=1}^n u_t[i]\log\left(\frac{\hat{x}'_t[i]}{\hat{x}'_{t-1}[i]}\right) - \frac{1}{2}\left\|\hat{x}_t - x_t\right\|_1^2 - \frac{1}{2}\left\|\hat{x}'_{t-1} - x_t\right\|_1^2\right\}$$
$$+ \sum_{t=1}^T \left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty \left\|x_t - \hat{x}_t\right\|_1 + 1. \tag{5.4.15}$$

Now note that we can rewrite the first sum in the above bound and get :

$$\sum_{t=1}^T \frac{1}{\eta_t}\sum_{i=1}^n u_t[i]\log\left(\frac{\hat{x}'_t[i]}{\hat{x}'_{t-1}[i]}\right) \le \sum_{t=2}^T \frac{\sum_{i=1}^n u_t[i]\log\left(\frac{1}{\hat{x}'_{t-1}[i]}\right)}{\eta_t} - \frac{\sum_{i=1}^n u_{t-1}[i]\log\left(\frac{1}{\hat{x}'_{t-1}[i]}\right)}{\eta_{t-1}}$$
$$+ \frac{\log(T^2 n)}{\eta_1}$$
$$\le \sum_{t=2}^T \frac{\sum_{i=1}^n \left(u_t[i] - u_{t-1}[i]\right)\log\left(\frac{1}{\hat{x}'_{t-1}[i]}\right)}{\eta_t}$$
$$+ \sum_{t=2}^T \sum_{i=1}^n u_{t-1}[i]\log\left(\frac{1}{\hat{x}'_{t-1}[i]}\right)\left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}\right) + \frac{\log(T^2 n)}{\eta_1}.$$

Since by definition of $\hat{x}'_{t-1}$, we are mixing in $1/T^2$ of the uniform distribution we have that for any i, $\hat{x}'_{t-1}[i] > \frac{1}{T^2 n}$ and, since η_t's are non-increasing, we continue bounding above as

$$\begin{aligned}
\sum_{t=1}^{T} \frac{1}{\eta_t} \sum_{i=1}^{n} u_t[i] \log\left(\frac{\hat{x}'_t[i]}{\hat{x}'_{t-1}[i]}\right) &\leq \log(T^2 n) \sum_{t=2}^{T} \frac{\|u_{t-1} - u_t\|_1}{\eta_t} \\
&\quad + \log(T^2 n) \sum_{t=2}^{T} \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}\right) + \frac{\log(T^2 n)}{\eta_1} \\
&\leq \log(T^2 n) \left(\sum_{t=2}^{T} \frac{\|u_{t-1} - u_t\|_1}{\eta_t} + \frac{1}{\eta_T} - \frac{1}{\eta_1}\right) + \frac{\log(T^2 n)}{\eta_1} \\
&\leq \log(T^2 n) \left(\sum_{t=2}^{T} \frac{\|u_{t-1} - u_t\|_1}{\eta_t} + \frac{1}{\eta_T}\right),
\end{aligned}$$

using the above in Equation 5.4.15 we get

$$\begin{aligned}
&\sum_{t=1}^{T} f_t^\top A_t x_t - f_t^\top A_t u_t \\
&\qquad \leq \log(T^2 n) \sum_{t=2}^{T} \frac{\|u_{t-1} - u_t\|_1}{\eta_t} - \frac{1}{2} \sum_{t=1}^{T} \frac{1}{\eta_t} \|\hat{x}_t - x_t\|_1^2 - \frac{1}{2} \sum_{t=1}^{T} \frac{1}{\eta_t} \left\|\hat{x}'_{t-1} - x_t\right\|_1^2 + 1 \\
&\qquad\quad + \sum_{t=1}^{T} \left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty \|x_t - \hat{x}_t\|_1 + \frac{\log(T^2 n)}{\eta_T} \\
&\qquad \leq \frac{\log(T^2 n)\left(C_T(u_1, \ldots, u_T) + 2\right)}{\eta_T} - \frac{1}{2} \sum_{t=1}^{T} \frac{1}{\eta_t} \|\hat{x}_t - x_t\|_1^2 - \frac{1}{2} \sum_{t=1}^{T} \frac{1}{\eta_t} \left\|\hat{x}'_{t-1} - x_t\right\|_1^2 \\
&\qquad\quad + \sum_{t=1}^{T} \left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty \|x_t - \hat{x}_t\|_1 . \qquad (5.4.16)
\end{aligned}$$

Notice that our choice of step size given by,

$$\begin{aligned}
\eta_t &= \min\left\{\log(T^2 n) \frac{L}{\sqrt{\sum_{i=1}^{t-1} \left\|f_i^\top A_i - f_{i-1}^\top A_{i-1}\right\|_\infty^2} + \sqrt{\sum_{i=1}^{t-2} \left\|f_i^\top A_i - f_{i-1}^\top A_{i-1}\right\|_\infty^2}}, \frac{1}{32L}\right\} \\
&= \min\left\{\log(T^2 n) \frac{L\left(\sqrt{\sum_{i=1}^{t-1} \left\|f_i^\top A_i - f_{i-1}^\top A_{i-1}\right\|_\infty^2} - \sqrt{\sum_{i=1}^{t-2} \left\|f_i^\top A_i - f_{i-1}^\top A_{i-1}\right\|_\infty^2}\right)}{\left\|f_{t-1}^\top A_{t-1} - f_{t-2}^\top A_{t-2}\right\|_\infty^2}, \frac{1}{32L}\right\}, \qquad (5.4.17)
\end{aligned}$$

guarantees that

$$\eta_t^{-1} = \max\left\{\frac{\sqrt{\sum_{i=1}^{t-1} \left\|f_i^\top A_i - f_{i-1}^\top A_{i-1}\right\|_\infty^2} + \sqrt{\sum_{i=1}^{t-2} \left\|f_i^\top A_i - f_{i-1}^\top A_{i-1}\right\|_\infty^2}}{\log(T^2 n) L}, 32L\right\}.$$

Using the step-size specified above in the bound 5.4.16, we get

$$\begin{aligned}
&\sum_{t=1}^{T} f_t^\top A_t x_t - \sum_{t=1}^{T} f_t^\top A_t u_t \\
&\qquad \leq \log(T^2 n)\left(C_T(u_1,\dots,u_T)+2\right)\left(\frac{2\sqrt{\sum_{t=1}^{T}\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty^2}}{\log(T^2 n)L}+32L\right) \\
&\qquad\qquad + \sum_{t=1}^{T}\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty \left\|x_t-\hat{x}_t\right\|_1 \\
&\qquad\qquad - 16L\sum_{t=1}^{T}\left\|\hat{x}_t - x_t\right\|_1^2 - 16L\sum_{t=1}^{T}\left\|\hat{x}'_{t-1}-x_t\right\|_1^2 . \qquad (5.4.18)
\end{aligned}$$

Now note that by triangle inequality, we have

$$\begin{aligned}
\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty &= \left\|f_t^\top A_t - f_t^\top A_{t-1} + f_t^\top A_{t-1} - f_{t-1}^\top A_{t-1}\right\|_\infty \\
&\leq \left\|A_{t-1}-A_t\right\|_\infty + \left\|f_t - f_{t-1}\right\|_1 \\
&\leq \left\|A_{t-1}-A_t\right\|_\infty + \left\|f_t - \hat{f}_{t-1}\right\|_1 + \left\|\hat{f}_{t-1}-f_{t-1}\right\|_1 ,
\end{aligned}$$

since the entries of matrix sequence $\{A_t\}_{t=1}^T$ are bounded by one. Using the bound above in (5.4.18) and splitting the product term, we see that

$$\begin{aligned}
&\sum_{t=1}^{T} f_t^\top A_t x_t - f_t^\top A_t u_t \\
&\qquad \leq \log(T^2 n)\left(C_T(u_1,\dots,u_T)+2\right)\left(\frac{2\sqrt{\sum_{t=1}^{T}\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty^2}}{\log(T^2 n)L}+32L\right) \\
&\qquad + 2\sum_{t=1}^{T}\left\|A_t - A_{t-1}\right\|_\infty - 8L\sum_{t=1}^{T}\left\|\hat{x}_t - x_t\right\|_1^2 - 16L\sum_{t=1}^{T}\left\|\hat{x}'_{t-1}-x_t\right\|_1^2 \\
&\qquad + \frac{1}{16L}\sum_{t=1}^{T}\left\|f_t - \hat{f}_{t-1}\right\|_1^2 + \frac{1}{16L}\sum_{t=1}^{T}\left\|\hat{f}_{t-1}-f_{t-1}\right\|_1^2 , \qquad (5.4.19)
\end{aligned}$$

where we used the simple inequality $ab \leq \frac{\rho}{2}a^2 + \frac{1}{2\rho}b^2$ for $\rho > 0$.

When Player II follows prescribed strategy In this case we would like to get convergence of payoffs to the average value of the games. To get this, using the notation $x_t^* = \operatorname*{argmin}_{x_t\in\Delta_n} f_t^\top A_t x_t$ and denoting the corresponding sequence regularity for Player I by C_T, we get

$$
\begin{aligned}
\sum_{t=1}^{T} f_t^\top A_t x_t - f_t^\top A_t x_t^* \\
&\le \log(T^2 n)\left(C_T + 2\right)\left(\frac{2\sqrt{\sum_{t=1}^{T}\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty^2}}{\log(T^2 n)L} + 32L\right) \\
&+ 2\sum_{t=1}^{T}\left\|A_t - A_{t-1}\right\|_\infty - 8L\sum_{t=1}^{T}\left\|\hat{x}_t - x_t\right\|_1^2 - 16L\sum_{t=1}^{T}\left\|\hat{x}'_{t-1} - x_t\right\|_1^2 \\
&+ \frac{1}{16L}\sum_{t=1}^{T}\left\|f_t - \hat{f}_{t-1}\right\|_1^2 + \frac{1}{16L}\sum_{t=1}^{T}\left\|\hat{f}_t - f_t\right\|_1^2 + \frac{1}{4L},
\end{aligned}
$$

where the term $\frac{1}{4L}$ appeared in the last line comparing to (5.4.19) is due to

$$
\frac{1}{16L}\sum_{t=1}^{T}\left\|\hat{f}_{t-1} - f_{t-1}\right\|_1^2 - \frac{1}{16L}\sum_{t=1}^{T}\left\|\hat{f}_t - f_t\right\|_1^2 \le \frac{1}{4L}.
$$

Using the same bound for Player 2 (using loss as $-f_t^\top A_t x_t$ on round t), as well as using $f_t^* = \underset{f_t \in \Delta_m}{\operatorname{argmin}} - f_t^\top A_t x_t$ and denoting the corresponding sequence regularity by C'_T, we have that

$$
\begin{aligned}
\sum_{t=1}^{T} f_t^\top A_t x_t - {f_t^*}^\top A_t x_t \ge - \log(T^2 m)\left(C'_T + 2\right)\left(\frac{2\sqrt{\sum_{t=1}^{T}\left\|A_t x_t - A_{t-1}x_{t-1}\right\|_\infty^2}}{\log(T^2 m)L} + 32L\right) \\
- 2\sum_{t=1}^{T}\left\|A_t - A_{t-1}\right\|_\infty + 8L\sum_{t=1}^{T}\left\|\hat{f}_t - f_t\right\|_1^2 + 16L\sum_{t=1}^{T}\left\|\hat{f}'_{t-1} - f_t\right\|_1^2 \\
- \frac{1}{16L}\sum_{t=1}^{T}\left\|x_t - \hat{x}_{t-1}\right\|_1^2 - \frac{1}{16L}\sum_{t=1}^{T}\left\|\hat{x}_t - x_t\right\|_1^2 - \frac{1}{4L}.
\end{aligned}
$$

Combining the two and noting that

$$
\begin{aligned}
{f_t^*}^\top A_t x_t = \sup_{f_t \in \Delta_m} f_t^\top A_t x_t &\ge \inf_{x_t \in \Delta_n} \sup_{f_t \in \Delta_m} f_t^\top A_t x_t \\
&= \sup_{f_t \in \Delta_m} \inf_{x_t \in \Delta_n} f_t^\top A_t x_t \ge \inf_{x_t \in \Delta_n} f_t^\top A_t x_t = f_t^\top A_t x_t^*,
\end{aligned}
$$

we get

$$\begin{aligned}\sum_{t=1}^{T} \sup_{f_t \in \Delta_m} f_t^\top A_t x_t \le \sum_{t=1}^{T} \inf_{x_t \in \Delta_n} \sup_{f_t \in \Delta_m} f_t^\top A_t x_t + \frac{256L}{T} + \frac{1}{2L} + 4 \sum_{t=1}^{T} \|A_t - A_{t-1}\|_\infty \\ + \log(T^2 n)\left(C_T + 2\right)\left(\frac{2\sqrt{\sum_{t=1}^{T} \left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty^2}}{\log(T^2 n)L} + 32L\right) \\ + \log(T^2 m)\left(C'_T + 2\right)\left(\frac{2\sqrt{\sum_{t=1}^{T} \|A_t x_t - A_{t-1} x_{t-1}\|_\infty^2}}{\log(T^2 m)L} + 32L\right) \\ + \left(\frac{1}{16L} - 8L\right)\sum_{t=1}^{T} \|\hat{x}_t - x_t\|_1^2 + \left(\frac{1}{16L} - 16L\right)\sum_{t=1}^{T} \|\hat{x}_{t-1} - x_t\|_1^2 \\ + \left(\frac{1}{16L} - 8L\right)\sum_{t=1}^{T} \left\|\hat{f}_t - f_t\right\|_1^2 + \left(\frac{1}{16L} - 16L\right)\sum_{t=1}^{T} \left\|\hat{f}_{t-1} - f_t\right\|_1^2,\end{aligned} \tag{5.4.20}$$

where the constant $256L/T$ appeared in the first line accounts for the identities

$$\|\hat{x}_{t-1} - x_t\|_1^2 - \left\|\hat{x}'_{t-1} - x_t\right\|_1^2 \le \frac{8}{T^2} \qquad \left\|\hat{f}_{t-1} - f_t\right\|_1^2 - \left\|\hat{f}'_{t-1} - f_t\right\|_1^2 \le \frac{8}{T^2}.$$

Using the triangle inequality again,

$$\begin{aligned}\sum_{t=1}^{T} \left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty^2 &= \sum_{t=1}^{T} \left\|f_t^\top A_t - f_t^\top A_{t-1} + f_t^\top A_{t-1} - f_{t-1}^\top A_{t-1}\right\|_\infty^2 \\ &\le 2\sum_{t=1}^{T} \|A_{t-1} - A_t\|_\infty^2 + 2\sum_{t=1}^{T} \|f_t - f_{t-1}\|_1^2 \\ &\le 2\sum_{t=1}^{T} \|A_{t-1} - A_t\|_\infty^2 \\ &\quad + 4\sum_{t=1}^{T} \left\|f_t - \hat{f}_{t-1}\right\|_1^2 + 4\sum_{t=1}^{T} \left\|\hat{f}_{t-1} - f_{t-1}\right\|_1^2,\end{aligned} \tag{5.4.21}$$

which also implies

$$\begin{aligned}
&\sqrt{\sum_{t=1}^{T}\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty^2} \\
&\qquad \le \sqrt{2\sum_{t=1}^{T}\|A_{t-1}-A_t\|_\infty^2 + 4\sum_{t=1}^{T}\left\|f_t-\hat{f}_{t-1}\right\|_1^2 + 4\sum_{t=1}^{T}\left\|\hat{f}_{t-1}-f_{t-1}\right\|_1^2} \\
&\qquad \le 2\sqrt{\sum_{t=1}^{T}\|A_{t-1}-A_t\|_\infty^2} + 2\sqrt{\sum_{t=1}^{T}\left\|f_t-\hat{f}_{t-1}\right\|_1^2 + \sum_{t=1}^{T}\left\|\hat{f}_{t-1}-f_{t-1}\right\|_1^2} \\
&\qquad \le 2\sqrt{\sum_{t=1}^{T}\|A_{t-1}-A_t\|_\infty^2} + 2 + 2\sum_{t=1}^{T}\left\|f_t-\hat{f}_{t-1}\right\|_1^2 + 2\sum_{t=1}^{T}\left\|\hat{f}_{t-1}-f_{t-1}\right\|_1^2 \\
&\qquad \le 2\sqrt{\sum_{t=1}^{T}\|A_{t-1}-A_t\|_\infty^2} + 10 + 2\sum_{t=1}^{T}\left\|f_t-\hat{f}_{t-1}\right\|_1^2 + 2\sum_{t=1}^{T}\left\|\hat{f}_t-f_t\right\|_1^2, \qquad (5.4.22)
\end{aligned}$$

where we used the bound $\sqrt{c} \le c+1$ for any $c \ge 0$ in the penultimate line. Similar bounds as Equations (5.4.21) and (5.4.22) hold for the other player as well. Using them in Equation 5.4.20 after some calculations, we conclude that

$$\begin{aligned}
\sum_{t=1}^{T}\sup_{f_t\in\Delta_m} f_t^\top A_t x_t \le{}& \sum_{t=1}^{T}\inf_{x_t\in\Delta_n}\sup_{f_t\in\Delta_m} f_t^\top A_t x_t + \frac{256L}{T} + \frac{1}{2L} + 4\sum_{t=1}^{T}\|A_{t-1}-A_t\|_\infty \\
&+ 32L\big(\log(T^2 n)C_T + \log(T^2 m)C_T' + 2\log(T^4 nm)\big) \\
&+ (C_T + C_T' + 4)\,\frac{20 + 4\sqrt{\sum_{t=1}^{T}\|A_{t-1}-A_t\|_\infty^2}}{L} \\
&+ 4\left(\frac{C_T+3}{L} - 2L\right)\left(\sum_{t=1}^{T}\left\|\hat{f}_t - f_t\right\|_1^2 + 2\sum_{t=1}^{T}\left\|\hat{f}_{t-1}-f_t\right\|_1^2\right) \\
&+ 4\left(\frac{C_T'+3}{L} - 2L\right)\left(\sum_{t=1}^{T}\|\hat{x}_t - x_t\|_1^2 + 2\sum_{t=1}^{T}\|\hat{x}_{t-1}-x_t\|_1^2\right).
\end{aligned}$$

When Player II is dishonest In this case we would like to bound Player I's regret regardless of the strategy adopted by Player II. Dropping one of the negative terms in Equation 5.4.16, we get :

$$\begin{aligned}
\sum_{t=1}^{T}\left(f_t^\top A_t x_t - f_t^\top A_t u_t\right) &\leq \frac{\log(T^2 n)\left(C_T(u_1,\dots,u_T)+2\right)}{\eta_T} - \frac{1}{2}\sum_{t=1}^{T}\frac{1}{\eta_t}\left\|\hat{x}_t - x_t\right\|_1^2 \\
&\quad + \sum_{t=1}^{T}\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty \left\|x_t - \hat{x}_t\right\|_1 \\
&\leq \frac{\log(T^2 n)\left(C_T(u_1,\dots,u_T)+2\right)}{\eta_T} - \frac{1}{2}\sum_{t=1}^{T}\frac{1}{\eta_t}\left\|\hat{x}_t - x_t\right\|_1^2 \\
&\quad + \sum_{t=1}^{T}\frac{\eta_{t+1}}{2}\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty^2 + \frac{1}{2}\sum_{t=1}^{T}\frac{1}{\eta_{t+1}}\left\|x_t - \hat{x}_t\right\|_1^2 .
\end{aligned} \tag{5.4.23}$$

Noting to the telescoping sum

$$\frac{1}{2}\sum_{t=1}^{T}\left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right)\left\|x_t - \hat{x}_t\right\|_1^2 \leq 2\sum_{t=1}^{T}\left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right) \leq \frac{2}{\eta_{T+1}},$$

as well as the choice of step-size (5.4.17) which entails

$$\begin{aligned}
&\sum_{t=1}^{T}\frac{\eta_{t+1}}{2}\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty^2 \\
&\qquad \leq \log(T^2 n)\frac{L}{2}\sum_{t=1}^{T}\sqrt{\sum_{i=1}^{t}\left\|f_i^\top A_i - f_{i-1}^\top A_{i-1}\right\|_\infty^2} - \sqrt{\sum_{i=1}^{t-1}\left\|f_i^\top A_i - f_{i-1}^\top A_{i-1}\right\|_\infty^2} \\
&\qquad \leq \log(T^2 n)\frac{L}{2}\sqrt{\sum_{t=1}^{T}\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty^2},
\end{aligned}$$

we bound (5.4.23) to obtain

$$
\begin{aligned}
&\sum_{t=1}^{T} f_t^\top A_t x_t - f_t^\top A_t u_t \\
&\qquad\qquad \le \frac{\log(T^2 n)\left(C_T(u_1,\ldots,u_T)+2\right)}{\eta_T} + \frac{2}{\eta_{T+1}} \\
&\qquad\qquad + \log(T^2 n)\frac{L}{2}\sqrt{\sum_{t=1}^{T}\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty^2} \\
&\qquad\qquad \le 2\log(T^2 n)\left(C_T(u_1,\ldots,u_T)+2\right)\left(32L + \frac{2\sqrt{\sum_{t=1}^{T}\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty^2}}{\log(T^2 n)L}\right) \\
&\qquad\qquad + \log(T^2 n)\frac{L}{2}\sqrt{\sum_{t=1}^{T}\left\|f_t^\top A_t - f_{t-1}^\top A_{t-1}\right\|_\infty^2}.
\end{aligned}
$$

A similar statement holds for Player II that her/his pay off converges at the provided rate to the average minimax equilibrium value. ■

Chapter 6

Concluding Remarks

6.1 Thesis Summary

In this thesis we addressed problems in the fields of online learning and statistical identification. In the first part of the thesis, we focused on environments where the leaner should statistically process the data for inference, whereas the second part was dedicated to online learning in which data arrive in a sequential fashion. In all problems studied in this thesis, the main objective was to understand data and its properties, and design efficient algorithms for inference of the *unknown*.

In the first part, we started with Chapter 2, and presented a distributed detection model where a network of agents aim to learn the underlying state of the world. As they cannot distinguish the true state in isolation, agents engage in a local communication. Each agent iteratively forms a belief about the state space using the collected data in its neighborhood. We analyzed the learning procedure for a *finite* time horizon. To study the efficiency of our algorithm versus its centralized counterpart, we brought forward the idea of KL cost. It turned out that network size, spectral gap, centrality of each agent and relative entropy of agents' signal structures are the key parameters that affect distributed detection. We further provided asymptotic analysis for *time-varying* network topologies. In Chapter 3, we considered an inverse problem where we reconstruct the topology of an *unknown* directed network of LTI systems. We proposed several reconstruction algorithms based on the power spectral properties of the network response to the input noise. Our first algorithm reconstructs the Boolean structure of a directed network based on a series of grounded dynamical responses. Our second algorithm recovers the exact structure of the network (including edge weights) when an eigenvalue-eigenvector pair of the connectivity matrix is known. This algorithm is useful, for example, when the connectivity matrix is a Laplacian matrix or the adjacency matrix of a regular graph. Apart from general directed networks, we also proposed more computationally efficient algorithms for reconstruction of both directed nonreciprocal networks and undirected networks.

The second part of the thesis focused on online learning in multi-player and one-player setting. In Chapter 4, we studied the MAB problem in the context of *multi-agent* networks. Each player sequentially pulls an arm, and receives a noisy payoff from a player-dependent distribution. Players want to detect the arm with highest average payoff among themselves (best global arm). Therefore, they communicate with each other to augment their imperfect observations with side information. Based on this model, we proposed a distributed online algorithm to compete with the best global arm. We further extended our results to sleeping bandits where a full set of arms is not available all the times. In both methods, the regret bound scales inversely in the spectral gap of the network, highlighting the impact of network structure. Interestingly, the regret scales down by the network size, which is an artifact of variance reduction through decentralizing the MAB problem. In Chapter 5, however, we considered an instance of *one-player* online learning. We proposed an online algorithm for dynamic environments, where the regret is measured with respect to *time-varying* benchmarks. Our proposed method is fully adaptive in the sense that the learner needs no prior knowledge of the environment. We derived a comprehensive upper bound on the dynamic regret capturing the interplay of regularity in the function sequence versus the comparator sequence. Interestingly, the regret bound adapts to the smaller quantity among the two, and selects the best of both worlds.

6.2 Future Directions

There are many open questions and interesting research directions relevant to what was presented in this thesis:

Distributed Detection in Fixed and Switching Network Topologies

In Section 2.5.2, we addressed a switching rule that works based on information of signals. Our convergence result holds for bidirectional communication. A potentially challenging problem is to investigate unidirectional communication, i.e., the case that *sending* and *receiving* information do not necessarily coincide. We have numerical experiments in support of convergence; however, the technical analysis is still under study.

Another interesting direction is optimal design of communication threshold τ. In particular, one can think of the following problem: given fix threshold τ, find the minimum (or expected) number of communication rounds for learning. This setting essentially addresses finite-time learning with an efficient communication protocol.

Competing with Structured Benchmarks in Online Optimization

The notion of regret with respect to arbitrary comparators has been studied in the past few years. Alternatively, one can consider the dynamic regret when the comparators are not arbitrary, and they follow an *unknown* dynamical model. Learning the dynamical model given different feedbacks (depending on the application of the problem) is an interesting subject of study. For instance, the comparator can follow an LTI dynamics where the learner partially observes the comparator. This scenario might be of potential interest to many control engineers to study linear or non-linear systems. A version of the problem has been investigated by [110] where the comparators potentially follow some dynamics. The authors proved regret bounds for the cases that the dynamical model is either *known* or *unknown* but *finite*. Going beyond the finite, unknown models is also an interesting line of research.

Stochastic Optimization in Non-stationary Environments

Our results in Chapter 5 were based on receiving *noiseless* feedback. That is, at each round of the algorithm, the learner could query a noiseless gradient. A complementary direction is to develop an adaptive algorithm that works in *stochastic* environments, i.e., an algorithm with low *expected* dynamic regret given a noisy access to loss functions and gradients. The main challenge in the problem is that noisy feedback does not quite specify the changes of the environment. For instance, even when the environment is stationary, the learner could incorrectly infer some non-stationarity due to noise. It is currently known that the regret bound in this setting can be expressed in terms of variability of loss functions [116]; however, an adaptive solution to the problem is still open.

Multi-player Zero-sum Games

We studied the application of dynamic regret to two-player zero-sum games in Chapter 5. Another interesting direction is to propose a framework which can be applied to multi-player games (games on networks or graphical games [121]). For instance, in a recent work of [122] a multi-player static framework is addressed. The extension of this problem to dynamic setting would be an interesting line of research.

Bibliography

[1] D. Watts and S. Strogatz, “Collective dynamics of small-world networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[2] A.-L. Barabási, *Linked: The New Science Of Networks*. Basic Books, 2002.

[3] M. Jackson, *“Social and Economic Networks”*. Princeton Univ Pr, 2008.

[4] R. R. Tenney and N. R. Sandell Jr, “Detection with distributed sensors,” *IEEE Transactions on Aerospace Electronic Systems*, vol. 17, pp. 501–510, 1981.

[5] J. N. Tsitsiklis *et al.*, “Decentralized detection,” *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.

[6] V. Borkar and P. P. Varaiya, “Asymptotic agreement in distributed estimation,” *IEEE Transactions on Automatic Control*, vol. 27, no. 3, pp. 650–655, 1982.

[7] S. Kar, J. Moura, and K. Ramanan, “Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication,” *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.

[8] O. Dekel, R. Gilad-Bachrach, O. Shamir, and L. Xiao, “Optimal distributed online prediction using mini-batches,” *The Journal of Machine Learning Research*, vol. 13, pp. 165–202, 2012.

[9] A. Nedic and A. Ozdaglar, “Distributed subgradient methods for multi-agent optimization,” *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[10] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, “On distributed averaging algorithms and quantization effects,” *IEEE Transactions on Automatic Control*, vol. 54, no. 11, pp. 2506–2517, 2009.

[11] I. Lobel and A. Ozdaglar, “Distributed subgradient methods over random networks,” in *Proc. Allerton Conf. Commun., Control, Comput*, 2008.

[12] S. Ram, A. Nedic, and V. Veeravalli, “Distributed stochastic subgradient projection algorithms for convex optimization,” *Journal of optimization theory and applications*, vol. 147, no. 3, pp. 516–545, 2010.

[13] J. Duchi, A. Agarwal, and M. Wainwright, “Dual averaging for distributed optimization: convergence analysis and network scaling,” *IEEE Transactions on Automatic Control*, pp. 592–607, March 2012.

[14] T. L. Lai and H. Robbins, “Asymptotically efficient adaptive allocation rules,” *Advances in applied mathematics*, vol. 6, no. 1, pp. 4–22, 1985.

[15] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, “The nonstochastic multiarmed bandit problem,” *SIAM Journal on Computing*, vol. 32, no. 1, pp. 48–77, 2002.

[16] P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time analysis of the multiarmed bandit problem,” *Machine learning*, vol. 47, no. 2-3, pp. 235–256, 2002.

[17] J.-Y. Audibert, R. Munos, and C. Szepesvári, “Exploration–exploitation tradeoff using vari-

ance estimates in multi-armed bandits," *Theoretical Computer Science*, vol. 410, no. 19, pp. 1876–1902, 2009.

[18] S. Bubeck and N. Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *arXiv preprint arXiv:1204.5721*, 2012.

[19] N. Cesa-Bianchi, G. Lugosi *et al.*, *Prediction, learning, and games*. Cambridge University Press Cambridge, 2006, vol. 1.

[20] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[21] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *International Conference on Machine Learning*, 2003.

[22] J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 407–416, 2003.

[23] F. Bullo, J. Cortés, and S. Martínez, *Distributed control of robotic networks: a mathematical approach to motion coordination algorithms*. Princeton Univ Pr, 2009.

[24] N. A. Atanasov, J. Le Ny, and G. J. Pappas, "Distributed algorithms for stochastic source seeking with mobile robot networks," *Journal of Dynamic Systems, Measurement, and Control*, 2014.

[25] S. Shahrampour, S. Rakhlin, and A. Jadbabaie, "Online learning of dynamic parameters in social networks," in *Advances in Neural Information Processing Systems*, 2013.

[26] J. Tsitsiklis, “Problems in decentralized decision making and computation.” DTIC Document, Tech. Rep., 1984.

[27] A. Jadbabaie, J. Lin, and A. S. Morse, “Coordination of groups of mobile autonomous agents using nearest neighbor rules,” *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[28] R. Olfati-Saber and R. M. Murray, “Consensus problems in networks of agents with switching topology and time-delays,” *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.

[29] 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, 2012.

[30] J. C. Duchi, A. Agarwal, and M. J. Wainwright, “Dual averaging for distributed optimization: convergence analysis and network scaling,” *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 592–606, 2012.

[31] S. Shahrampour and A. Jadbabaie, “Exponentially fast parameter estimation in networks using distributed dual averaging,” in *IEEE Conference on Decision and Control (CDC)*, 2013, pp. 6196–6201.

[32] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, “Distributed detection: Finite-time analysis and impact of network topology,” *arXiv preprint arXiv:1409.8606*, 2014.

[33] S. Shahrampour, M. A. Rahimian, and A. Jadbabaie, “Switching to learn,” in *American Control Conference (ACC)*, July 2015, pp. 2918–2923.

[34] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 1917–1932, 2011.

[35] D. Jakovetic, J. M. Moura, and J. Xavier, "Distributed detection over noisy networks: Large deviations analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4306–4320, 2012.

[36] D. Bajovic, D. Jakovetic, J. M. Moura, J. Xavier, and B. Sinopoli, "Large deviations performance of consensus+ innovations distributed detection with non-gaussian observations," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5987–6002, 2012.

[37] A. Lalitha, A. Sarwate, and T. Javidi, "Social learning and distributed hypothesis testing," in *International Symposium on Information Theory (ISIT)*, 2014, pp. 551–555.

[38] M. A. Rahimian, P. Molavi, and A. Jadbabaie, "(Non-) bayesian learning without recall," in *IEEE Conference on Decision and Control (CDC)*, 2014, pp. 5730–5735.

[39] M. A. Rahimian, S. Shahrampour, and A. Jadbabaie, "Learning without recall by random walks on directed graphs," *arXiv preprint arXiv:1509.04332*, 2015.

[40] A. Nedić, A. Olshevsky, and C. A. Uribe, "Nonasymptotic convergence rates for cooperative learning over time-varying directed graphs," *arXiv preprint arXiv:1410.1977*, 2014.

[41] ——, "Network independent rates in distributed learning," *arXiv preprint arXiv:1509.08574*, 2015.

[42] ——, "Fast convergence rates for distributed non-bayesian learning," *arXiv preprint arXiv:1508.05161*, 2015.

[43] K. Rahnama Rad and A. Tahbaz-Salehi, “Distributed parameter estimation in networks,” in *IEEE Conference on Decision and Control (CDC)*, 2010, pp. 5050–5055.

[44] A. Jadbabaie, P. Molavi, and A. Tahbaz-Salehi, “Information heterogeneity and the speed of learning in social networks,” *Columbia Business School Research Paper*, no. 13-28, 2013.

[45] K. Drakopoulos, A. Ozdaglar, and J. N. Tsitsiklis, “On learning with finite memory,” *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6859–6872, 2013.

[46] J. S. Rosenthal, “Convergence rates for markov chains,” *SIAM Review*, vol. 37, no. 3, pp. 387–405, 1995.

[47] J. D. Abernethy, E. Hazan, and A. Rakhlin, “Interior-point methods for full-information and bandit online learning,” *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4164–4175, 2012.

[48] A. Nemirovskii and D. Yudin, *Problem complexity and method efficiency in optimization*. Wiley (Chichester and New York), 1983.

[49] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*. American Mathematical Soc., 2009.

[50] S. Boyd, P. Diaconis, and L. Xiao, “Fastest mixing markov chain on a graph,” *SIAM review*, vol. 46, no. 4, pp. 667–689, 2004.

[51] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[52] M. G. Rabbat, R. D. Nowak, and J. A. Bucklew, “Generalized consensus computation in

networked systems with erasure links," in *IEEE Workshop on Signal Processing Advances in Wireless Communications*, 2005, pp. 1088–1092.

[53] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1867–1872, 2005.

[54] F. R. Chung, *Spectral graph theory*. American Mathematical Soc., 1997, vol. 92.

[55] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Transactions on Automatic Control*, vol. 50, no. 2, pp. 169–182, 2005.

[56] C. McDiarmid, "Concentration," in *Probabilistic methods for algorithmic discrete mathematics*. Springer, 1998, pp. 195–248.

[57] F. Geier, J. Timmer, and C. Fleck, "Reconstructing gene-regulatory networks from time series, knock-out data, and prior knowledge," *BMC Systems Biology*, vol. 1, no. 1, p. 11, 2007.

[58] A. Julius, M. Zavlanos, S. Boyd, and G. Pappas, "Genetic network identification using convex programming," *Systems Biology, IET*, vol. 3, no. 3, pp. 155–166, 2009.

[59] M. Timme, "Revealing network connectivity from response dynamics," *Physical Review Letters*, vol. 98, no. 22, p. 224101, 2007.

[60] D. Napoletani and T. Sauer, "Reconstructing the topology of sparsely connected dynamical networks," *Physical Review E*, vol. 77, no. 2, p. 26103, 2008.

[61] R. Mantegna and H. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, 2000.

[62] S. Tuna, "Conditions for synchronizability in arrays of coupled linear systems," *IEEE Transactions on Automatic Control*, vol. 54, no. 10, pp. 2416–2420, 2009.

[63] L. Scardovi and R. Sepulchre, "Synchronization in networks of identical linear systems," *Automatica*, vol. 45, no. 11, pp. 2557–2562, 2009.

[64] M. Nabi-Abdolyousefi and M. Mesbahi, "Network identification via node knockout," *IEEE Transactions on Automatic Control*, vol. 57, no. 12, pp. 3214–3219, 2012.

[65] S. Shahrampour and V. M. Preciado, "Topology identification of directed dynamical networks via power spectral analysis," *IEEE Transactions on Automatic Control*, vol. 60, no. 8, 2015.

[66] D. R. Brillinger, *Time series: data analysis and theory*. Siam, 1981, vol. 36.

[67] D. Materassi and M. Salapaka, "On the problem of reconstructing an unknown topology via locality properties of the wiener filter," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1765–1777, 2012.

[68] D. Tylavsky and G. Sohie, "Generalization of the matrix inversion lemma," *Proceedings of the IEEE*, vol. 74, no. 7, pp. 1050–1052, 1986.

[69] S. Shahrampour and V. M. Preciado, "Reconstruction of directed networks from consensus dynamics," in *IEEE American Control Conference (ACC)*, 2013, pp. 1685–1690.

[70] D. West, *Introduction to Graph Theory*. Prentice-Hall, 2001, vol. 2.

[71] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel method for nonlinear granger causality," *Physical Review Letters*, vol. 100, no. 14, p. 144103, 2008.

[72] C. W. Granger, “Investigating causal relations by econometric models and cross-spectral methods,” *Econometrica*, pp. 424–438, 1969.

[73] F. R. Bach and M. I. Jordan, “Learning graphical models for stationary time series,” *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2189–2199, 2004.

[74] J. Gonçalves and S. Warnick, “Necessary and sufficient conditions for dynamical structure reconstruction of LTI networks,” *IEEE Transactions on Automatic Control*, vol. 53, no. 7, pp. 1670–1674, 2008.

[75] Y. Yuan, G. Stan, S. Warnick, and J. Goncalves, “Robust dynamical network structure reconstruction,” *Automatica*, 2011.

[76] D. Hayden, Y. Yuan, and J. Goncalves, “Network reconstruction from intrinsic noise,” *arXiv preprint arXiv:1310.0375*, 2013.

[77] D. Materassi and G. Innocenti, “Topological identification in networks of dynamical systems,” *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1860–1871, 2010.

[78] M. Nabi-Abdolyousefi and M. Mesbahi, “Sieve method for consensus-type network tomography,” *IET Control Theory & Applications*, vol. 6, no. 12, pp. 1926–1932, 2012.

[79] M. Nabi-Abdolyousefi, M. Fazel, and M. Mesbahi, “A graph realization approach to network identification,” in *IEEE Conference on Decision and Control (CDC)*, Dec 2012, pp. 4642–4647.

[80] M. Fazlyab and V. M. Preciado, “Robust topology identification and control of lti networks,” in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2014, pp. 918–922.

[81] S. Nabavi, A. Chakrabortty, and P. P. Khargonekar, “A global identifiability condition for consensus networks with tree graphs,” *arXiv preprint arXiv:1412.0684*, 2014.

[82] S. Nabavi and A. Chakrabortty, “A graph-theoretic condition for global identifiability of weighted consensus networks,” *IEEE Transactions on Automatic Control*, 2015.

[83] F. Morbidi and A. Y. Kibangou, “A distributed solution to the network reconstruction problem,” *Systems & Control Letters*, vol. 70, pp. 85–91, 2014.

[84] T.-M. D. Tran and A. Y. Kibangou, “Distributed network topology reconstruction in presence of anonymous nodes,” in *European Signal Processing Conference (EUSIPCO 2015)*, 2015.

[85] P. Torres, J.-W. van Wingerden, and M. Verhaegen, “Po-moesp subspace identification of directed acyclic graphs with unknown topology,” *Automatica*, vol. 53, pp. 60–71, 2015.

[86] R. Kleinberg, A. Niculescu-Mizil, and Y. Sharma, “Regret bounds for sleeping experts and bandits,” *Machine learning*, vol. 80, no. 2-3, pp. 245–272, 2010.

[87] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, “Multi-armed bandits in multi-agent networks,” 2015.

[88] K. Liu and Q. Zhao, “Distributed learning in multi-armed bandit with multiple players,” *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5667–5681, 2010.

[89] C. Tekin and M. Liu, “Performance and convergence of multi-user online learning,” in *Game Theory for Networks*. Springer, 2012, pp. 321–336.

[90] D. Kalathil, N. Nayyar, and R. Jain, “Decentralized learning for multiplayer multiarmed bandits,” *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2331–2345, 2014.

[91] S. Buccapatnam, A. Eryilmaz, and N. B. Shroff, “Multi-armed bandits in the presence of side observations in social networks,” in *IEEE Conference on Decision and Control (CDC)*, 2013, pp. 7309–7314.

[92] S. Kar, H. V. Poor, and S. Cui, “Bandit problems in networks: Asymptotically efficient distributed allocation rules,” in *IEEE Conference on Decision and Control and European Control Conference*, 2011, pp. 1771–1778.

[93] S. Mannor and O. Shamir, “From bandits to experts: On the value of side-observations,” in *Advances in Neural Information Processing Systems*, 2011, pp. 684–692.

[94] S. Caron, B. Kveton, M. Lelarge, and S. Bhagat, “Leveraging side observations in stochastic bandits,” *Uncertainty in Artificial Intelligence*, 2012.

[95] N. Cesa-Bianchi, C. Gentile, and G. Zappella, “A gang of bandits,” in *Advances in Neural Information Processing Systems*, 2013, pp. 737–745.

[96] T. Lattimore and R. Munos, “Bounded regret for finite-armed structured bandits,” in *Advances in Neural Information Processing Systems*, 2014, pp. 550–558.

[97] P. B. Reverdy, V. Srivastava, and N. E. Leonard, “Modeling human decision making in generalized gaussian multiarmed bandits,” *Proceedings of the IEEE*, 2014.

[98] Y. Gai, B. Krishnamachari, and R. Jain, “Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards and individual observations,” *IEEE/ACM Transactions on Networking (TON)*, vol. 20, no. 5, pp. 1466–1478, 2012.

[99] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[100] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma, "Belief consensus and distributed hypothesis testing in sensor networks," in *Networked Embedded Sensing and Control*. Springer, 2006, pp. 169–182.

[101] A. DeSantis, G. Markowsky, and M. N. Wegman, "Learning probabilistic prediction functions," in *29th Annual Symposium on Foundations of Computer Science*. IEEE, 1988, pp. 110–119.

[102] N. Littlestone and M. K. Warmuth, "The weighted majority algorithm," *Information and computation*, vol. 108, no. 2, pp. 212–261, 1994.

[103] V. G. Vovk, "Aggregating strategies," in *Proc. Third Workshop on Computational Learning Theory*. Morgan Kaufmann, 1990, pp. 371–383.

[104] A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan, "Online optimization: Competing with dynamic comparators," in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, 2015, pp. 398–406.

[105] E. Hazan, A. Agarwal, and S. Kale, "Logarithmic regret algorithms for online convex optimization," *Machine Learning*, vol. 69, no. 2-3, pp. 169–192, 2007.

[106] J. Abernethy, P. L. Bartlett, A. Rakhlin, and A. Tewari, "Optimal strategies and minimax lower bounds for online convex games," in *Proceedings of the Nineteenth Annual Conference on Computational Learning Theory*, 2008.

[107] O. Bousquet and M. K. Warmuth, "Tracking a small set of experts by mixing past posteriors," *The Journal of Machine Learning Research*, vol. 3, pp. 363–396, 2003.

[108] N. Cesa-Bianchi, P. Gaillard, G. Lugosi, and G. Stoltz, “A new look at shifting regret,” *CoRR abs/1202.3323*, 2012.

[109] N. Buchbinder, S. Chen, J. Naor, and O. Shamir, “Unified algorithms for online learning and competitive analysis.” *Journal of Machine Learning Research-Proceedings Track*, vol. 23, pp. 5–1, 2012.

[110] E. C. Hall and R. M. Willett, “Online optimization in dynamic environments,” *arXiv preprint arXiv:1307.5944*, 2013.

[111] A. Rakhlin and K. Sridharan, “Online learning with predictable sequences,” in *Conference on Learning Theory*, 2013, pp. 993–1019.

[112] S. Rakhlin and K. Sridharan, “Optimization, learning, and games with predictable sequences,” in *Advances in Neural Information Processing Systems*, 2013, pp. 3066–3074.

[113] E. Hazan and S. Kale, “Extracting certainty from uncertainty: Regret bounded by variation in costs,” *Machine learning*, vol. 80, no. 2-3, pp. 165–188, 2010.

[114] C.-K. Chiang, T. Yang, C.-J. Lee, M. Mahdavi, C.-J. Lu, R. Jin, and S. Zhu, “Online optimization with gradual variations,” in *Conference on Learning Theory*, 2012.

[115] E. Hazan and S. Kale, “Better algorithms for benign bandits,” *The Journal of Machine Learning Research*, vol. 12, pp. 1287–1311, 2011.

[116] O. Besbes, Y. Gur, and A. Zeevi, “Non-stationary stochastic optimization,” *arXiv preprint arXiv:1307.5449*, 2013.

[117] C. Daskalakis, A. Deckelbaum, and A. Kim, “Near-optimal no-regret algorithms for zero-sum games,” *Games and Economic Behavior*, 2014.

[118] A. S. Nemirovski and D. B. Yudin, *Problem complexity and method efficiency in optimization*. Wiley (Chichester and New York), 1983.

[119] A. Beck and M. Teboulle, "Mirror descent and nonlinear projected subgradient methods for convex optimization," *Operations Research Letters*, vol. 31, no. 3, pp. 167–175, 2003.

[120] A. Nemirovski, "Prox-method with rate of convergence o (1/t) for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems," *SIAM Journal on Optimization*, vol. 15, no. 1, pp. 229–251, 2004.

[121] M. Kearns, "Graphical games," *Algorithmic game theory*, vol. 3, pp. 159–180, 2007.

[122] V. Syrgkanis, A. Agarwal, H. Luo, and R. E. Schapire, "Fast convergence of regularized learning in games," *arXiv preprint arXiv:1507.00407*, 2015.