UNIVERSITY OF CALIFORNIA

Los Angeles

Multi-Objective Learning over Adaptive Networks

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Electrical Engineering

by

Chung-Kai Yu

2017

ABSTRACT OF THE DISSERTATION

Multi-Objective Learning over Adaptive Networks

by

Chung-Kai Yu

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2017

Professor Ali H. Sayed, Chair

Motivated by the expanding interest in applications where online learning and decision making by networked agents is a necessity, this dissertation deals with the design of adaptive networks where agents need to carry out strategic decisions toward different but complementary objectives under uncertainty and randomness.

Adaptive networks consist of collections of agents with learning abilities that interact with each other locally in order to solve distributed processing and inference tasks in real-time. In our first constrained problem, we examine the design and evolution of adaptive networks in which agents have multiple but complementary objectives. In these scenarios, selfish learning strategies by individual agents can influence the network dynamics in adverse ways. This situation is even more challenging in nonstationary environments where the solution to the multi-objective optimization problem can drift with time due to changes in the statistical distribution of the data.

We specifically formulate multi-objective optimization problems where agents seek to minimize their individual costs subject to constraints that are locally coupled. The coupling arises because the individual costs and the constraints can be dependent on actions by other agents in the neighborhood. In these types of problems, the Nash equilibrium is a desired and stable solution since at this location no agent can benefit by unilaterally deviating from the equilibrium. We therefore focus on developing distributed online strategies that enable agents to approach the Nash equilibrium. We illustrate an application of the results to a

stochastic version of the network Cournot competition problem, which arises in a variety of useful problems such as in modeling economic trading with geographical considerations, power management over smart grids, and resource allocation protocols.

Using this formulation, we then extend earlier contributions on adaptive networks, which generally assume that the agents work together for a common global objective or when they observe data that is generated by a common target model or parameter vector. We relax this condition and consider a broader scenario where individual agents may only have access to partial information about the global target vector, i.e., each agent may be sensing only a subset of the entries of the global target, and the number of these entries can be different across the agents. We develop cooperative distributed techniques where agents are only required to share estimates of their common entries and still can benefit from neighboring agents. Since agents' interactions are limited to exchanging estimates of select few entries, communication overhead is significantly reduced.

We also examine the behavior of adaptive networks where information-sharing is subject to a positive communications cost over the edges linking the agents. In this situation, we show that if left unattended, the optimal strategy for the agents is to behave in a selfish manner and not to participate in the sharing of information. We hence develop mechanisms to help turn selfish agents into cooperative formations. In one method, we design an adaptive reputation protocol to adjust agents' reputation in accordance to their past actions, which can then be used to predict their subsequent actions. In a second method, we allow agents to decide with whom to cluster and share information. When the communication cost is small, the proposed mechanisms entice agents to cooperate and thus enhance the overall social benefit of the network.

The dissertation of Chung-Kai Yu is approved.

Mihaela van der Schaar

Danijela Cabric

Robert M'Closkey

Ali H. Sayed, Committee Chair

University of California, Los Angeles

2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

years.

2006   B.S. in Electrical Engineering

            National Taiwan University, Taipei, Taiwan.

2008   M.S. in Communication Engineering

            National Taiwan University, Taipei, Taiwan.

2008–2011   Research Assistant

            Wireless Broadband Communication System Laboratory

            National Taiwan University, Taipei, Taiwan.

2011–2016   Research and Teaching Assistant

            Department of Electrical Engineering

            University of California, Los Angeles.

2016–2017   Dissertation Year Fellowship

            University of California, Los Angeles.

CHAPTER 1

Motivation

1.1 Introduction

Motivated by the expanding interest in applications where online learning and decision making by networked agents is a necessity, this dissertation focuses on the design of adaptive networks where agents need to carry out strategic decisions toward different but complementary objectives under uncertainty and randomness. Adaptive networks consist of a collection of agents with learning abilities that interact with each other locally in order to solve distributed processing and inference tasks in real-time. In most prior works, agents are assumed to be cooperative and designed to follow certain distributed rules such as the consensus strategy (e.g., [1–9]) or the diffusion strategy (e.g., [10–17]). Most of these earlier studies deal with scenarios where agents cooperate with their neighbors to achieve a common single objective that is beneficial to the entire family of agents.

There have been recent works where multi-task scenarios are introduced and studied in some detail over adaptive networks [18–20]. In these studies, different agents may be interested in different objectives and most of the available analyses focus on mean-square-error formulations. This dissertation deals with more general optimization scenarios that involve various types of constraints such as (a) having agents with different but complementary objectives; (b) having agents with access to only partial information; and (c) having selfish agents that seek to reduce their own communication cost at the expense of the social good for the overall network.

The multiplicity of tasks and the differences in objectives, although related, can give rise to selfish behavior by agents and lead to individual strategies that can influence the

1

network dynamics in adverse ways. In such scenarios, the Nash equilibrium becomes a desired and stable solution since at this location no agent can benefit by unilaterally deviating from the solution. Therefore, one main goal of this work is to develop distributed learning strategies for agents to gradually and continuously learn the Nash equilibrium under random environments. This situation is even more challenging in nonstationary environments where the Nash equilibrium can drift with time due to changes in the statistical distribution of the data. Consequently, a key challenge in our formulation is that agents need to operate in response to streaming data and be able to respond to changes in the statistical properties of the data, the nature of the task, and even the behavior of neighboring agents.

Notation: Throughout the dissertation, we use lowercase letters to denote vectors and scalars, uppercase letters for matrices, plain letters for deterministic variables, and boldface letters for random variables. All vectors in our treatment are column vectors, with the exception of the regression vectors, $\boldsymbol{u}_{k,i}$. The symbol T denotes transposition, and the symbol $*$ denotes complex conjugation for scalars and complex-conjugate transposition for matrices.

1.2 Single-Objective Learning over Networks

In preparation for the treatment of these problems, we first review adaptive networks for single-task optimization, which consist of N agents cooperatively working to minimize a global cost function of the following form in a distributed manner:

$$\min_{w \in \mathbb{R}^M} \quad \sum_{k=1}^{N} J_k(w) \triangleq J^{\text{glob}}(w) \tag{1.1}$$

Here, the symbol $w \in \mathbb{R}^M$ denotes a vector of size $M \times 1$. We denote the minimizer of $J^{\text{glob}}(w)$ by w^o, i.e.,

$$w^o \triangleq \arg \min_{w \in \mathbb{R}^M} \quad J^{\text{glob}}(w) \tag{1.2}$$

Moreover, each $J_k(w) : \mathbb{R}^M \to \mathbb{R}$ denotes the scalar individual cost function at agent k and is assumed to be twice-differentiable and strongly-convex in w. When the cost functions
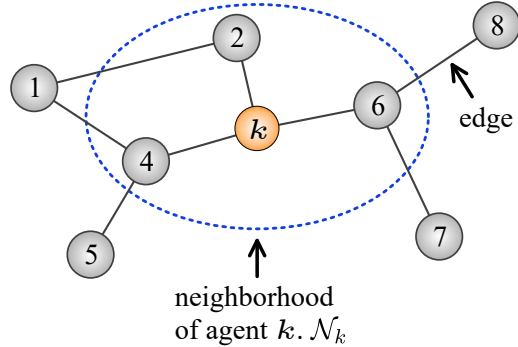
Figure 1.1: An example to illustrate the network topology with connected agents.

$\{J_k(w)\}$ are simultaneously minimized at the common minimizer w^o, all agents in the network will share the same objective of estimating w^o.

Various strategies and performance analyses have been put forward for the operation of such networks under streaming data. For example, in a centralized or batch strategy, the agents transmit the collected data for processing to a fusion center. One possible centralized implementation is to employ the gradient-descent strategy:

$$w_i = w_{i-1} - \frac{\mu}{N} \sum_{k=1}^{N} \nabla_{w^\mathsf{T}} J_k(w_{i-1}), \quad i \geq 0 \tag{1.3}$$

where $\mu > 0$ is the step-size and the notation $\nabla_{w^\mathsf{T}} J(a)$ denotes the gradient vector of the function $J(w)$ with respect to w^T and evaluated at $w = a$. After processing the collected data, the fusion center shares the results back with the distributed agents. While this centralized strategy is powerful, it suffers from some limitations. First, the transmission between agents and the fusion center can be costly in real-time applications. Second, privacy can be a critical concern in certain sensitive applications since agents may be reluctant to share their data with the fusion center. Furthermore, the centralized strategy is vulnerable since the network stops functioning if the fusion center fails. Network scalability is another critical problem because we will need a more powerful fusion center to process data for larger size networks.

In comparison, distributed strategies rely on localized interactions among the agents. Suppose that the agents are connected by a network topology. Let us denote the neighborhood of each agent k by $\mathcal{N}_k$, which includes k itself. We provide an example of a network

topology in Fig. 1.1. Two prominent classes of distributed strategies that can be used to compute local estimates $w_{k,i}$ in a distributed and online manner are consensus strategies [6,7] and diffusion strategies [10,11,14]. In the consensus strategy, each agent uses the following iteration to estimate w^o:

$$w_{k,i} = \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \; w_{\ell,i-1} - \mu_k \nabla_{w^\mathsf{T}} J_k(w_{k,i-1}), \quad i \geq 0 \tag{1.4}$$

The coefficients $\{a_{\ell k}\}$ are required to satisfy:

$$a_{\ell k} \geq 0, \quad \sum_{\ell=1}^{N} a_{\ell k} = 1, \quad \text{and} \quad a_{\ell k} = 0 \quad \text{if} \quad \ell \notin \mathcal{N}_k \tag{1.5}$$

The diffusion strategy, on the other hand, has several variants. In the adapt-then-combine (ATC) formulation, the agents update their estimates according to the following recursive construction:

$$(\text{ATC}) \begin{cases} \psi_{k,i} = w_{k,i-1} - \mu_k \nabla_{w_k^\mathsf{T}} J_k(w_{k,i-1}) & (1.6) \\ w_{k,i} = \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \; \psi_{\ell,i} & (1.7) \end{cases}$$

In the first step (1.6), an intermediate estimate $\psi_{k,i}$ is determined by adjusting the existing estimate $w_{k,i-1}$ using local data. The second step (1.7) uses non-negative coefficients $\{a_{\ell k}\}$ to combine the estimates from the neighbors. We can reverse the order of the two steps and obtain the combine-then-adapt (CTA) formulation:

$$(\text{CTA}) \begin{cases} \psi_{k,i} = \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \; w_{\ell,i-1} & (1.8) \\ w_{k,i} = \psi_{k,i} - \mu_k \nabla_{w_k^\mathsf{T}} J_k(\psi_{k,i}) & (1.9) \end{cases}$$

Compared to the consensus strategy (1.4), the diffusion strategies have been shown to have superior stability and mean-square performance properties [16,21].

1.3 Multi-Objective Learning over Networks

In formulation (1.1), all agents are interested in optimizing the same aggregate cost. In multi-objective learning, on the other hand, each agent k is interested in solving an optimization
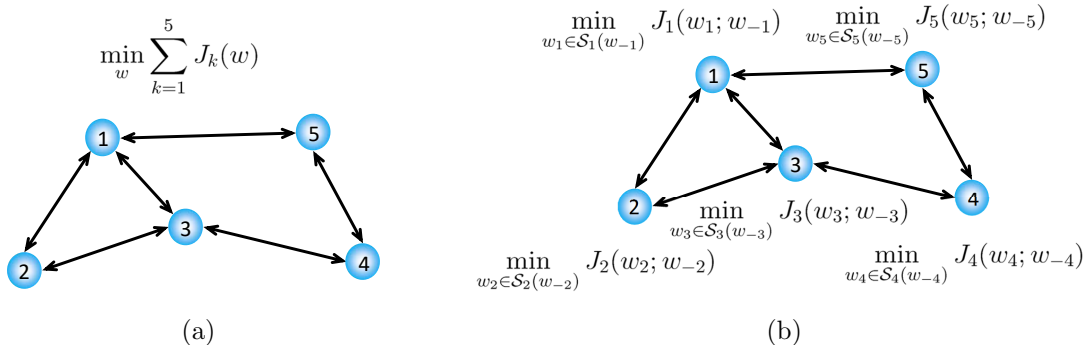
Figure 1.2: (a) Single-objective learning over networks. (b) Multi-objective learning over networks.

problem of the form:

$$\min_{w_k \in \mathcal{S}_k(w_{-k})} J_k(w_k; w_{-k}) \tag{1.10}$$

where we denote the unknown parameter (also called action) of agent k by the vector $w_k \in \mathbb{R}^{M_k \times 1}$ and collect the action profile of all other neighboring agents into the aggregate vector:

$$w_{-k} \triangleq \mathrm{col}\{w_\ell; \ell \in \mathcal{N}_k \setminus \{k\}\} \tag{1.11}$$

We note from (1.10) that each action w_k is required to belong to some feasible set $\mathcal{S}_k(w_{-k})$. Furthermore, the argument of $J_k(\cdot)$ does not depend solely on w_k but also on w_{-k}. Likewise, the argument of the feasible set depends on w_{-k}. Therefore, the neighbors' actions, represented by w_{-k}, can influence the selection of agent k for its action, w_k. In Fig. 1.2, we compare the optimization targets of single-objective and multi-objective learning over network topologies. Formulation (1.10) arises naturally in the modeling of many applications, as illustrated by the following examples.

Example 1.1. *(**Distributed power allocation**)* In this example, we explain one important application for formulation (1.10), of broad interest in many signal processing applications and communications scenarios; it relates to the problem of distributed power allocation over wireless networks — see, e.g., [22]. Referring to Fig. 1.3, assume that each agent k representing a femto-base station decides its transmission power w_k. The goal of k is to maximize the Shannon capacity function (transmission rate) to its macro-user terminal, while guaranteeing a minimum level of quality of service for all macro-user terminals in the network.
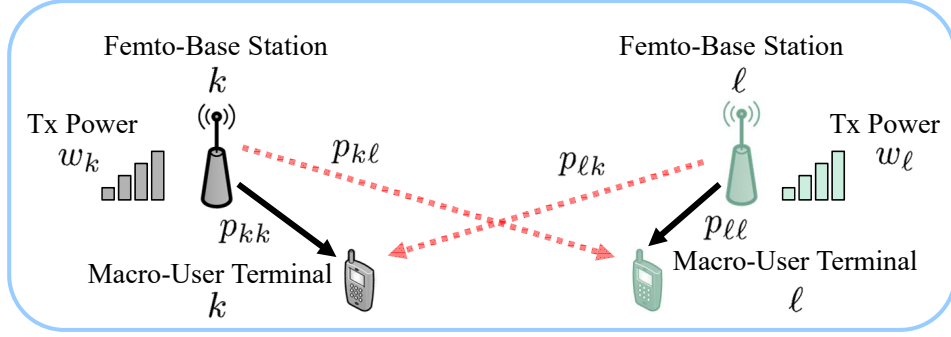
(a)
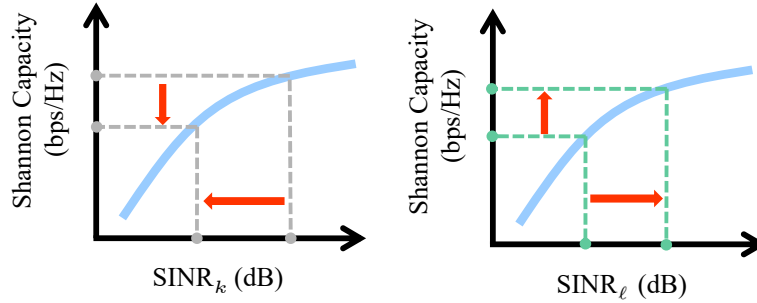
(b)

Figure 1.3: (a) Distributed power allocation with two femto-base stations k and ℓ. (b) The impact on Shannon capacity of macro-user terminals when transmission power of femto-base station ℓ increases.

Here we assume each femto-base station provides service to only one macro-user terminal. The problem can be mathematically formulated as follows:

$$\min_{w_k \in \mathbb{R}} \quad J_k(w_k; w_{-k})$$

$$\text{subject to} \quad w_k \geq 0, \quad g_f(w_k, w_{-k}) \leq 0, \quad f \in \mathcal{N}_k \tag{1.12}$$

where the cost function $J_k(\cdot)$ relates to the Shannon capacity, which is determined by the signal-to-interference and noise ratio (SINR) to the connecting macro-user terminal, and therefore, also determined by the neighboring agents' transmission powers w_{-k}. An example of $J_k(\cdot)$ with regularization would be

$$J_k(w_k; w_{-k}) = -\log(1 + \text{SINR}_k) + |w_k|^2 \tag{1.13}$$

6

where

$$\text{SINR}_k \triangleq \frac{p_{kk}w_k}{\sum\limits_{\ell \in \mathcal{N}_k \backslash \{k\}} p_{\ell k}w_\ell + \sigma_k^2} \tag{1.14}$$

and the $\{p_{\ell k}\}$ denote the channel gains from each femto-base station ℓ to macro-user terminal k, and σ_k^2 is the noise variance at that macro-user terminal. The first term in (1.13) is from the Shannon channel capacity with additive noise and interference [23] where the negative sign is because we are considering cost functions instead of payoff functions, and the second term is used to prevent the transmission power w_k from being too large. For example, in the scenario of Fig. 1.3 we consider two femto-base stations k and ℓ. If the transmission power of ℓ increases, e.g., by using a larger w_ℓ, the Shannon capacity of k increases since SINR_ℓ becomes large. At the same time, the interference from ℓ to k also becomes large, and therefore, SINR_k becomes small and the Shannon capacity of k decreases. The coupled constraints $\{g_f(w_k, w_{-k}) \leq 0\}$ guarantee that the SINR of each macro-user terminal inside the neighborhood $\mathcal{N}_k$ is above a certain threshold η, i.e., for any $f \in \mathcal{N}_k$,

$$\text{SINR}_f = \frac{p_{ff}w_f}{\sum\limits_{\ell \in \mathcal{N}_k \backslash \{f\}} p_{\ell f}w_\ell + \sigma_f^2} \geq \eta$$

$$\iff \quad g_f(w_k, w_{-k}) \triangleq \sum\limits_{\ell \in \mathcal{N}_k \backslash \{f\}} p_{\ell f}w_\ell - \frac{p_{ff}}{\eta}w_f + \sigma_f^2 \leq 0 \tag{1.15}$$

$\square$

Example 1.2. *(**Economic trading in geographical networks**)* A second example is economic trading with geographical network topologies where there exist several factories producing similar products and selling them to some common market place. This problem is also referred to as the networked Cournot competition problem in the literature [24, 25]. Figure 1.4 shows a simple Cournot network where the circles denote the factories and the rectangulars denote the common markets. We consider that each factory k has an individual cost function $J_k(w_k; w_{-k})$, defined as the production cost minus the revenue earned from the market places. In this case, the action w_k is a vector where each entry of w_k represents a quantity of products to be sold at one connected market place. Therefore, each factory k

7
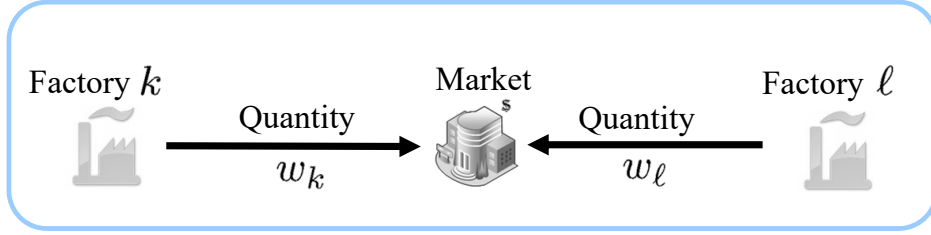
(a)
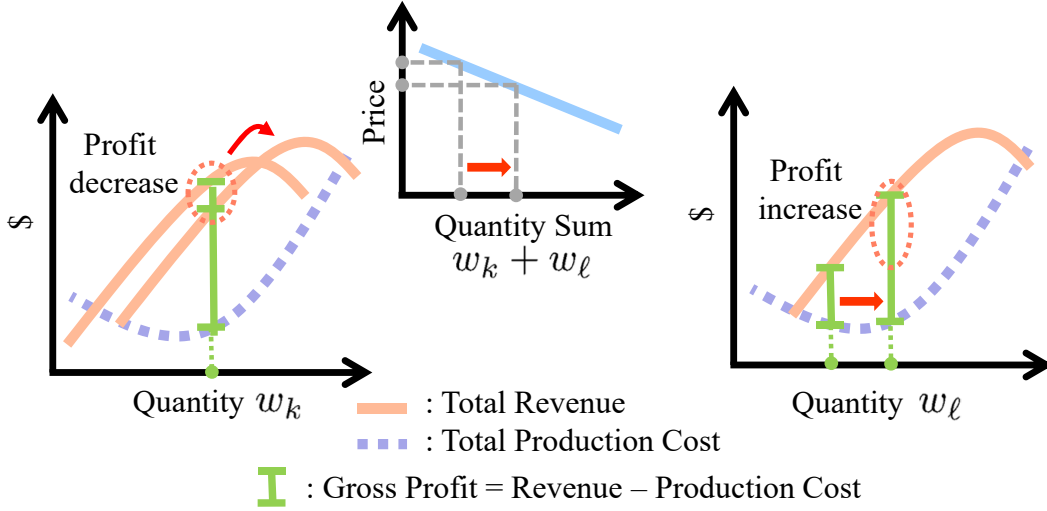
(b)

Figure 1.4: (a) Economic trading with two factories connected to a common market. (b) The impact on the market price and the gross profits of factories when the production quantity of factory ℓ increases.

solves the following optimization problem to find the optimal w_k:

$$\min_{w_k \in \mathbb{R}^{M_k}} \quad J_k(w_k; w_{-k})$$

$$\text{subject to} \quad w_k(m) \geq 0, \quad m = 1, \ldots, M_k, \quad r_\ell(w_k, w_{-k}) \leq \eta, \quad \ell \in \mathcal{M}_k \tag{1.16}$$

where $\mathcal{M}_k$ is the set of market places that factory k is connected to. The parameter $r_\ell(w_k, w_{-k}) = \sum_{k=1}^{N} w_k(\ell)$ is the total quantity of products sold in the market place ℓ. The individual cost function $J_k(\cdot)$ is given by

$$J_k(w_k; w_{-k}) = c(w_k) - \sum_{\ell \in \mathcal{M}_k} p(r_\ell(w_k, w_{-k})) w_k(\ell) \tag{1.17}$$

8

where $c(\cdot)$ is the production cost function and $p(\cdot)$ is the price function. Therefore, the first and second terms on the right-hand side of (1.17) represent the total production costs and the total revenue of factory k, respectively. Note that here we consider uniform production cost functions and uniform price functions for all factories and market places, which of course can be easily extended to more complex cases. One example for $c(\cdot)$ and $p(\cdot)$ is provided in Fig. 1.4 where we assume the production cost function is quadratic and the price function is linear. We also show the impact on the market price and the gross profits of factories when the production quantity of factory ℓ increases. In this situation, the profit of factory ℓ increases. At the same time, the price in the market decreases since $w_k + w_\ell$ becomes large, which affects the total revenue of factory k and makes its profit decreases. In formulation (1.16), the constraints enforce nonnegative entries in w_k and that for each market place there is some upper limit capacity η on the total amount of products to sell. $\qquad\square$

1.3.1 Generalized Nash Equilibrium

When we consider multi-objective optimization problems of the form (1.10), the concept of Nash equilibria is a desired and stable solution since at this location no agent can benefit by unilaterally deviating from the equilibrium.

Definition 1.1. *(**Nash equilibrium**) We say that a global action vector $w^\star \triangleq col\{w_1^\star, ..., w_N^\star\}$ is a Nash equilibrium if for each agent k, we have $w_k^\star \in \mathcal{S}_k(w_{-k}^\star)$ and*

$$J_k(w_k^\star; w_{-k}^\star) \leq J_k(w_k'; w_{-k}^\star) \tag{1.18}$$

for any feasible action $w_k' \in \mathcal{S}_k(w_{-k}^\star)$. $\qquad\square$

Therefore, a Nash equlibrium $w^\star$ satisfies that

$$w_k^\star \in \arg \min_{w_k \in \mathcal{S}_k(w_{-k}^\star)} J_k(w_k; w_{-k}^\star) \tag{1.19}$$

for each k. In other words, if a global action vector $w^\star$ is a Nash equlibrium, then the actions $\{w_k^\star\}$ simultaneously solve the multi-objective problem for each agent k in (1.10). Once the network operates at some Nash equilibrium $w^\star$, the network becomes stable, which

is defined as all agents achieving their own minimal cost values given the neighbors' actions $w_{-k}^\star$, and therefore, have no incentive to deviate from $w_k^\star$. Problem (1.10) is known as the generalized Nash equilibrium problem (GNEP) in the literature [26–28] since it generalizes the traditional Nash equilibrium problem with fixed feasible set $\mathcal{S}_k$ instead of general $\mathcal{S}_k(w_{-k})$ depending on w_{-k}. Throughout this dissertation, we will focus on a significant class of GNEP where the agents' actions are required to satisfy *shared* common constraints. This scenario arises naturally in many applications, e.g., Examples 1.1 and 1.2, and has more complete theoretical results than the general GNEP [27]. This specific class of GNEPs is sometimes referred to jointly convex GNEPs in the literature.

Definition 1.2. (*Jointly convex GNEP*) *Suppose that for each k, the individual cost function $J_k(w_k; w_{-k})$ is convex in w_k and the feasible set $\mathcal{S}_k(w_{-k})$ is closed and convex. We say that agents are solving a jointly convex GNEP if they have shared common constraints, i.e., there exists a constraint set $\mathcal{S}$ such that for each k we have*

$$\mathcal{S}_k(w_{-k}) = \{w_k : (w_k, w^{-k}) \in \mathcal{S}\} \tag{1.20}$$

where

$$w^{-k} \triangleq \mathrm{col}\{w_\ell; \ell \neq k\} \tag{1.21}$$

collects all actions in the network except w_k. $\qquad\square$

Equation (1.20) describes that the actions of agents are subject to a common feasible set $\mathcal{S}$. For example, let us consider a fully-connected network where each agent is a neighbor of all other agents. If the feasible set $\mathcal{S}_k(w_{-k})$ for each agent k is defined explicitly by inequalities $\{g_{k,q}(w_k, w^{-k}) \leq 0\}$ for $q = 1, ..., Q$, then the shared constraints mean that for all q, we have

$$g_{1,q}(w_1, w^{-1}) = g_{2,q}(w_2, w^{-2}) = ... = g_{N,q}(w_N, w^{-N}) = g_q(w) \tag{1.22}$$

where $w \triangleq \mathrm{col}\{w_1, ..., w_N\}$. In this case, the feasible set $\mathcal{S}$ is $\{w : g_q(w) \leq 0, q = 1, ..., Q\}$.

Let us consider the jointly convex GNEP and assume $\{J_k(w_k; w_{-k})\}$ for all k are once-differentiable. We first recall the minimum principle which states a fundamental optimality condition for convex optimization problems [29].

Definition 1.3. (***Minimum principle***) *Consider an optimization problem in the following form:*

$$\min_{x \in \mathcal{K}} \quad J(x) \tag{1.23}$$

where $J(x)$ is a convex and differentiable function and $\mathcal{K}$ is a convex set. Then, a feasible point $x^\star \in \mathcal{K}$ is an optimal solution if and only if

$$(x^a - x^\star)^\mathsf{T} \nabla_x J(x^\star) \geq 0, \qquad \forall x^a \in \mathcal{K} \tag{1.24}$$

$\square$

Then, from (1.19) and by (1.24), the generalized Nash equilibrium $w^\star \in \mathcal{S}$ should satisfy

$$(w_k^a - w_k^\star)^\mathsf{T} \nabla_{w_k} J_k(w_k^\star; w_{-k}^\star) \geq 0, \qquad \forall w_k^a \in \mathcal{S}_k(w_{-k}^\star) \tag{1.25}$$

If we sum (1.25) over all agents, we get

$$\sum_{k=1}^N (w_k^a - w_k^\star)^\mathsf{T} \nabla_{w_k} J_k(w_k^\star; w_{-k}^\star) = (w^a - w^\star)^\mathsf{T} F(w^\star) \geq 0, \qquad \forall w^a \in \mathcal{S} \tag{1.26}$$

where

$$w^a \triangleq \mathrm{col}\{w_1^a, ..., w_N^a\} \tag{1.27}$$

$$F(w) \triangleq \mathrm{col}\{\nabla_{w_1} J_1(w_1; w_{-1}), ..., \nabla_{w_N} J_N(w_N; w_{-N})\} \tag{1.28}$$

Equation (1.26) is called a Variational Inequality (VI), which focuses on the problem of finding the optimal $w^\star$. It is shown in [30] that the optimal VI solution $w^\star$ to (1.26) is a solution to (1.19), and therefore is a generalized Nash equilibrium to the multi-objective optimization problem (1.10) with the jointly convex assumption. The reverse direction is generally not true since if we scale some $\{J_k(w_k; w_{-k})\}$ in (1.10), the solution set of the new GNEP does not change but we get a weighted summation in (1.26), which may lead to different VI solutions. Consequently, not all solutions of (1.10) can be obtained by solving the variational inequality problem (1.26). These VI solutions are referred to variational equilibria or normalized equilibria. A variational equilibrium has a special interpretation in

11

terms of Lagrange multipliers of the corresponding KKT conditions of the GNEP (1.10); that is, it corresponds to the case when the Lagrange multipliers for all agents are the same — see [31] for more details. Furthermore, one sufficient condition for the existence of the solutions for the jointly convex GNEP problem is to ensure that the VI solutions in (1.26) exist, which is guaranteed by the following strongly-monotone condition on $F(w)$: for any two action vectors $w = w^\circ$ and $w = w^\bullet$, we must have [32, pp. 155]

$$(w^\circ - w^\bullet)^\mathsf{T} [F(w^\circ) - F(w^\bullet)] \geq \nu \|w^\circ - w^\bullet\|^2 \tag{1.29}$$

for some positive constant ν. We note that the solutions of the GNEP are usually non-unique. However, the solution of the VI problem (1.26) is unique under the coerciveness property [?, p. 14], which means that for some $w^{\text{ref}} \in \mathbb{R}^M$, we have

$$\lim_{\|w\| \to \infty} \frac{\left[F(w) - F(w^{\text{ref}})\right]^\mathsf{T} (w - w^{\text{ref}})}{\|w - w^{\text{ref}}\|} = \infty \tag{1.30}$$

The coerciveness property is also guaranteed by the strongly-monotone condition (1.29) since setting $w^\circ = w$ and $w^\bullet = w^{\text{ref}}$ we get

$$\lim_{\|w\| \to \infty} \frac{\left[F(w) - F(w^{\text{ref}})\right]^\mathsf{T} (w - w^{\text{ref}})}{\|w - w^{\text{ref}}\|} \geq \lim_{\|w\| \to \infty} \nu \|w - w^{\text{ref}}\| = \infty \tag{1.31}$$

As a result, the strongly-monotone condition ensures the existence of the VI solutions (and thus the generalized Nash equilibria) and the uniqueness of the VI solution. In order to avoid redundancy, here we neglect the proofs of these results, which can be obtained by similar arguments to those in Appendix 2.A. We summarize the solutions of jointly convex GNEP and VI problems in Fig. 1.5.

1.3.2 Pareto Optimality

Another notion of optimality is Pareto optimality, which is useful to compare with the aforementioned Nash equilibrium concept. For problems in the form (1.10), a global action vector, say, $w^o \triangleq \text{col}\{w_1^o, ..., w_N^o\}$, is regarded as Pareto optimal if none of the individual costs of agents can be improved without deteriorating at least one other individual cost.
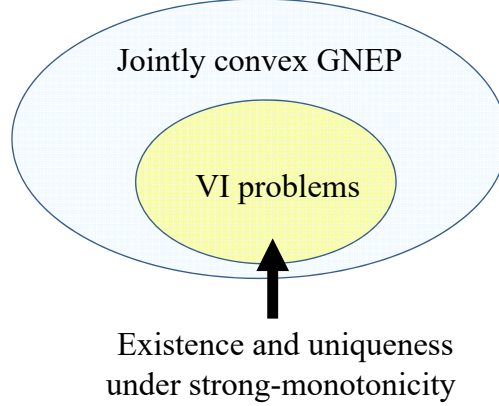
Figure 1.5: Illustration of the solutions of jointly convex GNEP and VI problems.

Definition 1.4. (***Pareto-optimal solution***) *We say a decision vector $w^o \triangleq col\{w_1^o, ..., w_N^o\}$ is Pareto-optimal if all $\{w_k^o\}$ are feasible and there does not exist another $w^\bullet \triangleq col\{w_1^\bullet, ..., w_N^\bullet\}$ with feasible $\{w_k^\bullet\}$ such that $J_k(w_k^\bullet; w_{-k}^\bullet) \leq J_k(w_k^o; w_{-k}^o)$ for all $k = 1, ..., N$ and $J_\ell(w_k^\bullet; w_{-k}^\bullet) < J_\ell(w_k^o; w_{-k}^o)$ for at least one index $\ell \in \{1, ..., N\}$.* $\square$

The Pareto-optimal solutions are also called Pareto-efficient. This is because if an action vector w is not Pareto-optimal, then there exists $w' \neq w$ which improves at least one agent' individual cost without making other agents' individual costs worse. We note that one useful technique to determine Pareto-optimal solutions is scalarization [16, 33, 34]. In this method, we formulate the following optimization problem with weighted aggregate cost functions:

$$\min_w \quad J^{\mathrm{par}}(w) \triangleq \sum_{k=1}^N a_k J_k(w_k; w_{-k}) \tag{1.32}$$

$$\text{subject to} \quad w_k \in \mathcal{S}_k(w_{-k}), \quad k = 1, \ldots, N$$

where $\{a_k\}$ are positive parameters. Given the $\{a_k\}$, every minimizer for the above optimization problem corresponds to a Pareto-optimal solution. Note that we can get different Pareto-optimal solutions by varying the parameters $\{a_k\}$ and resolving problem (1.32). We remark that as discussed in [33], some limiting Pareto-optimal solutions may not be obtained by the scalarization technique. Furthermore, a Nash equilibrium $w^\star$ that solves (1.10) for all k can be Pareto-optimal if $w^\star$ is also a solution to (1.32). However, a Nash equilibrium,

13
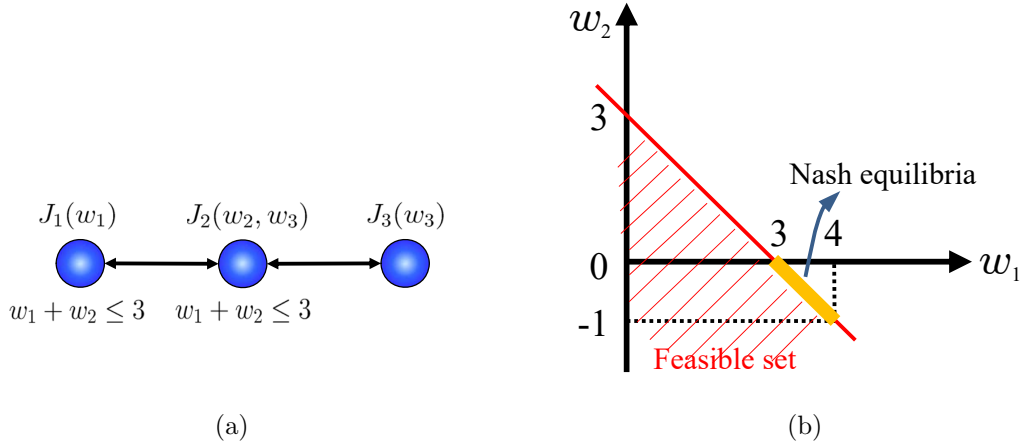
Figure 1.6: (a) Line topology with 3 agents. (b) Nash equilibria set for agents 1 and 2.

which is stable, is not necessarily Pareto-optimal or Pareto-efficient, and similarly, a Pareto-optimal solution is not necessarily a Nash equilibrium. We will encounter this situation later in Chapter 4, where we specifically consider discrete action values, and depending on the parameters in the game setting, the Nash equilibrium may or may not coincide with the Pareto-optimal solution. Let us consider the following example to illustrate the concepts of Nash equilibria and Pareto-optimal solutions.

Example 1.3. *(**Line topology**)* Suppose we have a three-agent network with a line topology as shown in Fig. 1.6a. Agents 1, 2, and 3 seek to solve the following constrained multi-objective optimization problems:

$$
\begin{cases}
\min_{w_1} \ J_1(w_1) \triangleq (w_1 - 4)^2 & \text{subject to} \ \ w_1 + w_2 \le 3 \\
\min_{w_2} \ J_2(w_2, w_3) \triangleq (w_2 + w_3)^2 & \text{subject to} \ \ w_1 + w_2 \le 3 \\
\min_{w_3} \ J_3(w_3) \triangleq w_3^2 &
\end{cases}
\tag{1.33}
$$

We note that agents 1 and 2 have a common and coupled constraint $w_1 + w_2 \le 3$, and the individual cost function of agent 2 depends on the action of agent 3. Furthermore, since in problem (1.33) the individual cost functions are convex and the constraints are affine, it satisfies the Slater's constraint qualification and thus we can utilize the Karush-Kuhn-Tucker (KKT) conditions [33] to find the Nash equilibria and Pareto-optimal solutions. From

14

Def. 1.1 and (1.19), a Nash equilibrium of problem (1.33), denoted by $w^\star \triangleq \text{col}\{w_1^\star, w_2^\star, w_3^\star\}$, should simultaneously satisfy

$$w_1^\star = \arg\min_{w_1} \ (w_1 - 4)^2 \quad \text{subject to} \ \ w_1 + w_2^\star \leq 3 \tag{1.34}$$

$$w_2^\star = \arg\min_{w_2} \ (w_2 + w_3^\star)^2 \ \ \text{subject to} \ \ w_1^\star + w_2 \leq 3 \tag{1.35}$$

$$w_3^\star = \arg\min_{w_3} \ w_3^2 \tag{1.36}$$

From (1.36), it is obvious for agent 3 that $w_3^\star = 0$ since $J_3(w_3) = w_3^2$ does not depend on w_1 and w_2 and there is no constraint for agent 3. Once we obtain $w_3^\star = 0$, we require the following KKT conditions for agents 1 and 2 to hold:

$$\nabla_{w_1}(w_1 - 4)^2|_{w_1^\star} + \lambda_1 \nabla_{w_1}(w_1 + w_2^\star - 3)|_{w_1^\star} = 0 \iff 2(w_1^\star - 4) + \lambda_1 = 0 \tag{1.37}$$

$$\lambda_1 \geq 0 \tag{1.38}$$

$$\lambda_1(w_1^\star + w_2^\star - 3) = 0 \tag{1.39}$$

$$w_1^\star + w_2^\star - 3 \leq 0 \tag{1.40}$$

and

$$\nabla_{w_2}(w_2 + w_3^\star)^2|_{w_2^\star} + \lambda_2 \nabla_{w_2}(w_1^\star + w_2 - 3)|_{w_2^\star} = 0 \iff 2w_2^\star + \lambda_2 = 0 \tag{1.41}$$

$$\lambda_2 \geq 0 \tag{1.42}$$

$$\lambda_2(w_1^\star + w_2^\star - 3) = 0 \tag{1.43}$$

$$w_1^\star + w_2^\star - 3 \leq 0 \tag{1.44}$$

where λ_1 and λ_2 are Lagrange multipliers. From (1.37)–(1.39), we get

$$\lambda_1 = 2(4 - w_1^\star) \geq 0 \iff w_1^\star \leq 4 \tag{1.45}$$

$$2(4 - w_1^\star)(w_1^\star + w_2^\star - 3) = 0 \tag{1.46}$$

and similarly from (1.41)–(1.43) we have

$$\lambda_2 = -2w_2^\star \geq 0 \iff w_2^\star \leq 0 \tag{1.47}$$

$$2w_2^\star(w_1^\star + w_2^\star - 3) = 0 \tag{1.48}$$

15

From the constraint $w_1^\star + w_2^\star - 3 \leq 0$ in (1.40) and (1.44), we know it can only be either $w_1^\star + w_2^\star - 3 < 0$ or $w_1^\star + w_2^\star - 3 = 0$. Assume that it is the former case. Then, by (1.46) and (1.48) it must hold that $w_1^\star = 4$ and $w_2^\star = 0$, which contradicts the assumption $w_1^\star + w_2^\star - 3 < 0$. Therefore, we conclude that $w_1^\star + w_2^\star - 3 = 0$ and arrive at

$$w_1^\star \leq 4, \qquad w_2^\star \leq 0, \qquad w_1^\star + w_2^\star - 3 = 0 \tag{1.49}$$

The Nash equilibria $w^\star$ are the intersection set of points that satisfy the above conditions, as we illustrate in Fig. 1.6b. One way to describe this set of Nash equilibria is

$$\{(w_1^\star, w_2^\star, w_3^\star) : w_1^\star + w_2^\star = 3, w_2^\star \in [-1, 0], w_3^\star = 0\} \tag{1.50}$$

Now, let us denote the Pareto-optimal solutions of agents 1, 2, and 3 by $w^o \triangleq \text{col}\{w_1^o, w_2^o, w_3^o\}$. Instead of solving the optimization problem (1.33), w^o can be obtained by solving the following weighted optimization problem with aggregating the individual costs:

$$\min_{w_1, w_2, w_3} \quad J^{\text{par}}(w_1, w_2, w_3) \triangleq a J_1(w_1) + b J_2(w_2, w_3) + c J_3(w_3) \tag{1.51}$$

$$= a(w_1 - 4)^2 + b(w_2 + w_3)^2 + c w_3^2$$

$$\text{subject to} \quad w_1 + w_2 \leq 3$$

where $a, b, c > 0$. The solution w^o to problem (1.51) should satisfy the following KKT conditions:

$$\nabla_w \left(a(w_1 - 4)^2 + b(w_2 + w_3)^2 + c w_3^2 \right)\big|_{w^o} + \lambda \nabla_w (w_1 + w_2 - 3)\big|_{w^o} = 0 \tag{1.52}$$

$$\lambda \geq 0 \tag{1.53}$$

$$\lambda(w_1^o + w_2^o - 3) = 0 \tag{1.54}$$

$$w_1^o + w_2^o \leq 3 \tag{1.55}$$

where λ is the Lagrange multiplier. From (1.52) we have

$$\begin{bmatrix} 2a(w_1^o - 4) \\ 2b(w_2^o + w_3^o) \\ 2b(w_2^o + w_3^o) + 2c w_3^o \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 0 \iff \begin{cases} 2a(w_1^o - 4) + \lambda = 0 \\ 2b(w_2^o + w_3^o) + \lambda = 0 \\ 2b(w_2^o + w_3^o) + 2c w_3^o = 0 \end{cases} \tag{1.56}$$

16

The third equation on the right-hand side of (1.17) gives that

$$w_3^o = \frac{-b}{b+c} w_2^o \tag{1.57}$$

which can be used in the second equation to get

$$2b \left(w_2^o - \frac{b}{b+c} w_2^o \right) + \lambda = 0 \quad \Longleftrightarrow \quad \lambda = -\frac{2bc}{b+c} w_2^o \tag{1.58}$$

Therefore, the first equation becomes

$$2a(w_1^o - 4) = \frac{2bc}{b+c} w_2^o \tag{1.59}$$

Combining (1.54) and (1.58) we have

$$w_2^o(w_1^o + w_2^o - 3) = 0 \tag{1.60}$$

Now, using similar arguments, this can only happen when either $w_2^o = 0$ or $w_1^o + w_2^o - 3 = 0$. Assume it is the former case $w_2^o = 0$. From (1.59) we get $w_1^o = 4$, which however does not satisfy the constraint condition in (1.55). Therefore, we conclude that $w_1^o + w_2^o - 3 = 0$ or equivalently, $w_1^o = 3 - w_2^o$. Substituting into (1.59), we then arrive at

$$2a(-w_2^o - 1) = \frac{2bc}{b+c} w_2^o \quad \Longleftrightarrow \quad w_2^o = \frac{-ab - ac}{ab + ac + bc} \tag{1.61}$$

and thus

$$w_1^o = 3 + \frac{ab + ac}{ab + ac + bc}, \qquad w_3^o = \frac{-b}{b+c} \cdot \frac{-ab - ac}{ab + ac + bc} = \frac{ab}{ab + ac + bc} \tag{1.62}$$

Consequently, we obtain the unique Pareto-optimal solution for every set of parameters (a, b, c) as

$$(w_1^o, w_2^o, w_3^o) = \left(3 + \frac{ab + ac}{ab + ac + bc}, \frac{-ab - ac}{ab + ac + bc}, \frac{ab}{ab + ac + bc} \right) \tag{1.63}$$

Comparing (1.63) with (1.50), we observe that since $b > 0$ and $c > 0$, the Pareto-optimal solutions obtained in (1.63) are not Nash equilibria. This means the Pareto-optimal solutions in (1.63) are not stable for agents targeting to minimize their own individual costs since, for example, agent 3 has the tendency to choose $w_3 = 0$ instead of $w_3^o = ab/(ab + ac + bc) > 0$ to obtain a smaller $J_3(w_3)$. $\qquad\square$

1.4   Organization

Therefore, this dissertation considers multi-objective learning over networks according to the general formulation of (1.10). From the previous example, we can see that Pareto-optimal solutions may not be stable in terms of the network dynamics since at these locations, agents still have incentives to deviate and to choose other actions. In comparison, the Nash equilibria are stable since no agent has an incentive to unilaterally deviate while Pareto-optimal solutions ensure efficient action vectors since there is no other action vector that can improve the cost of an individual agent without worsening the cost of some other agent. We will first purse stability and focus on solving Nash equilibria in Chapters 2 and 3. In Chapters 4, we will discuss a situation when Nash equilibria are not Pareto-optimal. In that case, we will explain how to design additional mechanisms so that the Nash equilibria in the modified multi-objective problem can approach to the Pareto-optimal solutions so that agents can arrive at a stable and efficient action vector.

In this dissertation, our main focus is on developing fully-distributed learning to solve the multi-objective formulation (1.10) where agents are only allowed to interact locally with the neighbors over the network topology. For the same reasons explained in Sec. 1.2 (scalability, robustness, privacy considerations), we do not consider centralized strategies, as already proposed in [35–39], which would require a fusion center collecting data from across the network. We are particularly interested in stochastic environments where the exact gradient information $\nabla_{w_k^\mathsf{T}} J_k(\cdot)$ is unavailable and can only be replaced by some approximate $\widehat{\nabla_{w_k^\mathsf{T}}} J_k(\cdot)$. The organization of this dissertation is as follows:

- **Chapter 2:** We consider a general formulation of multi-objective optimization problems over network topologies where agents seek to minimize their individual costs subject to constraints that are locally coupled. The coupling arises because the individual costs and the constraints can be dependent on actions by other agents in the neighborhood. In these types of problems, the Nash equilibrium is a desired and stable solution since at this location no agent can benefit by unilaterally deviating from the equilibrium. We therefore focus on developing distributed online strategies that enable

agents to approach the Nash equilibrium. We illustrate an application of the results to a stochastic version of the network Cournot competition problem, which arises in a variety of useful problems such as in modeling economic trading with geographical considerations, power management over smart grids, and resource allocation protocols.

- **Chapter 3:** Using the formulation developed in Chapter 2, we then extend earlier contributions on adaptive networks, which generally assume that the agents work together for a common global objective or when they observe data that is generated by a common target model or parameter vector. We relax this condition and consider a broader scenario where individual agents may only have access to partial information about the global target vector, i.e., each agent may be sensing only a subset of the entries of the global target, and the number of these entries can be different across the agents. We develop cooperative distributed techniques where agents are only required to share estimates of their common entries and still can benefit from neighboring agents. Since agents' interactions are limited to exchanging estimates of select few entries, communication overhead is significantly reduced.

- **Chapter 4:** In this chapter, we examine the behavior of adaptive networks where information sharing is subject to a positive communication cost over the edges linking the agents. In the absence of any incentives to cooperate, we show that the dominant strategy for all agents is for them not to participate in the sharing of information and thus Pareto inefficiency arises. We then develop a reputation protocol to summarize the opponent's past actions into a reputation score, which can then be used to form a belief about the opponent's subsequent actions. It is shown that the proposed reputation protocol can entice agents to cooperate and enhance the overall social benefit of the network. We perform a detailed mean-square-error analysis of the evolution of the network, which is verified by numerical simulations. The work has been successfully applied to information sharing over cognitive networks, and potential applications include online learning under communication bandwidth and/or latency constraints, and over social learning networks when the delivery of opinions involves some costs.

19

- **Chapter 5:** As an alternative to the reputation protocols considered in Chapter 4, we go beyond interactions among single agents and develop a clustering technique to entice cooperation by clusters of agents. In this method, agents are allowed to decide with whom to cluster and share information. The clustering concept is widely studied in the social sciences and game theory. It enables agents to drive their cooperative behavior by selecting their partners according to whether they can help them reduce their utility costs. One challenge is to select utility functions that can drive the clustering operation. Recent results on the performance of adaptive networks are exploited to great effect for this purpose. Another challenge is to deal with the intertwinement of the clustering and learning dynamics. In my design, each agent first evaluates the expected cost of its possible actions and decides on whether to propose cooperation to the other agent; if two agents agree on cooperation, they establish a link and become part of the same larger cluster. After the clusters are formed, agents share and process information over their sub-networks. We derive the conditions for clusters to unite under various cluster properties.

CHAPTER 2

Stochastic Generalized Nash Equilibrium Problems

This chapter examines a stochastic formulation of multi-objective optimization problems over network topologies, which is referred to as the stochastic generalized Nash equilibrium problem (GNEP) in the literature. In this formulation, agents seek to minimize their individual costs subject to constraints that are locally coupled in an environment of unknown statistical distribution. We focus on fully-distributed online learning by agents and employ penalized individual cost functions to deal with coupled constraints. Three stochastic gradient strategies are developed with constant step-sizes. We allow the agents to use heterogeneous step-sizes and show that the penalty solution is able to approach the Nash equilibrium in a stable manner within $O(\mu_{\max})$, for small step-size value $\mu_{\max}$ and sufficiently large penalty parameters. The operation of the algorithm is illustrated by considering the network Cournot competition problem.

2.1 Introduction

The generalized Nash equilibrium problem (GNEP) refers to a setting where each agent in a collection of agents seeks to minimize its own cost function subject to certain constraints *and* where both the cost function and the constraints are generally dependent on the actions selected by the other agents [26–28, 40–42]. The GNEP was first formally introduced in [40] and was called a social equilibrium problem. A special case of GNEPs was considered in the work [35] where all agents shared common constraints. GNEPs arise naturally in the modeling of many applications, ranging from market liberalization of electricity [42, 43], to natural gas [44], telecommunications [45], femto-cell power allocation [22], environmental

pollution control [46], and cloud computing [47, 48]. Useful overviews on GNEPs appear in [27, 28].

In these types of problems, the Nash equilibrium is a desired and stable solution since at the Nash equilibrium no agent can benefit by unilaterally deviating from the solution. However, Nash equilibrium solutions may not exist or may not be unique. For instance, it was shown in [26, 41] that the solution set of a GNEP can be characterized by solving a quasi-variational inequality (QVI), and it is rare that explicit results in QVIs can be utilized in GNEPs. Still, there is one common and important class of GNEPs that can be partially solved by solving a variational inequality (VI) [27, 29]. In this chapter, we focus on GNEPs with shared and coupled constraints since the theory of variational inequalities (VI) is more mature and has more useful results than the theory of quasi-variational inequalities (QVI).

In general, GNEP formulations do not admit closed-form solutions and many algorithms have been proposed to compute the solutions numerically. For example, GNEPs can be reformulated and solved using Nikaido-Isoda (NI) functions. Minimizing the NI can be achieved by means of gradient-descent algorithms [36] or relaxtion-based algorithms [37]. Likewise, using the Karush-Kuhn-Tucker (KKT) conditions, GNEPs can be solved numerically, as demonstrated in [38]. One can also resort to penalty-based reformulations where the original cost function is modified by including a penalty term. The purpose of the penalty term is to assign large penalties to deviations from the constraints. The works in [49, 50] consider exact penalty functions and focus on updating the penalty parameters incrementally until a certain stopping rule is satisfied.

In all these prior works [40]– [50], the individual cost functions are assumed to be *deterministic*. This means that, when seeking GNEP solutions, we are able to acquire exactly the NI functions or the gradient vectors as necessary. However, when the agents are subject to randomness in the environment, it is customary to define the cost functions in terms of *expectations* of certain loss functions. The expectation operations are in relation to the distribution of the random data, which is rarely known beforehand. This *stochastic* type of Nash games arises in many practical applications, e.g., in the transportation model of [51] and the signal transmission model for wireless networks in [22]. To deal with stochasticity,

Table 2.1: Comparing with Existing Works for Distributed Stochastic Problems.

	Optimization Target	Constraints	Feasibility Approach	Step-Sizes	Iterates Feasibility	Tracking Ability
Regularized SA [53]	Monotone individual cost	Shared and coupled	Using projection	Heterogeneous decaying	Feasible	No
Penalized Diffusion [54]	Strongly-convex aggregate cost	Decoupled	Using penalty functions	Uniform constant	Asymptotically feasible	Yes
This Chapter	Strongly-monotone individual costs	Shared and coupled	Using penalty functions	Heterogeneous constant	Asymptotically feasible	Yes

the sample average approximate (SAA) method was proposed in [52] to approximate the expectation of the individual cost functions. However, in this method, the equilibrium solutions are learned in an off-line manner and the GNEP needs to be re-solved for every given batch of samples.

In order to attain continuous learning in an online manner, the stochastic approximation (SA) method is a more suitable approach for differentiable cost functions, where the true gradient vectors are replaced by approximations. One stochastic implementation along these lines is considered in [53] albeit with a *vanishing* step-size parameter. The use of step-sizes that decay to zero is problematic in scenarios that require continuous adaptation and learning.

For example, in nonstationary environments, the Nash equilibrium will drift with time due to changes in the statistical distribution of subsequent changes in the locations of the minimizers of the cost functions. When the step-size approaches zero, as is the case with the rules considered in [55–58], adaptation stops and the stochastic gradient algorithm loses its ability to track the drift. The approach in [59] employs a decaying step-size to track the evolving minimizer of a non-stationary objective. However, in that work, the optimal sublinear regret is obtained under the condition that the variation budget V_T of the time-varying loss functions is sublinear with time. This condition implies that the variation in the loss functions should diminish with time, which is not applicable in the case where the

minimizer of the cost function drifts continuously. One example that does not satisfy the variation budget condition is discussed in [54]. In comparison, it is well-known that constant step-size adaptation in inherently capable of tracking moderate drifts due to nonstationarity in the data — see, e.g., the analysis in [16, 60, 61].

We therefore focus in this chapter on online and fully-distributed learning to solve the stochastic GNEPs where agents are only allowed to interact locally with their neighbors. We assume that such interactions are confined to neighboring agents over the network topology and are subject to some coupled constraints shared by all neighbors. That is, in addition to the stochastic setting, we build one additional topology layer on top of conventional GNEPs with shared constraints. One example for such stochastic GNEP scenarios linked to a geometric topology would be the femto-cell power allocation problem considered in [22], where distributed algorithms are proposed and designed for this specific application. In this chapter, we study general distributed learning strategies for the solution of GNEPs by networked agents. Motivated by results from [49, 50, 54, 62], we first resort to penalty functions to deal with the constraints in stochastic GNEPs. The penalty reformulation helps avoid the high computational complexity of conventional NI-based approaches or the requirement of projection steps. Traditionally, penalty methods focus on selecting penalty parameters [49, 50]. However, in order to cope with the stochastic nature of GNEPs, we fix the penalty parameters at constant but sufficiently large values, in a manner similar to [54, 62], and study the resulting performance under stochastic environments. We also focus on the use of *constant* step-sizes in the stochastic approximation methods to enable continuous adaptation and learning. When this is done, gradient noise seeps into the operation of the algorithm. By gradient noise we mean the difference between the true gradient vector and its approximation. In decaying step-size implementations, this gradient noise component is annihilated over time by the diminishing step-size parameter at the expense of a deteriorating tracking performance. In contrast, in the constant step-size implementation, the gradient noise process is persistently present in the operation of the algorithm. One main challenge in our analysis is to establish that the stochastic-gradient implementation is able to keep the influence of gradient noise under check and to deliver an accurate estimation of the Nash

equilibrium. Arriving at these conclusions for networked agents is one key contribution of this chapter. In Table 2.1 we list a summary of properties comparing our results to two other existing works for distributed stochastic problems.

We remark that there exist other techniques in the stochastic optimization literature to solve problems with the variational inequalities. For example, the works [63–65] consider a dual-averaging method, which requires the solution of an optimization problem at each iteration; this formulation would be useful in situations when the optimization problem can be solved in closed form. References [66, 67] consider stochastic mirror-based approaches, which assume the gradient noise has bounded variance. It is worth noting that the methods in these earlier references are not directly applicable to GNEP with shared constraints over networks, which is one critical contribution in this article.

In the simulations section, we will illustrate the theoretical results and apply the proposed algorithms to the constrained network Cournot competition problem, which is widely used in applications such as economic trading with geographical considerations, power management over smart grids, and resource allocation [25, 42, 43, 68]. We will assume there that factories and markets are connected in a Cournot network and suffer from some randomness in the parameters. We will see that the numerical results will match well with our theoretical analysis. We will also compare our algorithms with two projection-based algorithms from [53]: the distributed Arrow-Hurwicz method and the iterative Tikhonov regularization method. Table 2.2 provides a summary of the symbols used in the article for ease of reference.

2.2 Problem Setup

Consider a connected network of N agents indexed by the set $\mathcal{N} = \{1, ..., N\}$. The neighborhood of each agent k, denoted by $\mathcal{N}_k$, includes agent k and the neighboring agents connected to k. We denote the action of each agent k by a vector $w_k \in \mathbb{R}^{M_k}$ and associate with k an individual risk function denoted by $J_k(\cdot)$. The argument of $J_k(\cdot)$ does not depend solely on w_k but also on the action vectors of the neighboring agents. We collect the actions of all

Table 2.2: Summary of Main Symbols and Notation.

Symbol	Meaning	Equation
$J_k(\cdot), Q_k(\cdot)$	Individual cost and loss functions	(2.7)
$J_k^p(\cdot)$	Penalized individual cost function	(2.27)
$p_k(\cdot)$	Aggregated penalty function	(2.28)
$F(w)$	Block gradient vector	(2.12)
$F^p(w)$	Penalized block gradient vector	(2.34)
ν	Strongly-monotone parameter	(2.13)
δ	Lipschitz parameter	(2.14)
ρ	Penalty parameter	(2.27)
γ_k	Lipschitz gradient parameter	(2.64)
δ_p	Parameter related to γ_k	(2.65)
$\mu_{\max}$	Maximal step-size	(2.66)
t	Difference parameter for step-sizes	(2.67)
α	Gradient noise parameter	(2.76)
ν', ν''	Weighted strongly-monotone parameters	(2.68), (2.69)

agents in the neighborhood $\mathcal{N}_k$ into the block vector:

$$w^k = \mathrm{col}\{w_\ell; \ell \in \mathcal{N}_k\} \in \mathbb{R}^{M^k} \tag{2.1}$$

and the actions of all agents in the network $\mathcal{N}$ into:

$$w = \mathrm{col}\{w_1, \ldots, w_N\} \in \mathbb{R}^M \tag{2.2}$$

where

$$M^k \triangleq \sum_{\ell \in \mathcal{N}_k} M_\ell, \quad M \triangleq \sum_{\ell=1}^{N} M_\ell \tag{2.3}$$

For convenience, we also introduce the notation

$$w_{-k} \triangleq \mathrm{col}\{w_\ell; \ell \in \mathcal{N}_k \setminus \{k\}\} \tag{2.4}$$

to collect the actions of all other agents in $\mathcal{N}_k$, with the exception of agent k. Using this notation, we shall sometimes write $J_k(w_k; w_{-k})$ instead of $J_k(w^k)$ in order to make the

dependence on w_k explicit. We consider that the action of each agent k should satisfy a set of local constraints:

$$h_{k,u}(w^k) = 0, \quad u = 1, \ldots, U_k, \tag{2.5}$$

$$g_{k,q}(w^k) \leq 0, \quad q = 1, \ldots, L_k \tag{2.6}$$

The local constraint functions $\{h_{k,u}(w^k), g_{k,q}(w^k)\}$ at agent k are assumed to be differentiable and known to agent k. We also assume that the equality constraint functions $\{h_{k,u}(w^k)\}$ are affine and the inequality functions $\{g_{k,q}(w^k)\}$ are convex in w^k. We further assume that the constraints are shared by the neighbors, i.e., if the argument of any $h_{k,u}(w^k)$ or $g_{k,q}(w^k)$ at node k contains the action of some neighbor $\ell \in \mathcal{N}_k$, then agent ℓ is subject to the same constraint function, i.e., it will hold that $h_{\ell,u'}(w^\ell) = h_{k,u}(w^k)$ or $g_{\ell,q'}(w^\ell) = g_{k,q}(w^k)$ for some u' and q'. Figure 2.1 illustrates this setting for a network topology with 5 agents. An example of shared constraints is $g_{1,1}(w^1) = g_{2,1}(w^2) = g_{3,1}(w^3) \leq 0$, which is shared by the connected agents $1, 2$ and 3. We note that while there is no direct link between agents 2 and 4, the actions for these agents are coupled through the intermediate agent 3. Therefore, in general, the actions of agents are affected explicitly by the neighbors and also implicitly by other agents in the network. This scenario is common in applications [27, 29, 35, 50]. Each agent k then seeks an optimal action vector that solves the following constrained optimization problem [52, 53, 69]:

$$\min_{w_k \in \mathbb{R}^{M_k}} \quad J_k(w^k) \triangleq \mathbb{E}_{\boldsymbol{x}_k} Q_k(w^k; \boldsymbol{x}_k)$$

$$\text{subject to} \quad h_{k,u}(w^k) = 0, \quad u = 1, \ldots, U_k$$

$$g_{k,q}(w^k) \leq 0, \quad q = 1, \ldots, L_k \tag{2.7}$$

where $J_k(w^k)$ is assumed to be differentiable and strongly-convex in w_k, $Q_k(\cdot)$ is a scalar-valued loss function for agent k, and the expectation is taken over the distribution of the random data $\boldsymbol{x_k}$. For example, if we consider power allocation in wireless heterogeneous networks, the individual cost function $J_k(w^k)$ for each femto-base station k can represent the Shannon capacity function with channel uncertainty. Moreover, one constraint of $g_{k,q}(w^k)$ shared by neighboring femto-base stations can be used to guarantee that the average signal-

27

to-interference and noise ratio (SINR) at macro-user terminals is above a certain threshold [22]. Problem (2.7) is known as the stochastic generalized Nash equilibrium problem (GNEP). For convenience, we collect all *distinct* individual constraints across all agents into a global set denoted by

$$\mathcal{S} \triangleq \{w; h_u(w) = 0, g_q(w) \leq 0, 1 \leq u \leq U, 1 \leq q \leq L\} \tag{2.8}$$

by removing the repeated shared constraints. We assume that $\mathcal{S}$ is nonempty, which means that at least one solution w exists that satisfies the constraints in $\mathcal{S}$ and implies that the GNEP in (2.7) is feasible for each agent. Let us denote the feasible set of (2.7) by

$$\mathcal{S}_k(w_{-k}) \triangleq \{w_k; h_{k,u}(w^k) = 0, g_{k,q}(w^k) \leq 0, 1 \leq u \leq U_k, 1 \leq q \leq L_k\} \tag{2.9}$$

Without loss of generality, we assume that the input (domain) of $\mathcal{S}_k(w_{-k})$ satisfies all constraints in $\mathcal{S}$ that are independent of w_k. Therefore, any $w_k \in \mathcal{S}_k(w^{-k})$ shall satisfy the remaining constraints in $\mathcal{S}$ that are related to w_k, i.e., for each agent k we have

$$\mathcal{S}_k(w_{-k}) = \mathcal{S}_k(w^{-k}) = \{w_k; (w_k, w^{-k}) \in \mathcal{S}\} \tag{2.10}$$

where

$$w^{-k} \triangleq \mathrm{col}\{w_\ell; \ell \in \mathcal{N} \setminus \{k\}\} \tag{2.11}$$

since the actions of the agents who are not neighbors of agent k will not appear in any argument of the constraint functions $h_{k,u}(w^k)$ and $g_{k,q}(w^k)$. The conclusion in (2.10) shows that the scenario considered in this chapter satisfies the condition of GNEP with general shared common constraints [27].

Our objective now is to derive distributed learning strategies by which agents can adaptively learn to solve (2.7) using local observations of the actions of neighboring agents.

In preparation for our development, we collect the individual gradient vectors of $\{J_k(w^k)\}$ with respect to each w_k^T into

$$F(w) \triangleq \mathrm{col}\{\nabla_{w_1^\mathsf{T}} J_1(w^1), ..., \nabla_{w_N^\mathsf{T}} J_N(w^N)\} \tag{2.12}$$

and assume that this block column vector satisfies the following properties.

Agent	Neighborhood	Individual Cost	Constraints
1	$\mathcal{N}_1 = \{1,2,3,5\}$	$J_1(w^1) = \|w_1 + w_5\|^2$	$g_{1,1}(w^1) = \|w_1\|^2 + \|w_2 + w_3\|^2 - 2 \leq 0$
2	$\mathcal{N}_2 = \{1,2,3\}$	$J_2(w^2) = \|w_2\|^2$	$g_{2,1}(w^2) = \|w_1\|^2 + \|w_2 + w_3\|^2 - 2 \leq 0$ $g_{2,2}(w^2) = \|w_2 - w_3\|^2 - 5 \leq 0$
3	$\mathcal{N}_3 = \{1,2,3,4\}$	$J_3(w^3) = \|w_3\|^2 \cdot \|w_4\|^2$	$g_{3,1}(w^3) = \|w_1\|^2 + \|w_2 + w_3\|^2 - 2 \leq 0$ $g_{3,2}(w^3) = \|w_2 - w_3\|^2 - 5 \leq 0$
4	$\mathcal{N}_4 = \{3,4,5\}$	$J_4(w^4) = \|w_3\| + \|w_4\|^2$	$h_{4,1}(w^4) = \mathbb{1}_{M_4}^{\mathsf{T}} w_4 + \mathbb{1}_{M_5}^{\mathsf{T}} w_5 - 1 = 0$
5	$\mathcal{N}_5 = \{1,4,5\}$	$J_5(w^5) = \|w_1\| \cdot \|w_5\|^2$	$h_{5,1}(w^5) = \mathbb{1}_{M_4}^{\mathsf{T}} w_4 + \mathbb{1}_{M_5}^{\mathsf{T}} w_5 - 1 = 0$
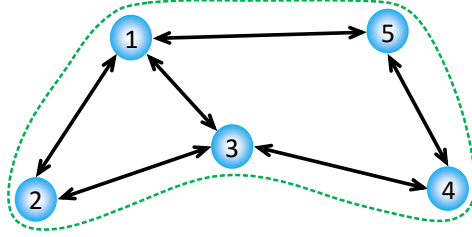
Figure 2.1: Illustration of the shared constraints over a network topology where $\mathbb{1}$ denotes the vector with all one entries.

Assumption 2.1. (*ν-Strongly Monotone*) *For any two action profiles $w = w^\circ$ and $w = w^\bullet$, it holds that*

$$(w^\circ - w^\bullet)^T [F(w^\circ) - F(w^\bullet)] \geq \nu \|w^\circ - w^\bullet\|^2 \tag{2.13}$$

for some positive constant ν. □

Assumption 2.2. (*δ-Lipschitz Continuous*) *The block column vector $F(w)$ is assumed to be Lipschitz continuous, i.e.,*

$$\|F(w^\circ) - F(w^\bullet)\| \leq \delta \|w^\circ - w^\bullet\| \tag{2.14}$$

for some positive constant δ. □

If we consider two action vectors w^a and w^b defined as:

$$w^a \triangleq \text{col}\{w_1, ..., w_k^\circ, ..., w_N\} \tag{2.15}$$

$$w^b \triangleq \text{col}\{w_1, ..., w_k^\bullet, ..., w_N\} \tag{2.16}$$

for some k, then using (2.13) we get

$$(w^a - w^b)^\mathsf{T}[F(w^a) - F(w^b)] = (w_k^\circ - w_k^\bullet)^\mathsf{T} \left[\nabla_{w_k^\mathsf{T}} J_k(w_k^\circ; w_{-k}) - \nabla_{w_k^\mathsf{T}} J_k(w_k^\bullet; w_{-k}) \right]$$

$$\geq \nu \|w_k^\circ - w_k^\bullet\|^2 \tag{2.17}$$

Therefore, Assumption 2.1 implies that each individual cost function $J_k(w^k)$ is strongly convex in w_k. Moreover, it holds that $\delta \geq \nu$ since from the Cauchy-Schwarz inequality we have

$$\nu \|w^\circ - w^\bullet\|^2 \leq (w^\circ - w^\bullet)^\mathsf{T}[F(w^\circ) - F(w^\bullet)]$$

$$\leq \|w^\circ - w^\bullet\| \cdot \|F(w^\circ) - F(w^\bullet)\|$$

$$\leq \delta \|w^\circ - w^\bullet\|^2 \tag{2.18}$$

Example 2.1. *(**Quadratic Risks**)* One useful example of a loss function is the quadratic loss, which can be expressed in the following form with the entries of $\boldsymbol{x}_k$ split into $\boldsymbol{x}_k \triangleq \{\boldsymbol{B}_k, \boldsymbol{b}_k, \varepsilon_k\}$:

$$Q_k(w^k; \boldsymbol{x}_k) = w^{k\mathsf{T}} \boldsymbol{B}_k w^k + \boldsymbol{b}_k^\mathsf{T} w^k + \boldsymbol{\varepsilon}_k$$

$$= \sum_{s \in \mathcal{N}_k} \sum_{\ell \in \mathcal{N}_k} w_s^\mathsf{T} \boldsymbol{B}_{s\ell}^k w_\ell + \sum_{\ell \in \mathcal{N}_k} \boldsymbol{b}_{k\ell}^\mathsf{T} w_\ell + \boldsymbol{\varepsilon}_k \tag{2.19}$$

where $\boldsymbol{B}_k$ is a random symmetric matrix of size $M^k \times M^k$, $\boldsymbol{b}_k$ is a random vector of size $1 \times M^k$, and $\boldsymbol{\varepsilon}_k$ is a random scalar variable with mean ε_k. In (2.19), we partitioned $\boldsymbol{B}_k$ and $\boldsymbol{b}_k$, respectively, into block matrices $\{\boldsymbol{B}_{s\ell}^k \in \mathbb{R}^{M_s \times M_\ell}\}$ and block vectors $\{\boldsymbol{b}_{k\ell} \in \mathbb{R}^{M_\ell \times 1}\}$ in conformity with the block structure of w^k. The random data $\{\boldsymbol{B}_k, \boldsymbol{b}_k, \boldsymbol{\varepsilon}_k\}$ are assumed to be independent of each other. Note that under (2.19), the gradient vector of $J_k(w^k)$ with respect to w_k^T is the $M_k \times 1$ vector given by

$$\nabla_{w_k^\mathsf{T}} J_k(w^k) = \sum_{\ell \in \mathcal{N}_k} 2B_{k\ell}^k w_\ell + b_{kk} \tag{2.20}$$

where we introduced the means $B_{k\ell}^k = \mathbb{E}\boldsymbol{B}_{k\ell}^k$ and $b_{kk} = \mathbb{E}\boldsymbol{b}_{kk}$. Collecting these individual gradient vectors we get

$$F(w) = Bw + b \tag{2.21}$$

where

$$B \triangleq \begin{bmatrix} 2B_{11}^1 & \cdots & 2B_{1N}^1 \\ \vdots & \ddots & \vdots \\ 2B_{N1}^N & \cdots & 2B_{NN}^N \end{bmatrix} \in \mathbb{R}^{M \times M}, \quad b \triangleq \begin{bmatrix} b_{11} \\ \vdots \\ b_{NN} \end{bmatrix} \in \mathbb{R}^{M \times 1} \tag{2.22}$$

Note that Assumption 2.1 will hold if there exists a positive constant ν such that for any $M \times 1$ vector a we have

$$a^\mathsf{T}\left(B - \nu I\right)a \geq 0 \quad \Longleftrightarrow \quad a^\mathsf{T}Ba \geq \nu\|a\|^2 \tag{2.23}$$

Since B is not necessarily symmetric, we know from [70, p. 259] that (2.23) holds if, and only if, the symmetric part of B satisfies:

$$\frac{1}{2}(B + B^\mathsf{T}) \geq \nu I \tag{2.24}$$

It follows from this condition that the largest singular value of B, denoted by $\sigma_{\max}$, should be greater than or equal to ν since

$$\sigma_{\max} = \|B\| \geq \left\|\frac{1}{2}(B + B^\mathsf{T})\right\| \geq \nu \tag{2.25}$$

From (2.21), it is easy to verify that Assumption 2.2 always holds for the quadratic loss function since

$$\|F(w^\circ) - F(w^\bullet)\| = \|B(w^\circ - w^\bullet)\| \leq \sigma_{\max}\|w^\circ - w^\bullet\| \tag{2.26}$$

$\square$

2.3 Stochastic Penalty-Based Learning

2.3.1 Penalty Approximation for Coupled Constraints

Solving the constrained optimization problem (2.7) is generally demanding and may not admit a closed-form solution. In this chapter, we resort to a penalty-based approach to

replace the original problem by an unconstrained optimization problem and then show that the solution to the penalized problem tends asymptotically with the penalty parameter to the desired solution to (2.7). Even more importantly, we will show that the penalty-based approach enables the agents to employ *adaptive* learning strategies, which instantaneously approximate the unknown random individual cost functions and endow the agents with the ability to track variations in the location of the Nash equilibrium due to changes that may occur in the constraint conditions or cost measures.

The main motivation for penalty methods is to assign a large penalty weight whenever constraints are violated and a smaller or zero weight when the constraints are satisfied [26,54,61,71]. More specifically, problem (2.7) is replaced by the following unconstrained formulation:

$$\min_{w_k \in \mathbb{R}^{M_k}} \quad J_k(w^k) + \rho p_k(w^k) \triangleq J_k^p(w^k) = J_k^p(w_k; w_{-k}) \tag{2.27}$$

where $\rho \geq 0$ is a penalty parameter, $p_k(w^k)$ denotes the penalty function for agent k and is assumed to be of the following aggregate form, with one penalty factor applied to each constraint:

$$p_k(w^k) = \sum_{u=1}^{U_k} \theta_{\mathrm{EP}}\left(h_{k,u}(w^k)\right) + \sum_{q=1}^{L_k} \theta_{\mathrm{IP}}\left(g_{k,q}(w^k)\right) \tag{2.28}$$

where $\theta_{\mathrm{EP}}(x)$ and $\theta_{\mathrm{IP}}(x)$ are convex functions. The equality penalty factor $\theta_{\mathrm{EP}}(x)$ returns zero value if the constraint is satisfied, i.e., when $h_{k,u}(w^k) = 0$, and introduces a large positive penalty if the constraint is violated, i.e., when $h_{k,u}(w^k) \neq 0$. For example, a continuous and differentiable choice for the equality penalty is the quadratic function:

$$\theta_{\mathrm{EP}}(x) = x^2 \tag{2.29}$$

Since $h_{k,u}(w^k)$ is affine, a convex choice of $\theta_{\mathrm{EP}}(\cdot)$ ensures the convexity of the function composition $\theta_{\mathrm{EP}}\left(h_{k,u}(w^k)\right)$. Similarly, the inequality penalty function $\theta_{\mathrm{IP}}(x)$ returns zero value if $g_{k,q}(w^k) \leq 0$, and introduces a large positive penalty if $g_{k,q}(w^k) > 0$. In the penalty method studied in [50], we get an exact Nash equilibrium solution to (2.7) as long as ρ is sufficiently large and we use the ℓ_1 penalty function [72]:

$$\theta_{\mathrm{IP}}^e(x) = \max\{0, x\} \tag{2.30}$$

32

However, using this penalty function makes the objective function in (2.27) non-differentiable, which limits the use of gradient-based adaptation rules [73]. To avoid this difficulty, we can employ the following half-quadratic penalty function [71], which is continuous, convex, non-decreasing, and once-differentiable:

$$\theta_{\text{IP}}(x) \triangleq \begin{cases} 0, & x \leq 0 \\ x^2/2, & x \geq 0 \end{cases} \tag{2.31}$$

Other choices for $\theta_{\text{IP}}(x)$ are of course possible, e.g., γ-norm [49], exponential and shifted logarithmic functions [74, 75], linear-quadratic functions [62], and others in [15, 54]. We note that a convex and nondecreasing choice of $\theta_{\text{IP}}(\cdot)$ results in a convex composite function $\theta_{\text{IP}}\left(g_{k,q}(w^k)\right)$ since $g_{k,q}(w^k)$ is convex. Consequently, the penalty function $p_k(w^k)$ defined in (2.28) is convex in w^k.

The penalized cost $J_k^p(w^k)$ is strongly-convex in w_k since $J_k(w^k)$ is strongly-convex in w_k, as seen in (2.17), and $p_k(w^k)$ is convex in w^k, and therefore in w_k. An action profile $w^\star = \text{col}\{w_1^\star, ..., w_N^\star\}$ that minimizes simultaneously all penalized costs $\{J_k^p(w^k)\}$ is called a Nash equilibrium for the penalized formulation (2.27), i.e., for each agent k, the Nash equilibrium $w^\star$ satisfies

$$J_k^p(w_k^\star; w_{-k}^\star) \leq J_k^p(w_k; w_{-k}^\star), \qquad \forall w_k \in \mathbb{R}^{M_k} \tag{2.32}$$

The following theorem ensures the existence and uniqueness of the Nash equilibrium.

Theorem 2.1. (**Existence and Uniqueness**): *Under Assumption 2.1 and for any convex choice of $\theta_{EP}(x)$ and any convex and nondecreasing choice of $\theta_{IP}(x)$, there exists a unique Nash equilibrium $w^\star$ for problem (2.27), and it satisfies*

$$F^p(w^\star) \triangleq F(w^\star) + \rho \nabla_{w^\mathsf{T}} p(w^\star) = 0 \tag{2.33}$$

where

$$F^p(w) \triangleq \text{col}\{\nabla_{w_1^\mathsf{T}} J_1^p(w^1), ..., \nabla_{w_N^\mathsf{T}} J_N^p(w^N)\} \tag{2.34}$$

$$\nabla_{w^\mathsf{T}} p(w) \triangleq \text{col}\{\nabla_{w_1^\mathsf{T}} p_1(w^1), ..., \nabla_{w_N^\mathsf{T}} p_N(w^N)\} \tag{2.35}$$

Proof. See Appendix 2.A. □

Now, for any ρ, let us denote the unique Nash equilibrium to the penalized optimization problem (2.27) by

$$w^\star(\rho) \triangleq \text{col}\{w_1^\star(\rho), ..., w_N^\star(\rho)\} \tag{2.36}$$

where

$$
\begin{aligned}
w_k^\star(\rho) &= \arg\min_{w_k \in \mathbb{R}^{M_k}} J_k^p(w_k; w_{-k}^\star(\rho)) \\
&= \arg\min_{w_k \in \mathbb{R}^{M_k}} J_k(w_k; w_{-k}^\star(\rho)) + \rho p_k(w_k; w_{-k}^\star(\rho))
\end{aligned} \tag{2.37}
$$

For convenience, we introduce the notation:

$$w_k^\star(\infty) \triangleq \lim_{\rho\to\infty} w_k^\star(\rho) \tag{2.38}$$

$$w_{-k}^\star(\infty) \triangleq \text{col}\{w_\ell^\star(\infty); \ell \in \mathcal{N}_k \setminus \{k\}\} \tag{2.39}$$

From the results in [54, p. 3930] and [71, Theorem 9.2.2], we know that given any w_{-k} and as ρ goes to infinity, we have

$$\inf_{w_k \in \mathcal{S}_k(w_{-k})} J_k(w_k; w_{-k}) = \lim_{\rho\to\infty} \inf_{w_k \in \mathbb{R}^{M_k}} J_k^p(w_k; w_{-k}) \tag{2.40}$$

and

$$J_k(w_k^o; w_{-k}) = \inf_{w_k \in \mathcal{S}_k(w_{-k})} J_k(w_k; w_{-k}) \tag{2.41}$$

where $w_k^o \in \mathcal{S}_k(w_{-k})$ is feasible for optimization problem (2.7) and satisfies

$$w_k^o \triangleq \lim_{\rho\to\infty} \arg\min_{w_k \in \mathbb{R}^{M_k}} J_k^p(w_k; w_{-k}) \tag{2.42}$$

Therefore, if we are given $w_{-k}^\star(\infty)$, we get

$$
\begin{aligned}
J_k(w_k^\star(\infty); w_{-k}^\star(\infty)) &= \inf_{w_k \in \mathcal{S}_k(w_{-k}^\star(\infty))} J_k(w_k; w_{-k}^\star(\infty)) \\
&= \lim_{\rho\to\infty} \inf_{w_k \in \mathbb{R}^{M_k}} J_k^p(w_k; w_{-k}^\star(\infty)) \\
&= J_k^p(w_k^\star(\infty); w_{-k}^\star(\infty))
\end{aligned} \tag{2.43}
$$

34

It then follows that $w^\star(\infty) \triangleq \text{col}\{w_1^\star(\infty), ..., w_N^\star(\infty)\}$ is an asymptotic Nash equilibrium of GNEP in (2.7) as ρ goes to infinity. Furthermore, for each agent k the value of the original cost J_k coincides with the value of the penalized cost J_k^p at $w^\star(\infty)$. Consequently, the Nash equilibrium for the penalized problem (2.27) can be made arbitrarily close to the set of Nash equilibria (if not unique) by choosing ρ large enough. Comparing with the variational equilibrium concept discussed in [30], the main difference here is that instead of solving an exact GNE directly, we introduce the differentiable penalty function $p(\cdot)$ to get an asymptotic solution, which is more practical computationally under stochastic environments as we will see in later sections.

2.3.2 Stochastic Learning Dynamics

The unknown statistical distribution of the data makes it impossible to solve the penalized optimization problem (2.27) analytically. As a result, a closed form solution to problem (2.27) is not generally possible. If this were possible, then the agents could learn $w^\star$ given knowledge of the other agents' actions; this solution method would lead to the best response dynamics [76]. Since this approach is rarely applicable, agents can instead appeal to learning strategies where they gradually approach the desired $w^\star$ through successive inference from streaming data. For example, one well-known gradient-descent solution to update the agents' actions at discrete-time instants i is to employ the following localized rule [77–79]:

$$
\begin{aligned}
w_{k,i} &= w_{k,i-1} - \mu_k \nabla_{w_k^\mathsf{T}} J_k^p(w_{i-1}^k) \\
&= w_{k,i-1} - \mu_k \left(\nabla_{w_k^\mathsf{T}} J_k(w_{i-1}^k) + \rho \nabla_{w_k^\mathsf{T}} p_k(w_{i-1}^k) \right)
\end{aligned} \tag{2.44}
$$

where μ_k is the step-size for agent k. Alternatively, motivated by the arguments from [54], one can implement (2.44) incrementally by using a two-step learning strategy to improve the individual costs and the penalty costs separately. For example, agent k can use an Adapt-then-Penalize (ATP) diffusion learning strategy to update first the iterate along the negative gradient direction of the individual cost $J_k(\cdot)$ and then apply the correction along

the gradient of the penalty term:

$$(\text{ATP}) \quad \begin{cases} \psi_{k,i} = w_{k,i-1} - \mu_k \nabla_{w_k^\mathsf{T}} J_k(w_{i-1}^k) & (2.45) \\[2ex] w_{k,i} = \psi_{k,i} - \mu_k \rho \nabla_{w_k^\mathsf{T}} p_k(\psi_i^k) & (2.46) \end{cases}$$

where $\psi_{k,i} \in \mathbb{R}^{M_k}$ is an intermediate action of agent k and, similar to w_i^k, the notation ψ_i^k collects the iterates $\psi_{\ell,i}$ from across the neighborhood of agent k. Agents can also switch the order of these two steps and use a Penalize-then-Adapt (PTA) diffusion learning strategy:

$$(\text{PTA}) \quad \begin{cases} \psi_{k,i} = w_{k,i-1} - \mu_k \rho \nabla_{w_k^\mathsf{T}} p_k(w_{i-1}^k) & (2.47) \\[2ex] w_{k,i} = \psi_{k,i} - \mu_k \nabla_{w_k^\mathsf{T}} J_k(\psi_i^k) & (2.48) \end{cases}$$

We note that in the gradient-based learning strategies of (2.44), (2.45)–(2.46), and (2.47)–(2.48), agents are assumed to be able to observe or acquire the intermediate actions taken by neighboring agents and then synchronously update their actions[1]. Furthermore, when implementing these strategies, each agent k requires knowledge of its own gradient quantities $\nabla_{w_k^\mathsf{T}} J_k(w^k)$ and $\nabla_{w_k^\mathsf{T}} p_k(w^k)$. When the exact statistics of the data $\boldsymbol{x}_k$ are unavailable, we need to resort to instantaneous realizations $\{\boldsymbol{x}_{k,i}\}$ of these random variables at each time i and estimate the gradient vectors by employing constructions based on the loss functions, i.e.,

$$\widehat{\nabla_{w_k^\mathsf{T}} J_k}(w^k) \triangleq \nabla_{w_k^\mathsf{T}} Q_k(w^k; \boldsymbol{x}_{k,i}) \tag{2.49}$$

Using these estimates, we arrive at the following stochastic gradient implementation:

$$\boldsymbol{w}_{k,i} = \boldsymbol{w}_{k,i-1} - \mu_k \nabla_{w_k^\mathsf{T}} Q_k(\boldsymbol{w}_{i-1}^k; \boldsymbol{x}_{k,i}) - \mu \rho \nabla_{w_k^\mathsf{T}} p_k(\boldsymbol{w}_{i-1}^k) \tag{2.50}$$

and the corresponding ATP and PTA diffusion versions:

$$\begin{pmatrix} \text{diffusion} \\ \text{ATP} \end{pmatrix} \begin{cases} \boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} - \mu_k \nabla_{w_k^\mathsf{T}} Q_k(\boldsymbol{w}_{i-1}^k; \boldsymbol{x}_{k,i}) & (2.51) \\[2ex] \boldsymbol{w}_{k,i} = \boldsymbol{\psi}_{k,i} - \mu_k \rho \nabla_{w_k^\mathsf{T}} p_k(\boldsymbol{\psi}_i^k) & (2.52) \end{cases}$$

and

$$\begin{pmatrix} \text{diffusion} \\ \text{PTA} \end{pmatrix} \begin{cases} \boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} - \mu_k \rho \nabla_{w_k^\mathsf{T}} p_k(\boldsymbol{w}_{i-1}^k) & (2.53) \\[2ex] \boldsymbol{w}_{k,i} = \boldsymbol{\psi}_{k,i} - \mu_k \nabla_{w_k^\mathsf{T}} Q_k(\boldsymbol{\psi}_i^k; \boldsymbol{x}_{k,i}) & (2.54) \end{cases}$$

---

[1]We remark that asynchronous adaptation and learning is also possible, see [20, 80] and the references therein. We focus here on synchronous operation.

Observe that we are denoting the weight iterates in boldface since they are now random quantities due to the randomness of $\widehat{\nabla_{w_k^{\mathsf{T}}}} J_k(w^k)$ resulting from the use of realizations $\boldsymbol{x}_{k,i}$. Note also that instead of diminishing step-sizes, we are considering constant step-sizes $\{\mu_k\}$ in order to endow the algorithms with a tracking mechanism that enables them to track variations in the statistical distribution of the data over time. If the step-sizes are uniform, i.e., $\mu_k = \mu$ for all k, we will show later in Sec. 2.4 that the diffusion ATP and PTA strategies are more stable than the stochastic gradient (2.50). Furthermore, we will observe in the simulations of Sec. 2.5.2 that the diffusion ATP and PTA strategies exhibit better mean-square error performance than the stochastic gradient (2.50).

Example 2.2. *(**Multitask diffusion adaptation**)* The formulations of multitask diffusion adaptation in [18–20] can also be regarded as a special case of the GNEP formulation for quadratic cost functions. In multitask scenarios, there exist clusters in the network with agents in the same cluster interested in the same objective or task (such as estimating a common vector). Cooperation is still warranted among agents and clusters because the multiple tasks can have some similarities. We can reformulate the multitask problem as an GNEP as follows:

$$\min_{w_k} \quad J_k(w^k)$$
$$\text{subject to} \quad w_k = w_\ell, \ \ell \in \mathcal{N}_k \cap \mathcal{C}_k \tag{2.55}$$

where $\mathcal{C}_k$ denotes the cluster that agent k belongs to. Note that the constraints are only on the neighboring agents belonging to cloud $\mathcal{C}_k$ since they have the same estimation target. Following [18–20], we consider a regularized mean-square-error risk of the form:

$$J_k(w^k) \triangleq \mathbb{E}|\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i} w_k|^2 + \sum_{\ell \in \mathcal{N}_k \backslash \mathcal{C}_k} \eta_{k\ell} \|w_k - w_\ell\|^2 \tag{2.56}$$

where the scalar $\boldsymbol{d}_k(i) \in \mathbb{R}$ and the regression vector $\boldsymbol{u}_{k,i} \in \mathbb{R}^{1 \times M}$ are the observation data, and $\{\eta_{k\ell} \geq 0\}$ are regularization parameters. Note that the regularization terms include only the neighboring agents in different clusters from $\mathcal{C}_k$. Let us rewrite the constraints as $\{w_k(m) - w_\ell(m) = 0\}$ for $\ell \in \mathcal{N}_k \cap \mathcal{C}_k$ and $m = 1, ..., M$, and then use the quadratic penalty

function (2.29) to get

$$p_k(w^k) = \sum_{\ell \in \mathcal{N}_k \cap \mathcal{C}_k} \sum_{m=1}^{M} (w_k(m) - w_\ell(m))^2 \tag{2.57}$$

with the gradient vector

$$\nabla_{w_k^\mathsf{T}} p_k(w^k) = \sum_{\ell \in \mathcal{N}_k \cap \mathcal{C}_k} 2(w_k - w_\ell) \tag{2.58}$$

Using the diffusion ATP strategy in (2.51)–(2.52), we then arrive at the multitask ATC algorithm derived in [18–20]:

$$\begin{pmatrix} \text{multitask} \\ \text{ATC} \end{pmatrix} \begin{cases} \boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} + \mu_k \boldsymbol{u}_{k,i}^\mathsf{T} [\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i} \boldsymbol{w}_{k,i-1}] \\ \qquad + \sum_{\ell \in \mathcal{N}_k \setminus \mathcal{C}(k)} \eta_{k\ell}(\boldsymbol{w}_{\ell,i-1} - \boldsymbol{w}_{k,i-1}) \tag{2.59} \\ \boldsymbol{w}_{k,i} = \sum_{\ell \in \mathcal{N}_k \cap \mathcal{C}_k} a_{\ell k} \boldsymbol{\psi}_{\ell,i} \tag{2.60} \end{cases}$$

where

$$a_{kk} \triangleq 1 - \sum_{\ell \in \mathcal{N}_k \cap \mathcal{C}_k} 2\mu_k \rho, \quad a_{\ell k} \triangleq 2\mu_k \rho, \quad \text{for } \ell \neq k \tag{2.61}$$

We note that for the case $\mathcal{C}_k = \mathcal{N}_k$, the multitask ATC algorithm (2.59)–(2.60) becomes a standard diffusion strategy [11, 12, 14–17, 81–83]. The consensus strategies [1, 2, 4, 6, 7] can also be derived by considering the stochastic gradient descent rule (2.50) and using similar arguments. □

2.4 Performance Analysis

We now examine the convergence and stability properties of the distributed stochastic algorithms (2.51)–(2.52) and (2.53)–(2.54). In particular, we examine how close their limiting point gets to the unique equilibrium point, $w^\star$. To continue, we introduce the following condition on the penalty function. This condition is not restrictive since the choice of the penalty function is under the designer's control.

Condition 2.1. (*Lipschitz gradients*) *Consider two arbitrary block vectors w° and $w^\bullet$ collecting all actions from all agents:*

$$w^\circ \triangleq col\{w_1^\circ, ..., w_N^\circ\}, \qquad w^\bullet \triangleq col\{w_1^\bullet, ..., w_N^\bullet\} \tag{2.62}$$

We denote the corresponding action vectors in $\mathcal{N}_k$ by

$$w_\circ^k \triangleq col\{w_\ell^\circ; \ell \in \mathcal{N}_k\}, \qquad w_\bullet^k \triangleq col\{w_\ell^\bullet; \ell \in \mathcal{N}_k\} \tag{2.63}$$

For each individual agent k, we assume that the gradient vector $\nabla_{w_k^T} p_k(\cdot)$ satisfies:

$$\left\| \nabla_{w_k^T} p_k(w_\circ^k) - \nabla_{w_k^T} p_k(w_\bullet^k) \right\| \leq \gamma_k \left\| w_\circ^k - w_\bullet^k \right\| \tag{2.64}$$

where γ_k is a positive constant. □

Note that $p_k(w^k)$ is not required to be twice-differentiable, which is weaker than the assumption used in [54]. Then, we have the following theorem.

Lemma 2.1. (*Lipschitz continuity*) *Under Condition 2.1 and Assumption 2.2, the penalized block gradient vector $F^p(w)$ is $(\delta + \rho\delta_p)$-Lipschitz continuous, i.e., for any w° and $w^\bullet$ we have*

$$\|F^p(w^\circ) - F^p(w^\bullet)\| \leq (\delta + \rho\delta_p)\|w^\circ - w^\bullet\| \tag{2.65}$$

where $\delta_p \triangleq \left(\sum_{k=1}^N \gamma_k^2\right)^{1/2}$.

Proof. See Appendix 2.B. □

In order to characterize the heterogeneous step-sizes, let us denote the maximal and minimal step-sizes, respectively, over the network by

$$\mu_{\max} \triangleq \max_{1 \leq k \leq N} \{\mu_k\} \tag{2.66}$$

$$\mu_{\min} \triangleq \min_{1 \leq k \leq N} \{\mu_k\} \triangleq (1 - t)\mu_{\max} \tag{2.67}$$

for some parameter $0 \leq t < 1$. A small value of t indicates that the step-sizes $\{\mu_k\}$ are clustered together. To continue, we establish the following lemma.

Lemma 2.2. (*Weighted strong monotonicity*) *The penalized block gradient vector $F^p(w)$ satisfies, for any two action profiles w° and $w^\bullet$,*

$$(w^\circ - w^\bullet)^T U[F^p(w^\circ) - F^p(w^\bullet)] \geq \mu_{max}\nu' \|w^\circ - w^\bullet\|^2 \tag{2.68}$$

where $U \triangleq diag\{\mu_1 I_{M_1}, ..., \mu_N I_{M_N}\}$ is a diagonal matrix with step-sizes in the diagonal positions and $\nu' \triangleq \nu - t(\delta + \rho\delta_p)$. Similarly, the block gradient vector $F(w)$ and the penalty gradient vector $\nabla_w \tau p(w)$ satisfy, respectively,

$$(w^\circ - w^\bullet)^T U[F(w^\circ) - F(w^\bullet)] \geq \mu_{max}\nu'' \|w^\circ - w^\bullet\|^2 \tag{2.69}$$

$$(w^\circ - w^\bullet)^T U[\nabla_w \tau p(w^\circ) - \nabla_w \tau p(w^\bullet)] \geq -t\mu_{max}\delta_p \|w^\circ - w^\bullet\|^2 \tag{2.70}$$

where $\nu'' \triangleq \nu - t\delta$.

Proof. See Appendix 2.C. □

Note that for uniform step-sizes we have $t = 0$ and thus $\nu' = \nu'' = 0$. Furthermore, ν' and ν'' are not necessarily positive unless t is small enough. We further introduce the gradient noise vector

$$\boldsymbol{s}_{k,i}(w^k) = \nabla_{w_k^\mathsf{T}} Q_k(w^k; \boldsymbol{x}_{k,i}) - \nabla_{w_k^\mathsf{T}} J_k(w^k) \tag{2.71}$$

and define the network vectors

$$\boldsymbol{s}_i(w) \triangleq \text{col}\{\boldsymbol{s}_{k,i}(w^1), \dots, \boldsymbol{s}_{N,i}(w^N)\} \tag{2.72}$$

$$\boldsymbol{Q}_i(w) \triangleq \text{col}\{\nabla_{w_1^\mathsf{T}} Q_1(w^1; \boldsymbol{x}_{1,i}), \dots, \nabla_{w_N^\mathsf{T}} Q_N(w^N; \boldsymbol{x}_{N,i})\} \tag{2.73}$$

where we simplified the notation $\boldsymbol{s}_{k,i}(w^k)$, $\boldsymbol{s}_i(w)$ and $\boldsymbol{Q}_i(w)$ by dropping $\{\boldsymbol{x}_{k,i}\}$ from their arguments. Then, it holds that

$$\boldsymbol{s}_i(w) = \boldsymbol{Q}_i(w) - F(w) \tag{2.74}$$

Note that given the action profile w, the randomness of $\boldsymbol{s}_{k,i}$, $\boldsymbol{s}_i$ and $\boldsymbol{Q}_i$ comes from the random data $\{\boldsymbol{x}_{k,i}\}$, and therefore we denote them in boldface. We denote by $\boldsymbol{\mathcal{F}}_{i-1}$ the collection of iterates $\{\boldsymbol{w}_{k,i-1}\}$ at all agents $k = 1, ..., N$ and up to time $i - 1$.

40

Assumption 2.3. *(**Gradient noise**) It is assumed that the first and second-order condi-tional moments of the gradient noise process satisfy:*

$$\mathbb{E}[\boldsymbol{s}_i(\boldsymbol{w}_{i-1})|\boldsymbol{\mathcal{F}}_{i-1}] = 0 \tag{2.75}$$

$$\mathbb{E}\left[\|\boldsymbol{s}_i(\boldsymbol{w}_{i-1})\|^2|\boldsymbol{\mathcal{F}}_{i-1}\right] \leq \alpha\|\boldsymbol{w}_{i-1}\|^2 + \beta \tag{2.76}$$

for some nonnegative constants α and β. □

It can be verified that conditions (2.75)–(2.76) are automatically satisfied for impor-tant cases of interest. For example, consider the case of quadratic losses in (2.19). Some straightforward algebra shows in this case that, using stationary realizations $\{\boldsymbol{B}_i, \boldsymbol{b}_i\}$ for the quantities $\{B, b\}$ in (2.21), we get the approximate block gradient vector as

$$\boldsymbol{Q}_i(\boldsymbol{w}_{i-1}) = \boldsymbol{B}_i\boldsymbol{w}_{i-1} + \boldsymbol{b}_i \tag{2.77}$$

so that

$$\boldsymbol{s}_i(\boldsymbol{w}_{i-1}) \triangleq -\widetilde{\boldsymbol{B}}_i\boldsymbol{w}_{i-1} - \widetilde{\boldsymbol{b}}_i \tag{2.78}$$

where $\widetilde{\boldsymbol{B}}_i \triangleq B - \boldsymbol{B}_i$ and $\widetilde{\boldsymbol{b}}_i \triangleq b - \boldsymbol{b}_i$. Note that $\mathbb{E}\widetilde{\boldsymbol{B}}_i = 0$ and $\mathbb{E}\widetilde{\boldsymbol{b}}_i = 0$ from the fact that $B = \mathbb{E}\boldsymbol{B}_i$ and $b = \mathbb{E}\boldsymbol{b}_i$. From the independence of $\boldsymbol{B}_i$, $\boldsymbol{b}_i$, and $\boldsymbol{w}_{i-1}$, Assumption 2.3 can be seen to be satisfied since

$$\mathbb{E}[\boldsymbol{s}_i(\boldsymbol{w}_{i-1})|\boldsymbol{\mathcal{F}}_{i-1}] = -\mathbb{E}[\widetilde{\boldsymbol{B}}_i] \cdot \boldsymbol{w}_{i-1} - \mathbb{E}\widetilde{\boldsymbol{b}}_i = 0 \tag{2.79}$$

$$\mathbb{E}\left[\|\boldsymbol{s}_i(\boldsymbol{w}_{i-1})\|^2|\boldsymbol{\mathcal{F}}_{i-1}\right] \leq \lambda_{\max}\left(\mathbb{E}[\widetilde{\boldsymbol{B}}_i^{\mathsf{T}}\widetilde{\boldsymbol{B}}_i]\right)\|\boldsymbol{w}_{i-1}\|^2 + \mathbb{E}\|\widetilde{\boldsymbol{b}}_i\|^2 \tag{2.80}$$

with $\alpha = \lambda_{\max}(\mathbb{E}[\widetilde{\boldsymbol{B}}_i^{\mathsf{T}}\widetilde{\boldsymbol{B}}_i])$ and $\beta = \mathbb{E}\|\widetilde{\boldsymbol{b}}_i\|^2$.

2.4.1 Stochastic Gradient Dynamics

We consider first the stochastic-gradient implementation (2.50). We can describe the evolu-tion of the dynamics of the algorithm in terms of the aggregate quantities $\boldsymbol{w}_i \triangleq \operatorname{col}\{\boldsymbol{w}_{1,i}, ..., \boldsymbol{w}_{N,i}\}$ by writing:

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} - U\boldsymbol{Q}_i(\boldsymbol{w}_{i-1}) - \rho U\nabla_{w^{\mathsf{T}}}p(\boldsymbol{w}_{i-1}) \tag{2.81}$$

Subtracting $w^\star$ from both sides of (2.81), introducing the error vector $\widetilde{\boldsymbol{w}}_i \triangleq w^\star - \boldsymbol{w}_i$ and using (2.33) we find that

$$\widetilde{\boldsymbol{w}}_i = \widetilde{\boldsymbol{w}}_{i-1} + UF^p(\boldsymbol{w}_{i-1}) + U\boldsymbol{s}_i(\boldsymbol{w}_{i-1}) \tag{2.82}$$

The following theorem now establishes that the network error is mean-square stable for sufficiently small step-sizes $\{\mu_k\}$ and variation parameter t.

Theorem 2.2. *(**Mean-square-error stability**) For the stochastic gradient implementation (2.50), if the step-sizes $\{\mu_k\}$ satisfy*

$$0 < \mu_{max} < \frac{2\nu'}{(\delta + \rho\delta_p)^2 + 2\alpha}, \quad t < \frac{\nu}{\delta + \rho\delta_p} \tag{2.83}$$

then it holds that

$$\limsup_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|^2 = O(\mu_{max}) \tag{2.84}$$

Proof. See Appendix 2.D. $\square$

2.4.2 Diffusion ATP and PTA Strategies

Let us consider next the deterministic ATP and PTA strategies (2.45)–(2.46) and (2.47)–(2.48), respectively, without gradient noise. Later, we re-incorporate the gradient noise and adjust the conclusions. Thus, note that in the noiseless case we can aggregate the recursions across all agents into the following unified description:

$$\phi_i = w_{i-1} - c_1 \rho U \nabla_{w^\mathsf{T}} p(w_{i-1}) \tag{2.85}$$

$$\psi_i = \phi_i - UF(\phi_i) \tag{2.86}$$

$$w_i = \psi_i - c_2 \rho U \nabla_{w^\mathsf{T}} p(\psi_i) \tag{2.87}$$

for some constants (c_1, c_2). By setting $(c_1, c_2) = (0, 1)$ we recover the ATP recursions (2.45)–(2.46) while for $(c_1, c_2) = (1, 0)$ we obtain the PTA recursions from (2.47)–(2.48). We thus note that the constants (c_1, c_2) satisfy the properties:

$$c_1^2 = c_1, \quad c_2^2 = c_2, \quad c_1 \cdot c_2 = 0, \quad c_1 + c_2 = 1 \tag{2.88}$$

The following result establishes that recursions (2.85)–(2.87) converge to a unique fixed point.

42

Theorem 2.3. (*Unique fixed point*) *The mapping from w_{i-1} to w_i in (2.85)–(2.87) converges to a unique fixed point, denoted by ψ^∞, for small step-sizes and for sufficiently large penalty parameters that satisfy:*

$$0 < \mu_{max} < \mu_o, \quad t < \frac{\nu}{\delta + \rho\delta_p}, \quad \rho > \frac{\delta}{\delta_p} \tag{2.89}$$

where

$$\mu_o \triangleq \min\left\{ \frac{2\nu'}{\delta^2 + \rho^2\delta_p^2 - 4t\nu''\rho\delta_p}, \frac{\nu' + \frac{t(\rho^2\delta_p^2 - \delta^2)}{\rho\delta_p}}{\delta^2} \right\} \tag{2.90}$$

Proof. See Appendix 2.E. □

We note that if the step-sizes are uniform, i.e., $\mu_k = \mu$ and $t = 0$, the step-size condition in (2.89) simplifies to

$$0 < \mu < \frac{2\nu}{\delta^2 + \rho^2\delta_p^2} \tag{2.91}$$

since

$$\rho > \frac{\delta}{\delta_p} \iff \rho^2\delta_p^2 > \delta^2 \iff \frac{\nu}{\delta^2} > \frac{2\nu}{\delta^2 + \rho^2\delta_p^2} \tag{2.92}$$

From Theorem 2.3 we know that there exists a unique fixed point for recursion (2.85)–(2.87), which means that we can write

$$\phi^\infty = w^\infty - c_1\rho U \nabla_{w^\top} p(w^\infty) \tag{2.93}$$

$$\psi^\infty = \phi^\infty - U F(\phi^\infty) \tag{2.94}$$

$$w^\infty = \psi^\infty - c_2\rho U \nabla_{w^\top} p(\psi^\infty) \tag{2.95}$$

where we are denoting the network fixed-point vectors by w^∞, ψ^∞ and ϕ^∞. Similarly, we can express the diffusion (stochastic) versions of the ATP and PTA strategies in (2.51)–(2.52) and (2.53)–(2.54) in the form:

$$\boldsymbol{\phi}_i = \boldsymbol{w}_{i-1} - c_1\rho U \nabla_{w^\top} p(\boldsymbol{w}_{i-1}) \tag{2.96}$$

$$\boldsymbol{\psi}_i = \boldsymbol{\phi}_i - U \boldsymbol{Q}_i(\boldsymbol{\phi}_i) \tag{2.97}$$

$$\boldsymbol{w}_i = \boldsymbol{\psi}_i - c_2\rho U \nabla_{w^\top} p(\boldsymbol{\psi}_i) \tag{2.98}$$

43

Let $\widetilde{\boldsymbol{w}}_i^\infty \triangleq w^\infty - \boldsymbol{w}_i$ denote the fixed-point error resulting from (2.96)–(2.98). The following theorem shows that the variance of this error is bounded.

Theorem 2.4. (*Bounded MSE*) *For the stochastic recursion (2.96)–(2.98), if the step-sizes $\{\mu_k\}$ and the penalty parameter ρ satisfy*

$$0 < \mu_{max} < \mu_o', \quad t < \frac{\nu}{\delta + \rho\delta_p}, \quad \rho > \frac{\sqrt{\delta^2 + 2\alpha}}{\delta_p} \tag{2.99}$$

where

$$\mu_o' \triangleq \min \left\{ \frac{2\nu'}{\delta^2 + 2\alpha + \rho^2\delta_p^2 - 4t\nu''\rho\delta_p}, \frac{\nu' + \frac{t(\rho^2\delta_p^2 - (\delta^2 + 2\alpha))}{\rho\delta_p}}{\delta^2 + 2\alpha} \right\} \tag{2.100}$$

then it holds that for sufficiently small step-sizes

$$\limsup_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_i^\infty\|^2 = O(\mu_{max}) \tag{2.101}$$

Proof. See Appendix 2.F. □

It is easy to verify that if the step-sizes are uniform, the step-size condition in (2.99) becomes

$$0 < \mu < \frac{2\nu}{\delta^2 + 2\alpha + \rho^2\delta_p^2} \tag{2.102}$$

We note from $\alpha \geq 0$ that $\mu_o' \leq \mu_o$, which means that condition (2.99) for the stochastic recursion implies condition (2.89) for the deterministic recursion. Therefore, any $\mu_{\max}$ satisfying (2.99) ensures the existence of the fixed point w^∞. However, the fixed point w^∞ is generally different from the desired Nash equilibrium $w^\star$. In the following theorem, we examine the bias $\widetilde{w} \triangleq w^\star - w^\infty$. We show that for small $\mu_{\max}$, the norm of the bias is asymptotically upper bounded by $O(\mu_{\max})$.

Theorem 2.5. (*Small bias*) *For sufficiently small step-sizes $\{\mu_k\}$ satisfying the following conditions:*

$$0 < \mu_{max} < \mu_o, \quad t < \frac{\nu}{\delta + \rho\delta_p}, \quad \rho > \frac{\delta}{\delta_p} \tag{2.103}$$

it holds that

$$\lim_{\mu_{max}\to 0}\sup \frac{\|w^\star - w^\infty\|}{\mu_{max}} \le c\rho \tag{2.104}$$

where c is a constant independent of μ_{max}. Therefore, for sufficiently small μ_{max} we can write

$$\lim_{i\to\infty}\sup \mathbb{E}\|w^\star - \boldsymbol{w}_i\|^2 \le 2\lim_{i\to\infty}\sup \mathbb{E}\|w^\infty - \boldsymbol{w}_i\|^2 + 2\|w^\star - w^\infty\|^2$$

$$= O(\mu_{max}) + O(\mu_{max}^2\rho^2) \tag{2.105}$$

Proof. See Appendix 2.G. $\qquad\square$

In Figure 2.2, we illustrate the relation between $\boldsymbol{w}_i$, $w^\star$, and w^∞ in steady-state for sufficiently small step-sizes. We note that $\boldsymbol{w}_i$, $w^\star$, and w^∞ asymptotically approach to the Nash equilibrium set of the original GNEP (2.7) as $\rho \to \infty$ and $\mu_{\max} \to 0$. We note that condition (2.99) implies conditions (2.89) and (2.103). That is, as long as the step-sizes $\{\mu_k\}$ and the penalty parameter ρ satisfy (2.99), the diffusion ATP and PTA learning strategies have fixed points, bounded MSE, and small bias. Furthermore, comparing (2.102) with (2.83) we observe that by using uniform step-sizes, the diffusion ATP and PTA learning strategies are more stable than the stochastic gradient dynamic strategy (2.50) since they are allowed to use a larger step-size, which would assist with faster convergence performance. We will observe this in the simulations later. For the special case in Example 2, this conclusion conforms with the results in [21] that the diffusion strategies are more stable than the consensus strategy.

2.5 Case Study and Simulations

2.5.1 Stochastic Network Cournot Competition

In this section, we consider the stochastic network Cournot competition problem [24, 25, 42, 43, 84] with shared constraints. We assume that the environment is stochastically dynamic in the following manner. Suppose that we have a network with N factories, regarded as the agents discussed in this chapter, and L markets connected to the factories. Each factory k

45
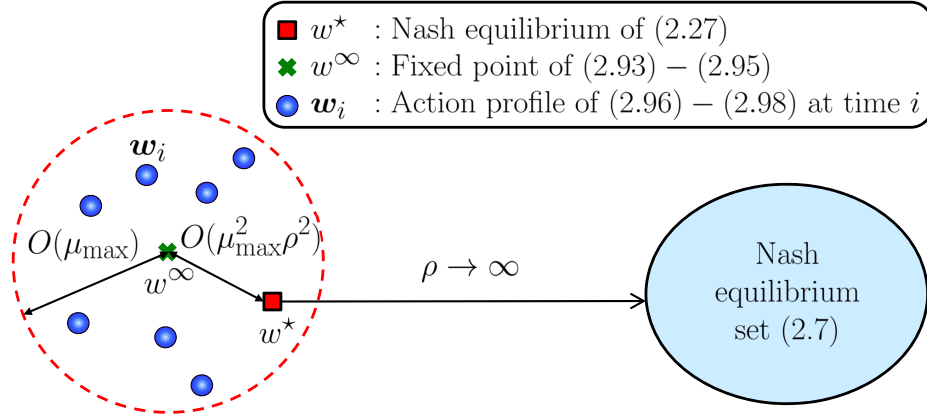
Figure 2.2: Illustration of the relations between $\boldsymbol{w}_i$, $w^\star$, and w^∞ in steady-state for sufficiently small step-sizes. The notation $O(\mu_{\max})$ and $O(\mu_{\max}^2 \rho^2)$ in the drawing represent the squared distances $\mathbb{E}\|w^\infty - \boldsymbol{w}_i\|^2$ and $\|w^\star - w^\infty\|$, respectively.

needs to determine a continuous-valued and nonnegative quantity of products to be produced and delivered to each connected market, which is defined as the action of factory k denoted by $w_k = [w_k(1), ..., w_k(M_k)]^\mathsf{T}$ where we assumed M_k markets are connected to factory k. For each factory k, there exists a random quadratic production cost function to generate $\sum_{n=1}^{M_k} w_k(n)$ amount of products, i.e., the production cost function for each factory is given by

$$\boldsymbol{C}_k(w_k) = (x_k + \boldsymbol{v}_{x,k}) \left(\sum_{m=1}^{M_k} w_k(m) \right)^2 \tag{2.106}$$

for some parameter $x_k > 0$ and random disturbance $\boldsymbol{v}_{x,k}$ with zero mean. Furthermore, the price of products sold in each market ℓ is assumed to follow a linear function:

$$\boldsymbol{P}_\ell(r(\ell)) = q_\ell - (y_\ell + \boldsymbol{v}_{y,\ell})r(\ell) \tag{2.107}$$

where $q_\ell > 0$ and $y_\ell > 0$ are the pricing parameters, the random disturbance $\boldsymbol{v}_{y,\ell}$ is zero-mean, and $r(\ell)$ is the total amount of products delivered to market ℓ by all connected factories, i.e.,

$$r(\ell) = \sum_{k=1, w_k(u) \sqsubset \ell}^{N} w_k(u) \tag{2.108}$$

46

where we write $w_k(u) \sqsubset \ell$ to represent that $w_k(u)$ is the quantity that factory k delivers to market ℓ. Note that in order to be consistent with the notation in (2.7), the index u in $w_k(u)$ can be different from the index ℓ denoted for markets. Consequently, each factory k has an individual cost function as follows:

$$J_k(w^k) = \mathbb{E}\left(\boldsymbol{C}_k(w_k) - \sum_{\ell=1,u\sqsubset\ell}^{L} w_k(u) \cdot \boldsymbol{P}_\ell(r(\ell))\right)$$

$$= x_k \left(\sum_{m=1}^{M_k} w_k(m)\right)^2 - \sum_{\ell=1,w_k(u)\sqsubset\ell}^{L} w_k(u)\left(q_\ell - y_\ell \cdot r(\ell)\right) \tag{2.109}$$

Note that the loss functions in the individual cost functions can be rewritten in the quadratic form (2.19). Now, let us show that $\{J_k(w^k)\}$ in the network Cournot competition are strongly monotone. For each $u \sqsubset \ell$ we have the components in $\nabla_{w_k^{\mathsf{T}}} J_k(w^k)$ as

$$\frac{\partial J_k(w^k)}{\partial w_k(u)} = 2x_k \sum_{m=1}^{M_k} w_k(m) - q_\ell + y_\ell\left[w_k(u) + r(\ell)\right] \tag{2.110}$$

If we collect these components into the long block vector $F(w)$, we get the form in (2.21) where the $(m,n)-$th entry in each block of matrix B is given by

$$B_{kk}^k(m,n) = \begin{cases} x_k + y_\ell, & \text{if } m = n \text{ s.t. } w_k(m) \sqsubset \ell \\ x_k, & \text{if } m \neq n \end{cases} \tag{2.111}$$

$$B_{kq}^k(m,n) = \begin{cases} y_\ell, & \text{if } w_k(m) \sqsubset \ell \text{ and } w_q(n) \sqsubset \ell \\ 0, & \text{otherwise} \end{cases}, \quad k \neq q \tag{2.112}$$

It is easy to check that matrix B can be expressed as

$$B = XX^{\mathsf{T}} + Y_1 Y_1^{\mathsf{T}} + Y_2 Y_2^{\mathsf{T}} \tag{2.113}$$

where X is an $M \times M$ diagonal matrix with diagonal entries $\{\sqrt{y_{\ell_{ku}}}\}$ for $w_k(u) \sqsubset \ell_{ku}$, Y_1 is a $M \times N$ block diagonal matrix with $N \times N$ blocks in which the $(k,k)-$th diagonal block is $Y_{1,kk} = \left[\sqrt{2x_k}, ..., \sqrt{2x_k}\right]^{\mathsf{T}}$ of size $M_k \times 1$, and Y_2 is a $M \times L$ block matrix with $N \times L$ blocks in which the $(k,\ell)-$th block is a vector of size $M_k \times 1$ and defined as

$$Y_{2,k\ell}(m) = \begin{cases} \sqrt{y_\ell}, & \text{if } w_k(m) \sqsubset \ell \\ 0, & \text{otherwise} \end{cases} \tag{2.114}$$
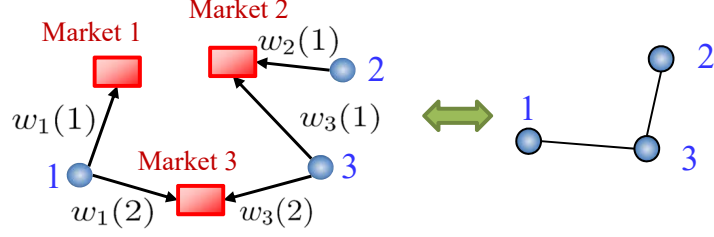
Figure 2.3: An example to illustrate the network Cournot competition and the equivalent network topology.

Therefore, we find that B has the following property for any $M \times 1$ vector a:

$$a^{\mathsf{T}} B a = a^{\mathsf{T}} X X^{\mathsf{T}} a + \|Y_1^{\mathsf{T}} a\|^2 + \|Y_2^{\mathsf{T}} a\|^2 \geq x_{\min} \cdot \|a\|^2 \tag{2.115}$$

where $x_{\min} \triangleq \min_{1 \leq \ell \leq L} x_\ell$. Consequently, the network Cournot competition with individual cost functions in (2.109) satisfies the strongly monotone property (2.13).

Example 2.3. *(**Cournot network with** 3 **agents**)* An illustrative example with $N = 3$ factories and $L = 3$ markets is provided in Fig. 2.3. Following the notations for the quantities at each link, we have the individual cost functions for the factories as

$$J_1(w^1) = x_1 \left[w_1(1) + w_1(2) \right]^2 - w_1(1) \cdot (q_1 - y_1 \cdot w_1(1))$$
$$- w_1(2) \cdot (q_2 - y_2 \cdot [w_1(2) + w_3(2)])$$
$$J_2(w^2) = x_2 [w_2(1)]^2 - w_2(1) \cdot (q_2 - y_2 \cdot [w_1(2) + w_3(2)])$$
$$J_3(w^3) = x_3 \left[w_3(1) + w_3(2) \right]^2$$
$$- w_3(1) \cdot (q_2 - y_2 \cdot [w_1(2) + w_3(2)])$$
$$- w_3(2) \cdot (q_3 - y_3 \cdot [w_2(1) + w_3(1)])$$

Therefore, we get

$$\frac{\partial J_1(w^1)}{\partial w_1(1)} = 2x_1 \left[w_1(1) + w_1(2) \right] - q_1 + 2y_1 w_1(1)$$

$$\frac{\partial J_1(w^1)}{\partial w_1(2)} = 2x_1 \left[w_1(1) + w_1(2) \right] - q_3 + 2y_3 w_1(2) + y_3 w_3(2)$$

$$\frac{\partial J_2(w^2)}{\partial w_2(1)} = 2x_2 w_2(1) - q_2 + 2y_2 w_2(1) + y_2 w_3(1)$$

$$\frac{\partial J_3(w^3)}{\partial w_3(1)} = 2x_3 \left[w_3(1) + w_3(2) \right] - q_2 + 2y_2 w_3(1) + y_2 w_2(1)$$

$$\frac{\partial J_3(w^3)}{\partial w_3(2)} = 2x_3 \left[w_3(1) + w_3(2) \right] - q_3 + 2y_3 w_3(2) + y_3 w_1(2)$$

It can be then verified that matrix B is given by

$$B = \begin{bmatrix} 2x_1 + 2y_1 & 2x_1 & 0 & 0 & 0 \\ 2x_1 & 2x_1 + 2y_3 & 0 & 0 & y_3 \\ 0 & 0 & 2x_2 + 2y_2 & y_2 & 0 \\ 0 & 0 & y_2 & 2x_3 + 2y_2 & 2x_3 \\ 0 & y_3 & 0 & 2x_3 & 2x_3 + 2y_3 \end{bmatrix}$$

$$= XX^\mathsf{T} + Y_1 Y_1^\mathsf{T} + Y_2 Y_2^\mathsf{T} \tag{2.116}$$

where

$$X = \text{diag}\{\sqrt{y_1}, \sqrt{y_3}, \sqrt{y_2}, \sqrt{y_2}, \sqrt{y_3}\} \tag{2.117}$$

$$Y_1 = \begin{bmatrix} \sqrt{2x_1} & \sqrt{2x_1} & 0 & 0 & 0 \\ 0 & 0 & \sqrt{2x_2} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{2x_3} & \sqrt{2x_3} \end{bmatrix}^\mathsf{T} \tag{2.118}$$

$$Y_2 = \begin{bmatrix} \sqrt{y_1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{y_2} & \sqrt{y_2} & 0 \\ 0 & \sqrt{y_3} & 0 & 0 & \sqrt{y_3} \end{bmatrix}^\mathsf{T} \tag{2.119}$$

$\square$

2.5.2 Numerical Results

In the simulations, we consider a network with $N = 20$ factories and $L = 7$ markets which are connected as shown in Fig. 2.4. For each individual cost function $J_k(w^k)$, we set $x_k = 4$, $q_\ell = 12$, and $y_\ell = 4$ for all k and ℓ in (2.109). For the stochastic setting, the realizations of random noises $\boldsymbol{v}_{x,k}$ and $\boldsymbol{v}_{y,\ell}$ for all k and ℓ are generated at each time instant i, and are assumed to be temporally and spatially independent. We further assume that both $\boldsymbol{v}_{x,k}$ and $\boldsymbol{v}_{y,\ell}$ are uniformly distributed between $[-4, 4]$. The step-sizes are assumed to be uniform, i.e., $\mu_k = \mu$ for all k.

The action w_k of each factory k needs to be determined under the following constraints. The quantity of products delivered to each market has to be nonnegative and each market ℓ has an upper limit capacity h_ℓ of products, i.e., for $m = 1, ..., M_k$ and $\ell = 1, ..., L$,

$$w_k(m) \geq 0, \quad r(\ell) = \sum_{k=1, w_k(u) \sqsubset \ell}^{N} w_k(u) \leq h_\ell \tag{2.120}$$

where h_ℓ is set to be 1 in the experiments. Furthermore, we apply the quadratic penalty function in (2.31) to each constraint in the algorithms. We remark that the proposed penalty methods give only asymptotically feasible solutions, which could be improved by imposing harsher penalty or considering stricter constraints than (2.120). However, we rely on (2.120) in the simulations to examine the numerical performance regardless of solution feasibility.

We first set the penalty parameter ρ to 200 and vary the step-size μ for the stochastic gradient dynamic (2.50), ATP strategy (2.51)-(2.52), and PTA strategy (2.53)-(2.54). In Fig. 2.5, we study the mean-square-deviation (MSD) performance, defined as $\mathbb{E}\|w^\infty - \boldsymbol{w}_i\|^2$, for each algorithm toward its fixed point. Note that for the stochastic gradient case we have $w^\infty = w^\star$. We can see that with a smaller step-size μ, the three algorithms exhibit smaller steady-state MSD values while converging slower, and their differences vanish with smaller μ as well. It is worthwhile to note though that the diffusion ATP and PTA strategies generally outperform the stochastic gradient dynamic. Furthermore, the ATP and PTA strategies allow larger ranges of step-sizes, as we can see that for $\mu = 0.0065$ these two strategies converge while the stochastic gradient dynamic does not. In Fig. 2.6, we observe that for
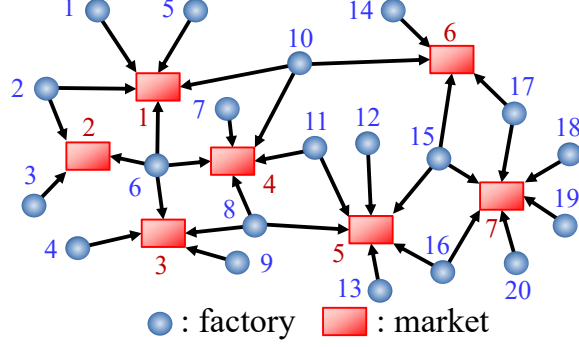
Figure 2.4: Network structure used for the simulations of the network Cournot competition.

sufficiently small step-sizes, the steady-state MSD values of diffusion ATP and PTA decrease linearly with respect to μ, as we expect from (2.101). The bias between the fixed points w^∞ and the Nash equilibrium $w^\star$ is shown in Fig. 2.7. We can see that the bias $\|w^\star - w^\infty\|$ is linear with respect to the step-size μ and the slope becomes steep when ρ increases, which verifies the result in (2.104). Comparing diffusion ATP and PTA strategies using sufficiently small step-sizes, we find that diffusion ATP exhibits smaller steady-state MSD values than diffusion PTA; on the other hand, diffusion PTA shows smaller bias values than diffusion ATP. This result would depend on the structure of the individual costs and the shared constraints, and the selection of the penalty functions θ_{IP} and θ_{EP}. However, as the step-size decreases, the difference between diffusion ATP and PTA strategies becomes small in terms of the steady-state MSD and bias.

For comparisons, we simulate two related projection-based stochastic algorithms discussed in [53], i.e., the distributed Arrow-Hurwicz method and the iterative Tikhonov regularization. Both algorithms use a decaying and uniform step-size $\mu_i = (1000 + i)^{-0.841}$. The distributed Arrow-Hurwicz method consists of the following two steps:

$$\begin{cases} \boldsymbol{w}_{k,i} = \Pi_{\mathbb{R}^+} \left[\boldsymbol{w}_{k,i-1} - \mu_i \left(\widehat{\nabla_{w_k^{\mathsf{T}}}} J_k(\boldsymbol{w}_{k,i-1}) \sum_{\ell=1}^L \boldsymbol{\lambda}_{\ell,i-1}(\boldsymbol{r}_i(\ell) - h_\ell) \right) \right] \\ \boldsymbol{\lambda}_{\ell,i} = \Pi_{\mathbb{R}^+} \left[\boldsymbol{\lambda}_{\ell,i-1} + \mu_i (\boldsymbol{r}_i(\ell) - h_\ell) \right] \end{cases} \tag{2.121}$$

where $\boldsymbol{r}_i(\ell)$ denotes the random realization for $r(\ell)$ at time i. On the other hand, the iterative

51
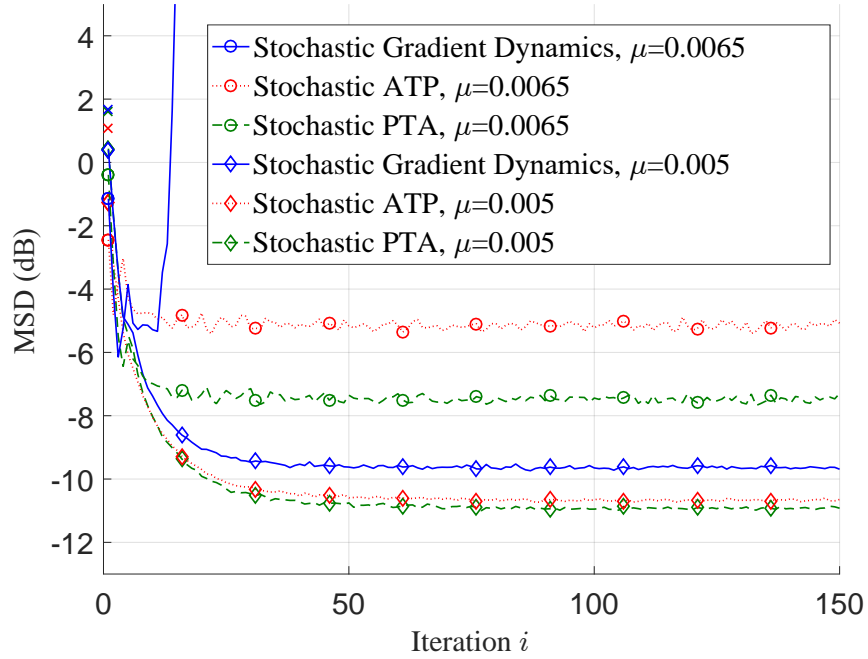
Figure 2.5: MSD learning curves for the stochastic gradient dynamic, diffusion ATP, and diffusion PTA with different step-sizes μ.
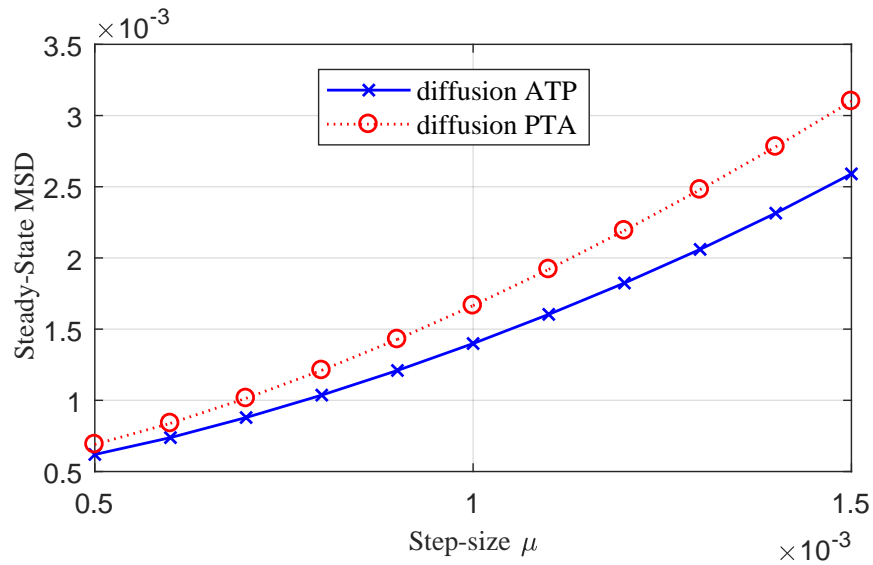
Figure 2.6: The steady-state MSD for diffusion ATP and diffusion PTA.
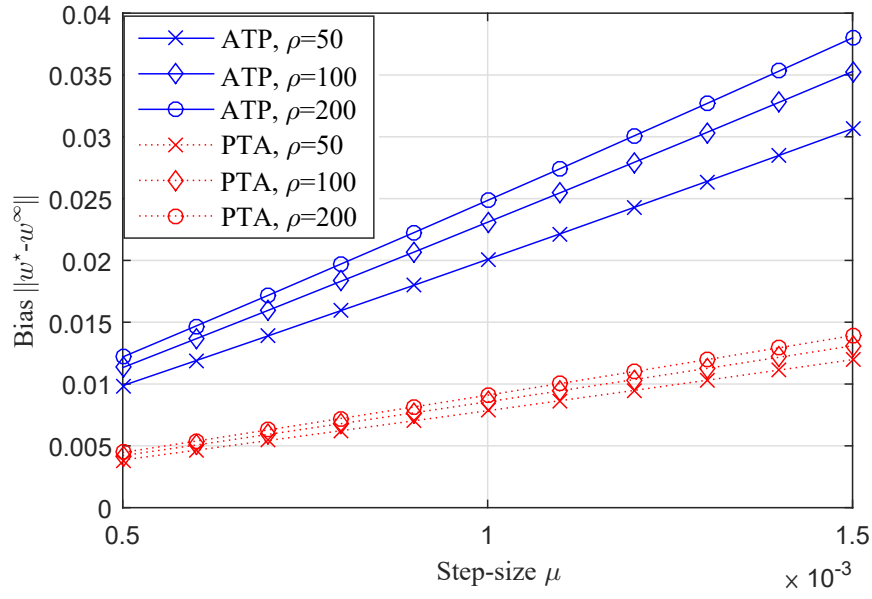
Figure 2.7: The bias distance $\|w^\star - w^\infty\|$ from the Nash equilibrium $w^\star$ to fixed points w^∞ for diffusion ATP and diffusion PTA.
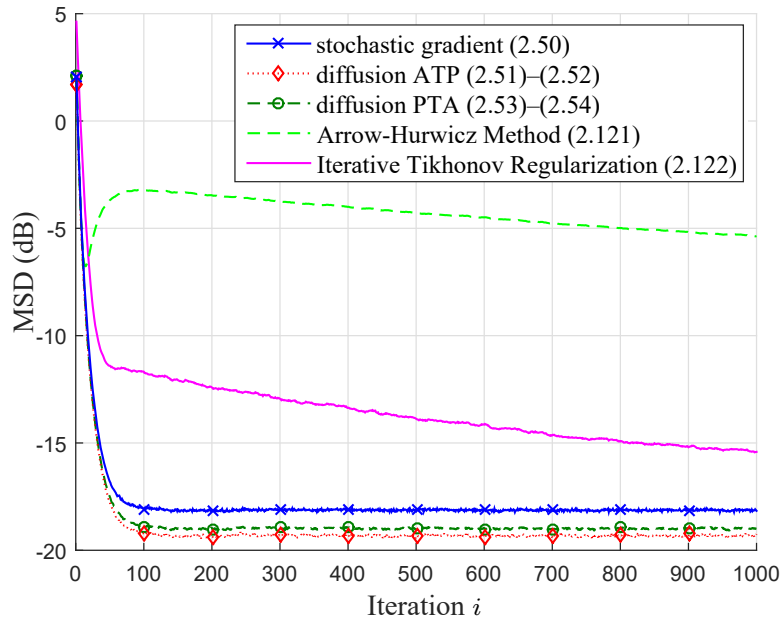
Figure 2.8: Comparisons of MSD learning curves for algorithms.

Tikhonov regularization follows these two steps:

$$\begin{cases} \boldsymbol{w}_{k,i} = \Pi_{\mathbb{R}^+}\Big[\boldsymbol{w}_{k,i-1} - \mu_i\Big(\epsilon_i\boldsymbol{w}_{k,i-1} + \widehat{\nabla_{w_k^\mathsf{T}}}J_k(\boldsymbol{w}_{k,i-1})\sum_{\ell=1}^{L}\boldsymbol{\lambda}_{\ell,i-1}(\boldsymbol{r}_i(\ell) - h_\ell)\Big)\Big] \\ \boldsymbol{\lambda}_{\ell,i} = \Pi_{\mathbb{R}^+}\left[\boldsymbol{\lambda}_{\ell,i-1} + \mu_i(\boldsymbol{r}_i(\ell) - h_\ell) - \mu_i\epsilon_i\boldsymbol{\lambda}_{\ell,i-1}\right] \end{cases} \tag{2.122}$$

where $\epsilon_i = (1000 + i)^{-0.1}$ is the regularization parameter which also vanishes with time i. We note that these two algorithms rely on the additional use of L Lagrange multiplier(s) to deal with the shared constraints, which require some additional "bridge nodes" for implementation. Furthermore, the projection step incurs additional computation complexity. These two problems do not appear in our penalty-based algorithms proposed in this chapter. In Fig. 2.8, we simulate the MSD learning curves for these algorithms. In order to make a fair comparison, we set the step-size $\mu = 0.003$ for the penalty-based strategies such that all algorithms have the same initial value of the step-sizes. The penalty parameter ρ is set to 200. We observe that the stochastic gradient dynamic, ATP, and PTA strategies converge much faster than the distributed Arrow-Hurwicz method and the iterative Tikhonov regularization. Furthermore, the distributed Arrow-Hurwicz and the iterative Tikhonov regularization methods have larger steady-state MSD values than the three penalty-based algorithms, while their MSD values can slowly and continuously improve at the expense of loss of tracking capability since the step-size goes to zero asymptotically.

2.6 Concluding Remarks

This chapter focuses on GNEPs with shared constraints over network topologies in stochastic environments. We develop three fully-distributed online learning strategies which asymptotically approach the set of generalized Nash equilibrium for small constant step-sizes and sufficiently large penalty parameters. An interesting future work would be to explore how the converging point of our algorithms in the set of GNE(s) relate to the variational equilibrium obtained by KKT conditions with identical Lagrange multipliers [85, 86]. Another possibility for future work is to explore the use of sub-gradient methods would be useful for sub-differentiable penalty functions and/or individual cost functions [7, 63, 87]. Asynchronous adaptation learning [20, 80] is also a useful extension so that agents do not need to

execute the update of actions simultaneously.

APPENDICES

2.A   Proof of Theorem 2.1

We introduce the aggregate penalty function

$$p(w) \triangleq \sum_{u=1}^{U} \theta_{\mathrm{EP}}(h_u(w)) + \sum_{q=1}^{L} \theta_{\mathrm{IP}}(g_q(w)) \tag{2.123}$$

and note that

$$\nabla_{w_k^\mathsf{T}} p(w) = \sum_{u=1}^{U} \nabla_{h_u(w)} \theta_{\mathrm{EP}}(h_u(w)) \cdot \nabla_{w_k^\mathsf{T}} h_u(w) + \sum_{q=1}^{L} \nabla_{g_q(w)} \theta_{\mathrm{IP}}(g_q(w)) \cdot \nabla_{w_k^\mathsf{T}} g_q(w) \tag{2.124}$$

and

$$\begin{aligned}
\nabla_{w_k^\mathsf{T}} p_k(w^k) = & \sum_{u=1}^{U_k} \nabla_{h_{k,u}(w^k)} \theta_{\mathrm{EP}}(h_{k,u}(w^k)) \cdot \nabla_{w_k^\mathsf{T}} h_{k,u}(w^k) \\
& + \sum_{q=1}^{L_k} \nabla_{g_{k,q}(w^k)} \theta_{\mathrm{IP}}(g_{k,q}(w^k)) \cdot \nabla_{w_k^\mathsf{T}} g_{k,q}(w^k)
\end{aligned} \tag{2.125}$$

Recall that, as defined in (2.8), the global constraint functions $\{h_u(w)\}$ and $\{g_q(w)\}$ are distinctly collected and include all $\{h_{k,u}(w^k)\}$ and $\{g_{k,q}(w^k)\}$ in the network. Therefore, if a global constraint function $h_u(w)$ or $g_q(w)$ relates to some action w_k, agent k is subject to the same constraint function, say, $h_{k,u'}(w^k) = h_u(w)$ or $g_{k,q'}(w^k) = g_q(w)$. That is, we can find one-to-one mapping from every nonzero $\nabla_{w_k^\mathsf{T}} h_u(w)$ or $\nabla_{w_k^\mathsf{T}} g_q(w)$ in (2.124) to some $\nabla_{w_k^\mathsf{T}} h_{k,u'}(w^k)$ or $\nabla_{w_k^\mathsf{T}} g_{k,q'}(w^k)$ in (2.125), which means that we have

$$\nabla_{w_k^\mathsf{T}} p_k(w^k) = \nabla_{w_k^\mathsf{T}} p(w) \tag{2.126}$$

so that

$$\nabla_{w^\mathsf{T}} p(w) = \mathrm{col}\{\nabla_{w_1^\mathsf{T}} p_1(w^1), ..., \nabla_{w_N^\mathsf{T}} p_N(w^N)\} \tag{2.127}$$

From (2.21) and (2.33) we can write

$$\begin{aligned}
F^p(w) & \triangleq \mathrm{col}\{\nabla_{w_1^\mathsf{T}} J_1^p(w^1), ..., \nabla_{w_N^\mathsf{T}} J_N^p(w^N)\} \\
& = F(w) + \rho \nabla_{w^\mathsf{T}} p(w)
\end{aligned} \tag{2.128}$$

56

Since the sum of convex functions is also convex, we know that $p(w)$ is convex and, therefore, for any w^a and w^b:

$$(w^a - w^b)^\mathsf{T}[\nabla_{w^\mathsf{T}} p(w^a) - \nabla_{w^\mathsf{T}} p(w^b)] \geq 0 \tag{2.129}$$

Using (2.13) we get

$$(w^a - w^b)^\mathsf{T}[F^p(w^a) - F^p(w^b)]$$
$$= (w^a - w^b)^\mathsf{T}[F(w^a) - F(w^b) + \nabla_{w^\mathsf{T}} p(w^a) - \nabla_{w^\mathsf{T}} p(w^b)]$$
$$\geq \nu\|w^a - w^b\|^2 \tag{2.130}$$

It follows that the penalized mapping $F^p : \mathbb{R}^M \to \mathbb{R}^M$ is strongly monotone. In order to examine the existence of a Nash equilibrium, we need to show that the strong monotonicity of $F^p(w)$ satisfies the coerciveness property [88, p. 14], i.e., for some $w^{\text{ref}} \in \mathbb{R}^M$,

$$\lim_{\|w\| \to \infty} \frac{\left[F^p(w) - F^p(w^{\text{ref}})\right]^\mathsf{T}(w - w^{\text{ref}})}{\|w - w^{\text{ref}}\|} = \infty \tag{2.131}$$

Using (2.130) and setting $w^a = w$ and $w^b = w^{\text{ref}}$ we get

$$\lim_{\|w\| \to \infty} \frac{\left[F^p(w) - F^p(w^{\text{ref}})\right]^\mathsf{T}(w - w^{\text{ref}})}{\|w - w^{\text{ref}}\|} \geq \lim_{\|w\| \to \infty} \nu\|w - w^{\text{ref}}\| = \infty \tag{2.132}$$

which shows that $F^p(w)$ satisfies the coerciveness property in (2.131) with w^{ref}. We then conclude the existence of solutions to problem (2.27).

The uniqueness of the Nash equilibrium is also guaranteed by the strong monotonicity following [32, Theorem 2.3.3]. Since $J_k^p(w_k; w_{-k})$ is convex and differentiable in w_k, from the optimality criterion [33] we know that the Nash equilibrium satisfies

$$(w_k' - w_k^\star)^\mathsf{T} \nabla_{w_k} J_k^p(w_k^\star; w_{-k}^\star) \geq 0 \tag{2.133}$$

for all feasible w_k'. Summing up these conditions over all agents we get

$$\sum_{k=1}^{N} (w_k' - w_k^\star)^\mathsf{T} \nabla_{w_k} J_k^p(w_k^\star; w_{-k}^\star) = (w' - w^\star)^\mathsf{T} F^p(w^\star) \geq 0 \tag{2.134}$$

Let us first assume the existence of two distinct solutions, $w^\star \neq w^\dagger \in \mathbb{R}^M$. Then, for any $w' \in \mathbb{R}^M$ they will satisfy

$$(w' - w^\star)^\mathsf{T} F^p(w^\star) \geq 0, \quad (w' - w^\dagger)^\mathsf{T} F^p(w^\dagger) \geq 0 \tag{2.135}$$

Setting $w' = w^\dagger$ in the first inequality and $w' = w^\star$ in the second inequality we get

$$(w^\dagger - w^\star)^\mathsf{T} F^p(w^\star) \geq 0, \quad (w^\star - w^\dagger)^\mathsf{T} F^p(w^\dagger) \geq 0 \tag{2.136}$$

By adding these two inequalities, we arrive at

$$(w^\dagger - w^\star)^\mathsf{T}[F^p(w^\dagger) - F^p(w^\star)] \leq 0 \tag{2.137}$$

which contradicts the strong monotonicity of $F^p(w)$. We thus conclude that the Nash equilibrium is unique.

Now, from the optimality criterion [33] and given $w^\star_{-k}$, we note that $w^\star_k$ is optimal if, and only if,

$$\nabla_{w_k} J^p_k(w^\star_k; w^\star_{-k}) = 0 \tag{2.138}$$

Collecting these conditions for all agents we obtain

$$F^p(w^\star) = F(w^\star) + \rho \nabla_{w^\mathsf{T}} p(w^\star) = 0 \tag{2.139}$$

2.B Proof of Lemma 2.1

Using Condition 2.1, we have

$$
\begin{aligned}
\|\nabla_{w^\mathsf{T}} p(w^\circ) - \nabla_{w^\mathsf{T}} p(w^\bullet)\|^2 &= \sum_{k=1}^N \left\| \nabla_{w^\mathsf{T}_k} p_k(w^k_\circ) - \nabla_{w^\mathsf{T}_k} p_k(w^k_\bullet) \right\|^2 \\
&\leq \sum_{k=1}^N \gamma_k^2 \|w^k_\circ - w^k_\bullet\|^2 \\
&\leq \delta_p^2 \|w^\circ - w^\bullet\|^2
\end{aligned}
\tag{2.140}
$$

where we used the fact that $\|w^k_\circ - w^k_\bullet\|^2 \leq \|w^\circ - w^\bullet\|^2$. Then, it follows that

$$
\begin{aligned}
\|F^p(w^\circ) - F^p(w^\bullet)\| &\leq \|F(w^\circ) - F(w^\bullet)\| + \rho\|\nabla_{w^\mathsf{T}} p(w^\circ) - \nabla_{w^\mathsf{T}} p(w^\bullet)\| \\
&\leq (\delta + \rho\delta_p)\|w^\circ - w^\bullet\|
\end{aligned}
\tag{2.141}
$$

as claimed.

2.C   Proof of Lemma 2.2

We first note that

$$(w^\circ - w^\bullet)^\mathsf{T} U[F^p(w^\circ) - F^p(w^\bullet)]$$

$$= \sum_{k=1}^{N} \mu_k (w_k^\circ - w_k^\bullet)^\mathsf{T} \left[\nabla_{w_k^\mathsf{T}} J_k^p(w_\circ^k) - \nabla_{w_k^\mathsf{T}} J_k^p(w_\bullet^k) \right]$$

$$= \mu_{\max} \sum_{k=1}^{N} (w_k^\circ - w_k^\bullet)^\mathsf{T} \left[\nabla_{w_k^\mathsf{T}} J_k^p(w_\circ^k) - \nabla_{w_k^\mathsf{T}} J_k^p(w_\bullet^k) \right]$$

$$- \sum_{k=1}^{N} (\mu_{\max} - \mu_k)(w_k^\circ - w_k^\bullet)^\mathsf{T} \left[\nabla_{w_k^\mathsf{T}} J_k^p(w_\circ^k) - \nabla_{w_k^\mathsf{T}} J_k^p(w_\bullet^k) \right] \tag{2.142}$$

Using the Cauchy-Schwartz inequality we get

$$\sum_{k=1}^{N} (\mu_{\max} - \mu_k)(w_k^\circ - w_k^\bullet)^\mathsf{T} \left[\nabla_{w_k^\mathsf{T}} J_k^p(w_\circ^k) - \nabla_{w_k^\mathsf{T}} J_k^p(w_\bullet^k) \right]$$

$$\leq (\mu_{\max} - \mu_{\min}) \sum_{k=1}^{N} (w_k^\circ - w_k^\bullet)^\mathsf{T} \left[\nabla_{w_k^\mathsf{T}} J_k^p(w_\circ^k) - \nabla_{w_k^\mathsf{T}} J_k^p(w_\bullet^k) \right]$$

$$\leq t\mu_{\max} \sum_{k=1}^{N} \|w_k^\circ - w_k^\bullet\| \cdot \left\| \nabla_{w_k^\mathsf{T}} J_k^p(w_\circ^k) - \nabla_{w_k^\mathsf{T}} J_k^p(w_\bullet^k) \right\|$$

$$\overset{(a)}{\leq} t\mu_{\max} \left(\sum_{k=1}^{N} \|w_k^\circ - w_k^\bullet\|^2 \right)^{\frac{1}{2}}$$

$$\times \left(\sum_{k=1}^{N} \left\| \nabla_{w_k^\mathsf{T}} J_k^p(w_\circ^k) - \nabla_{w_k^\mathsf{T}} J_k^p(w_\bullet^k) \right\|^2 \right)^{\frac{1}{2}}$$

$$= t\mu_{\max} \|w^\circ - w^\bullet\| \cdot \|F^p(w^\circ) - F^p(w^\bullet)\| \tag{2.143}$$

where (a) is obtained from Hölder's inequality [89]. By (2.143) we have

$$(w^\circ - w^\bullet)^\mathsf{T} U[F^p(w^\circ) - F^p(w^\bullet)]$$

$$\geq \mu_{\max}(w^\circ - w^\bullet)^\mathsf{T}[F^p(w^\circ) - F^p(w^\bullet)] - t\mu_{\max}\|w^\circ - w^\bullet\| \cdot \|F^p(w^\circ) - F^p(w^\bullet)\|$$

$$\geq \mu_{\max}\nu\|w^\circ - w^\bullet\|^2 - t\mu_{\max}(\delta + \rho\delta_p)\|w^\circ - w^\bullet\|^2$$

$$= \mu_{\max}[\nu - t(\delta + \rho\delta_p)] \cdot \|w^\circ - w^\bullet\|^2 \tag{2.144}$$

where we used the strong monotonicity property (2.130) and the Lipschitz continuous property (2.141). Similarly, we can express

$$(w^\circ - w^\bullet)^\mathsf{T} U[F(w^\circ) - F(w^\bullet)]$$

$$\geq \mu_{\max}(w^\circ - w^\bullet)^\mathsf{T}[F(w^\circ) - F(w^\bullet)] - t\mu_{\max}\|w^\circ - w^\bullet\| \cdot \|F(w^\circ) - F(w^\bullet)\|$$

$$\geq \mu_{\max}(\nu - t\delta)\|w^\circ - w^\bullet\|^2 \tag{2.145}$$

and

$$(w^\circ - w^\bullet)^\mathsf{T} U[\nabla_{w^\mathsf{T}} p(w^\circ) - \nabla_{w^\mathsf{T}} p(w^\bullet)]$$

$$\geq \mu_{\max}(w^\circ - w^\bullet)^\mathsf{T}[\nabla_{w^\mathsf{T}} p(w^\circ) - \nabla_{w^\mathsf{T}} p(w^\bullet)]$$

$$- t\mu_{\max}\|w^\circ - w^\bullet\| \cdot \|\nabla_{w^\mathsf{T}} p(w^\circ) - \nabla_{w^\mathsf{T}} p(w^\bullet)\|$$

$$\geq -t\mu_{\max}\delta_p\|w^\circ - w^\bullet\|^2 \tag{2.146}$$

where we used Assumptions 2.1 and 2.2, the convexity property (2.129), and the Lipschitz continuous property (2.140).

2.D   Proof of Theorem 2.2

We first note that assumption (2.76) can be rewritten as

$$\mathbb{E}\left[\|\boldsymbol{s}_i(\boldsymbol{w}_{i-1})\|^2|\boldsymbol{\mathcal{F}}_{i-1}\right] \leq \alpha\|\boldsymbol{w}_{i-1} - w^\star + w^\star\|^2 + \beta$$

$$\leq 2\alpha\|\widetilde{\boldsymbol{w}}_{i-1}\|^2 + \beta' \tag{2.147}$$

where we used $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$ and introduced $\beta' \triangleq \beta + 2\alpha\|w^\star\|^2$. Then, using properties (2.75), (2.76), (2.65), (2.68), and the fact that $F^p(w^\star) = 0$, we can express the mean-square error $\mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|^2$ from (2.82) as

$$\mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|^2 = \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}\|^2 - 2\,\mathbb{E}\left[\widetilde{\boldsymbol{w}}_{i-1}^\mathsf{T} U(F^p(w^\star) - F^p(\boldsymbol{w}_{i-1}))\right]$$

$$+ \mathbb{E}\|F^p(w^\star) - F^p(\boldsymbol{w}_{i-1})\|_{U^2}^2 + \mathbb{E}\|\boldsymbol{s}_i(\boldsymbol{w}_{i-1})\|_{U^2}^2$$

$$\leq \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}\|^2 - 2\,\mathbb{E}\left[\widetilde{\boldsymbol{w}}_{i-1}^\mathsf{T} U(F^p(w^\star) - F^p(\boldsymbol{w}_{i-1}))\right]$$

$$+ \mu_{\max}^2 \mathbb{E}\|F^p(w^\star) - F^p(\boldsymbol{w}_{i-1})\|^2 + \mu_{\max}^2 \mathbb{E}\|\boldsymbol{s}_i(\boldsymbol{w}_{i-1})\|^2$$

$$\leq \left(1 - 2\mu_{\max}\nu' + \mu_{\max}^2[(\delta + \rho\delta_p)^2 + 2\alpha]\right)\mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}\|^2 + \mu_{\max}^2\beta \tag{2.148}$$

60

Note that from $\delta \geq \nu$ in (2.18) we have

$$
1 - 2\mu_{\text{max}}\nu' + \mu_{\text{max}}^2[(\delta + \rho\delta_p)^2 + 2\alpha]
$$

$$
= 1 - 2\mu_{\text{max}}[\nu - t(\delta + \rho\delta_p)] + \mu_{\text{max}}^2[(\delta + \rho\delta_p)^2 + 2\alpha]
$$

$$
= (1 - \mu_{\text{max}}\nu)^2 + \mu_{\text{max}}^2((\delta + \rho\delta_p)^2 + 2\alpha - \nu^2) + 2\mu_{\text{max}}t(\delta + \rho\delta_p) \geq 0 \qquad (2.149)
$$

Therefore, the mean-square error is stable asymptotically, as $i \to \infty$, when the step-size μ_{max} satisfies

$$
|1 - 2\mu_{\text{max}}\nu' + \mu_{\text{max}}^2[(\delta + \rho\delta_p)^2 + 2\alpha]| < 1
$$

$$
\Longleftrightarrow \quad -1 < 1 - 2\mu_{\text{max}}\nu' + \mu_{\text{max}}^2[(\delta + \rho\delta_p)^2 + 2\alpha] < 1
$$

$$
\Longleftrightarrow \quad 0 < \mu_{\text{max}} < \frac{2\nu'}{(\delta + \rho\delta_p)^2 + 2\alpha} \qquad (2.150)
$$

when ν' is positive, i.e.,

$$
\nu' = \nu - t(\delta + \rho\delta_p) > 0 \quad \Longleftrightarrow \quad t < \frac{\nu}{\delta + \rho\delta_p} \qquad (2.151)
$$

This leads to the conditions in (2.83), and the resulting mean-squared error is upper bounded by

$$
\limsup_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|^2 \leq \frac{\mu_{\text{max}}\beta}{2\nu' - \mu_{\text{max}}[(\delta + \rho\delta_p)^2 + 2\alpha]} = O(\mu_{\text{max}}) \qquad (2.152)
$$

2.E   Proof of Theorem 2.3

Let us consider two unequal vectors w_{i-1}° and $w_{i-1}^\bullet$ with corresponding vectors $\{\phi_i^\circ, \psi_i^\circ, w_i^\circ\}$ and $\{\phi_i^\bullet, \psi_i^\bullet, w_i^\bullet\}$ in implementation (2.85)–(2.87). The squared Euclidean distance between ϕ_i° and $\phi_i^\bullet$ is given by

$$
\|\phi_i^\circ - \phi_i^\bullet\|^2 = \|(w_{i-1}^\circ - w_{i-1}^\bullet) - c_1\rho U[\nabla_{w^\top}p(w_{i-1}^\circ) - \nabla_{w^\top}p(w_{i-1}^\bullet)]\|^2
$$

$$
\leq \|w_{i-1}^\circ - w_{i-1}^\bullet\|^2 + c_1^2\mu_{\text{max}}^2\rho^2\|\nabla_{w^\top}p(w_{i-1}^\circ) - \nabla_{w^\top}p(w_{i-1}^\bullet)\|^2
$$

$$
- 2c_1\rho(w_{i-1}^\circ - w_{i-1}^\bullet)^\top U[\nabla_{w^\top}p(w_{i-1}^\circ) - \nabla_{w^\top}p(w_{i-1}^\bullet)]
$$

$$
\leq (1 + 2c_1t\mu_{\text{max}}\rho\delta_p + c_1\mu_{\text{max}}^2\rho^2\delta_p^2)\|w_{i-1}^\circ - w_{i-1}^\bullet\|^2 \qquad (2.153)
$$

61

where we used the properties (2.140), (2.70), and $c_1^2 = c_1$ from (2.88). Using similar arguments we have

$$\|w_i^\circ - w_i^\bullet\|^2 \le (1 + 2c_2 t \mu_{\max} \rho \delta_p + c_2 \mu_{\max}^2 \rho^2 \delta_p^2) \|\psi_i^\circ - \psi_i^\bullet\|^2 \tag{2.154}$$

For ψ_i° and $\psi_i^\bullet$, we can write

$$\begin{aligned}
\|\psi_i^\circ - \psi_i^\bullet\|^2 &= \|(\phi_i^\circ - \phi_i^\bullet) - U(F(\phi_i^\circ) - F(\phi_i^\bullet))\|^2 \\
&\le \|\phi_i^\circ - \phi_i^\bullet\|^2 - 2(\phi_i^\circ - \phi_i^\bullet)^\mathsf{T} U(F(\phi_i^\circ) - F(\phi_i^\bullet)) + \mu_{\max}^2 \|F(\phi_i^\circ) - F(\phi_i^\bullet)\|^2 \\
&\le (1 - 2\mu_{\max} \nu'' + \mu_{\max}^2 \delta^2) \|\phi_i^\circ - \phi_i^\bullet\|^2 \tag{2.155}
\end{aligned}$$

where we used (2.69) and Assumption 2.2. Combining (2.153), (2.154), and (2.155) we get

$$\begin{aligned}
\|w_i^\circ - w_i^\bullet\|^2 &\le (1 + 2c_1 t \mu_{\max} \rho \delta_p + c_1 \mu_{\max}^2 \rho^2 \delta_p^2)(1 + 2c_2 t \mu_{\max} \rho \delta_p + c_2 \mu_{\max}^2 \rho^2 \delta_p^2) \\
&\quad \times (1 - 2\mu_{\max} \nu'' + \mu_{\max}^2 \delta^2) \|w_{i-1}^\circ - w_{i-1}^\bullet\|^2 \\
&= (1 + 2t \mu_{\max} \rho \delta_p + \mu_{\max}^2 \rho^2 \delta_p^2)(1 - 2\mu_{\max} \nu'' + \mu_{\max}^2 \delta^2) \|w_{i-1}^\circ - w_{i-1}^\bullet\|^2 \tag{2.156}
\end{aligned}$$

The mapping $w_{i-1} \mapsto w_i$ is a contraction if

$$\begin{aligned}
&|(1 + 2t \mu_{\max} \rho \delta_p + \mu_{\max}^2 \rho^2 \delta_p^2)(1 - 2\mu_{\max} \nu'' + \mu_{\max}^2 \delta^2)| < 1 \\
\iff\quad &-1 < (1 + 2t \mu_{\max} \rho \delta_p + \mu_{\max}^2 \rho^2 \delta_p^2)(1 - 2\mu_{\max} \nu'' + \mu_{\max}^2 \delta^2) < 1 \tag{2.157}
\end{aligned}$$

We note that

$$1 - 2\mu_{\max} \nu'' + \mu_{\max}^2 \delta^2 = (1 - \mu_{\max} \nu)^2 + \mu_{\max}^2 (\delta^2 - \nu^2) + 2\mu_{\max} t \delta \ge 0 \tag{2.158}$$

Therefore, the inequality on the left-hand side of (2.157) always holds. Expanding the product of the two terms and using $\nu' = \nu'' - t\rho\delta_p$ we get for the inequality on the right-hand side of (2.157) that we must have

$$1 - a_1 < 1 \iff a_1 > 0 \tag{2.159}$$

where

$$a_1 \triangleq 2\mu_{\max} \nu' - \mu_{\max}^2 (\delta^2 + \rho^2 \delta_p^2 - 4t\nu'' \rho \delta_p) + \mu_{\max}^3 (\rho^2 \delta_p^2 \nu'' - t\rho \delta_p \delta^2) - \mu_{\max}^4 \rho^2 \delta_p^2 \delta^2 \tag{2.160}$$

62

Therefore, if we can guarantee

$$\nu' > 0 \tag{2.161}$$

$$\delta^2 + \rho^2 \delta_p^2 - 4t\nu'' \rho \delta_p > 0 \tag{2.162}$$

$$\mu_{\max}^3 (\rho^2 \delta_p^2 \nu'' - t\rho\delta_p \delta^2) > \mu_{\max}^4 \rho^2 \delta_p^2 \delta^2 \tag{2.163}$$

then the condition $a_1 > 0$ is satisfied if

$$2\mu_{\max}\nu' - \mu_{\max}^2(\delta^2 + \rho^2\delta_p^2 - 4t\nu''\rho\delta_p) > 0 \tag{2.164}$$

which means

$$\mu_{\max} < \frac{2\nu'}{\delta^2 + \rho^2\delta_p^2 - 4t\nu''\rho\delta_p} \tag{2.165}$$

Let us examine conditions (2.161)–(2.163). From (2.151) we know that the condition of a positive ν' holds if

$$t < \frac{\nu}{\delta + \rho\delta_p} \tag{2.166}$$

For the second condition (2.162), we now show that if ρ is sufficiently large such that

$$\rho > \delta/\delta_p \iff \rho\delta_p > \delta \tag{2.167}$$

then

$$f(t) \triangleq \delta^2 + \rho^2\delta_p^2 - 4t\nu''\rho\delta_p$$
$$= 4t^2\delta\rho\delta_p - 4t\nu\rho\delta_p + \delta^2 + \rho^2\delta_p^2 > \delta^2 > 0 \tag{2.168}$$

where we used $\nu'' = \nu - t\delta$. Note that $f(t)$ is a quadratic function of t and has a minimum at $t^o = \nu/(2\delta)$. Therefore, it is required that

$$f(t^o) = \delta^2 + \rho^2\delta_p^2 - \frac{\nu^2}{\delta}\rho\delta_p > \delta^2 \iff \rho\delta_p > \frac{\nu^2}{\delta} \tag{2.169}$$

Under condition (2.167) and from the fact $\delta \geq \nu$, we get

$$\delta\rho\delta_p > \delta^2 \geq \nu^2 \implies \rho\delta_p > \frac{\nu^2}{\delta} \tag{2.170}$$

63

which ensures $f(t) > \delta^2 > 0$ for any t. For the third condition (2.163), we first note that we need the left-hand side of (2.163) to be positive, which requires

$$\rho^2 \delta_p^2 \nu'' - t\rho\delta_p \delta^2 > 0 \quad \Longleftrightarrow \quad \rho\delta_p \nu' + t(\rho^2 \delta_p^2 - \delta^2) > 0 \tag{2.171}$$

where we used $\nu'' = \nu' + t\rho\delta_p$. By (2.167) and (2.161) we know that (2.171) holds. Then, we have

$$\rho^2 \delta_p^2 \nu'' - t\rho\delta_p \delta^2 > \mu_{\max}\rho^2 \delta_p^2 \delta^2 \quad \Longleftrightarrow \quad \rho\delta_p \nu' + t(\rho^2 \delta_p^2 - \delta^2) > \mu_{\max}\rho\delta_p \delta^2$$

$$\Longleftrightarrow \quad \mu_{\max} < \frac{\nu' + \frac{t(\rho^2 \delta_p^2 - \delta^2)}{\rho\delta_p}}{\delta^2} \tag{2.172}$$

Therefore, we arrive at the following sufficient conditions for the convergence of (2.85)–(2.87):

$$0 < \mu_{\max} < \mu_o, \quad t < \frac{\nu}{\delta + \rho\delta_p}, \quad \rho > \frac{\delta}{\delta_p} \tag{2.173}$$

where

$$\mu_o \triangleq \min \left\{ \frac{2\nu'}{\delta^2 + \rho^2 \delta_p^2 - 4t\nu''\rho\delta_p}, \frac{\nu' + \frac{t(\rho^2 \delta_p^2 - \delta^2)}{\rho\delta_p}}{\delta^2} \right\} \tag{2.174}$$

2.F   Proof of Theorem 2.4

Subtracting (2.96)–(2.98) from (2.93)–(2.95) we get

$$\widetilde{\boldsymbol{\phi}}_i^\infty = \widetilde{\boldsymbol{w}}_{i-1}^\infty - c_1 \rho U \left[\nabla_{w^\intercal} p(w^\infty) - \nabla_{w^\intercal} p(\boldsymbol{w}_{i-1}) \right] \tag{2.175}$$

$$\widetilde{\boldsymbol{\psi}}_i^\infty = \widetilde{\boldsymbol{\phi}}_i^\infty - U \left[F(\phi^\infty) - F(\boldsymbol{\phi}_i) \right] + U \boldsymbol{s}_i(\boldsymbol{\phi}_i) \tag{2.176}$$

$$\widetilde{\boldsymbol{w}}_i^\infty = \widetilde{\boldsymbol{\psi}}_i^\infty - c_2 \rho U \left[\nabla_{w^\intercal} p(\psi^\infty) - \nabla_{w^\intercal} p(\boldsymbol{\psi}_i) \right] \tag{2.177}$$

where $\widetilde{\boldsymbol{\phi}}_i^\infty \triangleq \phi^\infty - \boldsymbol{\phi}_i$, $\widetilde{\boldsymbol{\psi}}_i^\infty \triangleq \psi^\infty - \boldsymbol{\psi}_i$, and $\widetilde{\boldsymbol{w}}_i^\infty \triangleq w^\infty - \boldsymbol{w}_i$. From (2.175) we have

$$\mathbb{E}\|\widetilde{\boldsymbol{\phi}}_i^\infty\|^2 \leq \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}^\infty\|^2 + c_1^2 \mu_{\max}^2 \rho^2 \mathbb{E}\|\nabla_{w^\intercal} p(w^\infty) - \nabla_{w^\intercal} p(\boldsymbol{w}_{i-1})\|^2$$

$$- 2c_1 \rho \mathbb{E}\left[\widetilde{\boldsymbol{w}}_{i-1}^{\infty\intercal} U \left[\nabla_{w^\intercal} p(w^\infty) - \nabla_{w^\intercal} p(\boldsymbol{w}_{i-1}) \right] \right]$$

$$\leq (1 + 2c_1 t\mu_{\max}\rho\delta_p + c_1 \mu_{\max}^2 \rho^2 \delta_p^2) \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}^\infty\|^2 \tag{2.178}$$

and, similarly, from (2.177) we obtain

$$\mathbb{E}\|\widetilde{\boldsymbol{w}}_i^\infty\|^2 \leq (1 + 2c_2 t\mu_{\max}\rho\delta_p + c_2\mu_{\max}^2\rho^2\delta_p^2)\mathbb{E}\|\widetilde{\boldsymbol{\psi}}_i^\infty\|^2 \tag{2.179}$$

Similar to (2.147), we can rewrite assumption (2.76) as

$$\mathbb{E}\left[\|\boldsymbol{s}_i(\boldsymbol{w}_{i-1})\|^2|\boldsymbol{\mathcal{F}}_{i-1}\right] \leq 2\alpha\|\widetilde{\boldsymbol{w}}_{i-1}^\infty\|^2 + \beta'' \tag{2.180}$$

for $\beta'' \triangleq \beta + 2\alpha\|w^\infty\|^2$. Then, from (2.176) we obtain:

$$\begin{aligned}
\mathbb{E}\|\widetilde{\boldsymbol{\psi}}_i^\infty\|^2 &\leq \mathbb{E}\|\widetilde{\boldsymbol{\phi}}_i^\infty\|^2 + \mu_{\max}^2\mathbb{E}\|F(\phi^\infty) - F(\boldsymbol{\phi}_i)\|^2 \\
&\quad + \mu_{\max}^2\mathbb{E}\|\boldsymbol{s}(\boldsymbol{\phi}_i)\|^2 - 2\mathbb{E}\left[\widetilde{\boldsymbol{\phi}}_i^{\infty\mathsf{T}}U[F(\phi^\infty) - F(\boldsymbol{\phi}_i)]\right] \\
&\leq \left(1 - 2\mu_{\max}\nu'' + \mu_{\max}^2(\delta^2 + 2\alpha)\right)\mathbb{E}\|\widetilde{\boldsymbol{\phi}}_i^\infty\|^2 + \mu_{\max}^2\beta'' \tag{2.181}
\end{aligned}$$

Therefore, we can combine (2.178)–(2.181) to get

$$\begin{aligned}
\mathbb{E}\|\widetilde{\boldsymbol{w}}_i^\infty\|^2 &\leq (1 + 2c_1 t\mu_{\max}\rho\delta_p + c_1\mu_{\max}^2\rho^2\delta_p^2)(1 + 2c_2 t\mu_{\max}\rho\delta_p + c_2\mu_{\max}^2\rho^2\delta_p^2) \\
&\quad \times \left(1 - 2\mu_{\max}\nu'' + \mu_{\max}^2(\delta^2 + 2\alpha)\right)\mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}^\infty\|^2 \\
&\quad + \mu_{\max}^2(1 + 2c_2 t\mu_{\max}\rho\delta_p + c_2\mu_{\max}^2\rho^2\delta_p^2)\beta'' \\
&= (1 + 2t\mu_{\max}\rho\delta_p + \mu_{\max}^2\rho^2\delta_p^2)\left(1 - 2\mu_{\max}\nu'' + \mu_{\max}^2(\delta^2 + 2\alpha)\right) \\
&\quad \times \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}^\infty\|^2 + \mu_{\max}^2(1 + 2c_2 t\mu_{\max}\rho\delta_p + c_2\mu_{\max}^2\rho^2\delta_p^2)\beta'' \tag{2.182}
\end{aligned}$$

We expand the product of the two terms as

$$(1 + 2t\mu_{\max}\rho\delta_p + \mu_{\max}^2\rho^2\delta_p^2)(1 - 2\mu_{\max}\nu'' + \mu_{\max}^2(\delta^2 + 2\alpha)) \triangleq 1 - a_2 \tag{2.183}$$

where

$$\begin{aligned}
a_2 &\triangleq 2\mu_{\max}\nu' - \mu_{\max}^2(\delta^2 + 2\alpha + \rho^2\delta_p^2 - 4t\nu''\rho\delta_p) \\
&\quad + \mu_{\max}^3(\rho^2\delta_p^2\nu'' - t\rho\delta_p(\delta^2 + 2\alpha)) - \mu_{\max}^4\rho^2\delta_p^2(\delta^2 + 2\alpha) \tag{2.184}
\end{aligned}$$

Then, the mean-square error $\mathbb{E}\|\widetilde{\boldsymbol{w}}_i^\infty\|^2$ converges asymptotically as $i \to \infty$ if we have $|1-a_2| < 1$, which requires $a_2 > 0$ since from (2.158) we know $1-a_2 \geq 0$. Following a similar argument

65

to the one presented in Appendix 2.E, we obtain that the following conditions ensure the convergence of $\mathbb{E}\|\widetilde{\boldsymbol{w}}_i^\infty\|^2$:

$$\nu' > 0 \tag{2.185}$$

$$\delta^2 + 2\alpha + \rho^2\delta_p^2 - 4t\nu''\rho\delta_p > 0 \tag{2.186}$$

$$\mu_{\max}^3(\rho^2\delta_p^2\nu'' - t\rho\delta_p(\delta^2 + 2\alpha)) > \mu_{\max}^4\rho^2\delta_p^2(\delta^2 + 2\alpha) \tag{2.187}$$

$$2\mu_{\max}\nu' - \mu_{\max}^2(\delta^2 + 2\alpha + \rho^2\delta_p^2 - 4t\nu''\rho\delta_p) > 0 \tag{2.188}$$

The first two yield the same results in (2.166) and (2.167), i.e.,

$$t < \frac{\nu}{\delta + \rho\delta_p}, \qquad \rho > \frac{\delta}{\delta_p} \tag{2.189}$$

For the third condition we need to ensure

$$\rho^2\delta_p^2\nu'' - t\rho\delta_p(\delta^2 + 2\alpha) > 0 \quad \Longleftrightarrow \quad \rho\delta_p\nu' + t(\rho^2\delta_p^2 - (\delta^2 + 2\alpha)) > 0 \tag{2.190}$$

A stricter condition on ρ is therefore required:

$$\rho > \frac{\sqrt{\delta^2 + 2\alpha}}{\delta_p} \tag{2.191}$$

We then get

$$\rho\delta_p\nu'' - t(\delta^2 + 2\alpha) > \mu_{\max}\rho\delta_p(\delta^2 + 2\alpha) \quad \Longleftrightarrow \quad \mu_{\max} < \frac{\nu' + \frac{t(\rho^2\delta_p^2 - (\delta^2 + 2\alpha))}{\rho\delta_p}}{\delta^2 + 2\alpha} \tag{2.192}$$

Combining the last condition (2.188), we get the step-size condition as

$$0 < \mu_{\max} < \mu_o' \tag{2.193}$$

where

$$\mu_o' \triangleq \min\left\{\frac{2\nu'}{\delta^2 + 2\alpha + \rho^2\delta_p^2 - 4t\nu''\rho\delta_p}, \frac{\nu' + \frac{t(\rho^2\delta_p^2 - (\delta^2 + 2\alpha))}{\rho\delta_p}}{\delta^2 + 2\alpha}\right\} \tag{2.194}$$

Therefore, under conditions (2.189), (2.191), and (2.193), the recursion (2.182) is stable and the resulting mean-square error is upper bounded by

$$\begin{aligned}
\limsup_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_i^\infty\|^2 &\leq \frac{\mu_{\max}^2(1 + 2c_2 t\mu_{\max}\rho\delta_p + c_2\mu_{\max}^2\rho^2\delta_p^2)\beta''}{a_2} \\
&= \frac{\mu_{\max}(1 + 2c_2 t\mu_{\max}\rho\delta_p + c_2\mu_{\max}^2\rho^2\delta_p^2)\beta''}{2\nu' - O(\mu_{\max})} \\
&= O(\mu_{\max}) \tag{2.195}
\end{aligned}$$

for sufficiently small step-sizes.

2.G Proof of Theorem 2.5

We recall from (2.33) that the Nash equilibrium $w^\star$ satisfies the relation:

$$
\begin{aligned}
w^\star &= w^\star - U\left[F(w^\star) + \rho\nabla_{w^\intercal}p(w^\star)\right] \\
&= w^\star - UF(w^\star) - c_1\rho U\nabla_{w^\intercal}p(w^\star) - c_2\rho U\nabla_{w^\intercal}p(w^\star) \\
&= \phi^\star - UF(w^\star) - c_2\rho U\nabla_{w^\intercal}p(w^\star) \\
&= \psi^\star - c_2\rho U\nabla_{w^\intercal}p(w^\star)
\end{aligned}
\tag{2.196}
$$

where we introduced two auxiliary variables $\phi^\star$ and $\psi^\star$:

$$
\phi^\star = w^\star - c_1\rho U\nabla_{w^\intercal}p(w^\star)
\tag{2.197}
$$

$$
\psi^\star = \phi^\star - UF(w^\star)
\tag{2.198}
$$

If we further introduce the error vectors $\widetilde{\phi} \triangleq \phi^\star - \phi^\infty$, $\widetilde{\psi} \triangleq \psi^\star - \psi^\infty$, and $\widetilde{w} \triangleq w^\star - w^\infty$, then using (2.196)–(2.198) we have

$$
\widetilde{\phi} = \widetilde{w} - c_1\rho U[\nabla_{w^\intercal}p(w^\star) - \nabla_{w^\intercal}p(w^\infty)]
\tag{2.199}
$$

$$
\widetilde{\psi} = \widetilde{\phi} - U\left[F(w^\star) - F(\phi^\infty)\right]
\tag{2.200}
$$

$$
\widetilde{w} = \widetilde{\psi} - c_2\rho U[\nabla_{w^\intercal}p(w^\star) - \nabla_{w^\intercal}p(\psi^\infty)]
\tag{2.201}
$$

From (2.199), the squared norm of $\widetilde{\phi}$ satisfies

$$
\begin{aligned}
\|\widetilde{\phi}\|^2 &\leq \|\widetilde{w}\|^2 - 2c_1\rho\widetilde{w}^\mathsf{T}U[\nabla_{w^\intercal}p(w^\star) - \nabla_{w^\intercal}p(w^\infty)] \\
&\quad + c_1\mu_{\max}^2\rho^2\|\nabla_{w^\intercal}p(w^\star) - \nabla_{w^\intercal}p(w^\infty)\|^2 \\
&\leq \mathcal{Y}_1\|\widetilde{w}\|^2
\end{aligned}
\tag{2.202}
$$

where we used (2.140) and (2.70) and introduced

$$
\mathcal{Y}_1 \triangleq 1 + 2c_1 t\mu_{\max}\rho\delta_p + c_1\mu_{\max}^2\rho^2\delta_p^2
\tag{2.203}
$$

From (2.200), the squared norm of $\widetilde{\psi}$ satisfies

$$
\|\widetilde{\psi}\|^2 = \|\widetilde{\phi}\|^2 - 2\widetilde{\phi}^\mathsf{T}U[F(w^\star) - F(\phi^\infty)] + \|U[F(w^\star) - F(\phi^\infty)]\|^2
\tag{2.204}
$$

67

We note that

$$-2\widetilde{\phi}^{\mathsf{T}}U[F(w^\star) - F(\phi^\infty)] = -2\widetilde{\phi}^{\mathsf{T}}U[F(\phi^\star) - F(\phi^\infty)] - 2\widetilde{\phi}^{\mathsf{T}}U[F(w^\star) - F(\phi^\star)]$$

$$\stackrel{(a)}{\leq} -2\mu_{\max}\nu''\|\widetilde{\phi}\|^2 + 2\mu_{\max}\|F(w^\star) - F(\phi^\star)\| \cdot \|\widetilde{\phi}\|$$

$$\stackrel{(b)}{\leq} -2\mu_{\max}\nu''\|\widetilde{\phi}\|^2 + 2c_1\mu_{\max}^2\delta\|F(w^\star)\| \cdot \|\widetilde{\phi}\| \qquad (2.205)$$

where step (a) is from (2.69) and Hölder's inequality and step (b) is due to

$$\|F(w^\star) - F(\phi^\star)\| \leq \delta\|w^\star - \phi^\star\| \leq c_1\mu_{\max}\delta\|F(w^\star)\| \qquad (2.206)$$

since from (2.197) and (2.33) we have

$$\|w^\star - \phi^\star\| = \|c_1\rho U\nabla_{w^{\mathsf{T}}}p(w^\star)\| \leq c_1\mu_{\max}\|F(w^\star)\| \qquad (2.207)$$

We further note that

$$\|U[F(w^\star) - F(\phi^\infty)]\|^2 \leq \mu_{\max}^2\delta^2\|w^\star - \phi^\infty\|^2$$

$$\leq \mu_{\max}^2\delta^2\|\widetilde{\phi}\|^2 + 2c_1\mu_{\max}^3\delta^2\|F(w^\star)\| \cdot \|\widetilde{\phi}\|$$

$$+ c_1\mu_{\max}^4\delta^2\|F(w^\star)\|^2 \qquad (2.208)$$

where we used the fact $w^\star - \phi^\infty = \phi^\star - \phi^\infty + w^\star - \phi^\star$ and

$$\|w^\star - \phi^\infty\|^2 = \|\widetilde{\phi}\|^2 + 2\widetilde{\phi}^{\mathsf{T}}(w^\star - \phi^\star) + \|w^\star - \phi^\star\|^2$$

$$\leq \|\widetilde{\phi}\|^2 + 2\|w^\star - \phi^\star\| \cdot \|\widetilde{\phi}\| + c_1\mu_{\max}^2\|F(w^\star)\|^2$$

$$\leq \|\widetilde{\phi}\|^2 + 2c_1\mu_{\max}\|F(w^\star)\| \cdot \|\widetilde{\phi}\| + c_1\mu_{\max}^2\|F(w^\star)\|^2 \qquad (2.209)$$

Using (2.205) and (2.208) we get

$$\|\widetilde{\psi}\|^2 \leq \mathcal{X}\|\widetilde{\phi}\|^2 + 2c_1\mu_{\max}^2(1 + \mu_{\max}\delta)\delta\|F(w^\star)\| \cdot \|\widetilde{\phi}\| + c_1\mu_{\max}^4\delta^2\|F(w^\star)\|^2 \qquad (2.210)$$

where we introduced

$$\mathcal{X} \triangleq 1 - 2\mu_{\max}\nu'' + \mu_{\max}^2\delta^2 \geq 0 \qquad (2.211)$$

Note that $\mathcal{X}$ is always nonnegative by (2.158). Similarly, from (2.201) we have

$$
\begin{aligned}
\|\widetilde{w}\|^2 &\le \|\widetilde{\psi}\|^2 + c_2 \mu_{\max}^2 \rho^2 \|\nabla_{w^\mathsf{T}} p(w^\star) - \nabla_{w^\mathsf{T}} p(\psi^\infty)\|^2 \\
&\quad - 2c_2 \rho \widetilde{\psi}^\mathsf{T} U [\nabla_{w^\mathsf{T}} p(w^\star) - \nabla_{w^\mathsf{T}} p(\psi^\infty)] \\
&\le \|\widetilde{\psi}\|^2 + c_2 \mu_{\max}^2 \rho^2 \delta_p^2 \|w^\star - \psi^\infty\|^2 - 2c_2 \rho \widetilde{\psi}^\mathsf{T} U [\nabla_{w^\mathsf{T}} p(w^\star) - \nabla_{w^\mathsf{T}} p(\psi^\star)] \\
&\quad + 2c_2 t \mu_{\max} \rho \delta_p \|\widetilde{\psi}\|^2
\end{aligned}
\tag{2.212}
$$

where we rewrote

$$
\nabla_{w^\mathsf{T}} p(w^\star) - \nabla_{w^\mathsf{T}} p(w^\infty) = \nabla_{w^\mathsf{T}} p(w^\star) - \nabla_{w^\mathsf{T}} p(\psi^\star) + \nabla_{w^\mathsf{T}} p(\psi^\star) - \nabla_{w^\mathsf{T}} p(w^\infty)
\tag{2.213}
$$

and used (2.140) and (2.70). By (2.196) we know that

$$
\begin{aligned}
\|w^\star - \psi^\star\| &= \|c_2 \rho U \nabla_{w^\mathsf{T}} p(w^\star)\| \\
&\le c_2 \mu_{\max} \|\rho \nabla_{w^\mathsf{T}} p(w^\star)\| \\
&= c_2 \mu_{\max} \|F(w^\star)\|
\end{aligned}
\tag{2.214}
$$

Then, it follows that

$$
\begin{aligned}
\|w^\star - \psi^\infty\|^2 &= \|\widetilde{\psi}\|^2 + 2\widetilde{\psi}^\mathsf{T}(w^\star - \psi^\star) + \|w^\star - \psi^\star\|^2 \\
&\le \|\widetilde{\psi}\|^2 + 2\|w^\star - \psi^\star\| \cdot \|\widetilde{\psi}\| + c_2 \mu_{\max}^2 \|F(w^\star)\|^2 \\
&\le \|\widetilde{\psi}\|^2 + 2c_2 \mu_{\max} \|F(w^\star)\| \cdot \|\widetilde{\psi}\| + c_2 \mu_{\max}^2 \|F(w^\star)\|^2
\end{aligned}
\tag{2.215}
$$

Furthermore, we can use the Cauchy-Schwartz inequality and the Lipschitz-continuous assumption again to write

$$
\begin{aligned}
-2c_2 \rho \widetilde{\psi}^\mathsf{T} U [\nabla_{w^\mathsf{T}} p(w^\star) - \nabla_{w^\mathsf{T}} p(\psi^\star)] &\le 2c_2 \rho \mu_{\max} \|\nabla_{w^\mathsf{T}} p(w^\star) - \nabla_{w^\mathsf{T}} p(\psi^\star)\| \cdot \|\widetilde{\psi}\| \\
&\le 2c_2 \mu_{\max} \rho \delta_p \|w^\star - \psi^\star\| \cdot \|\widetilde{\psi}\| \\
&\le 2c_2 \mu_{\max}^2 \rho \delta_p \|F(w^\star)\| \cdot \|\widetilde{\psi}\|
\end{aligned}
\tag{2.216}
$$

where the last inequality is by (2.214). Substituting (2.215) and (2.216) into (2.212), we get

$$
\|\widetilde{w}\|^2 \le \mathcal{Y}_2 \|\widetilde{\psi}\|^2 + 2c_2 \mu_{\max}^2 (1 + \mu_{\max} \rho \delta_p) \rho \delta_p \|F(w^\star)\| \cdot \|\widetilde{\psi}\| + c_2 \mu_{\max}^4 \rho^2 \delta_p^2 \|F(w^\star)\|^2
\tag{2.217}
$$

69

where we introduced

$$\mathcal{Y}_2 \triangleq 1 + 2c_2 t \mu_{\max} \rho \delta_p + c_2 \mu_{\max}^2 \rho^2 \delta_p^2 \tag{2.218}$$

To continue, we note the following properties:

$$\mathcal{Y}_1 \mathcal{Y}_2 = 1 + 2t \mu_{\max} \rho \delta_p + \mu_{\max}^2 \rho^2 \delta_p^2 \triangleq \mathcal{Y} \tag{2.219}$$

$$c_1 \mathcal{Y}_2 = c_1, \qquad c_2 \mathcal{Y}_1 = c_2 \mathcal{Y} \tag{2.220}$$

$$c_1 \mathcal{Y}_1 = c_1 \mathcal{Y}, \qquad c_2 \mathcal{Y}_2 = c_2 \mathcal{Y} \tag{2.221}$$

$$c_2 \|\widetilde{\psi}\|^2 = c_2 \mathcal{X} \|\widetilde{w}\|^2 \iff c_2 \|\widetilde{\psi}\| = c_2 \sqrt{\mathcal{X}} \|\widetilde{w}\| \tag{2.222}$$

by recalling $c_1 \cdot c_2 = 1$ and $c_1 + c_2 = 1$ in (2.88). Combining (2.202), (2.210) and (2.217) we obtain

$$
\begin{aligned}
\|\widetilde{w}\|^2 &\leq \mathcal{Y}_1 \mathcal{Y}_2 \mathcal{X} \|\widetilde{w}\|^2 + 2c_1 \mu_{\max}^2 (1 + \mu_{\max} \delta) \delta \|F(w^\star)\| \sqrt{\mathcal{Y}_1} \cdot \|\widetilde{w}\| \\
&\quad + 2c_2 \mu_{\max}^2 (1 + \mu_{\max} \rho \delta_p) \rho \delta_p \|F(w^\star)\| \sqrt{\mathcal{Y}_2} \cdot \|\widetilde{w}\| \\
&\quad + c_1 \mu_{\max}^4 \delta^2 \|F(w^\star)\|^2 + c_2 \mu_{\max}^4 \rho^2 \delta_p^2 \|F(w^\star)\|^2 \\
&= \mathcal{Y} \mathcal{X} \|\widetilde{w}\|^2 + 2\mu_{\max}^2 \|F(w^\star)\| \sqrt{\mathcal{Y}} \mathcal{Z} \cdot \|\widetilde{w}\| + \mu_{\max}^4 (c_1 \delta^2 + c_2 \rho^2 \delta_p^2) \cdot \|F(w^\star)\|^2 \quad (2.223)
\end{aligned}
$$

where

$$\mathcal{Z} \triangleq c_1 (1 + \mu_{\max} \delta) \delta + c_2 (1 + \mu_{\max} \rho \delta_p) \rho \delta_p \tag{2.224}$$

Noting $1 - \mathcal{Y} \mathcal{X} = a_1$ as defined in (2.160), we can rewrite (2.223) as

$$a_1 \|\widetilde{w}\|^2 - 2b \|\widetilde{w}\| \leq \eta \tag{2.225}$$

where

$$b \triangleq 2\mu_{\max}^2 \|F(w^\star)\| \sqrt{\mathcal{Y}} \mathcal{Z} \geq 0 \tag{2.226}$$

$$\eta \triangleq \mu_{\max}^4 (c_1 \delta^2 + c_2 \rho^2 \delta_p^2) \cdot \|F(w^\star)\|^2 \geq 0 \tag{2.227}$$

From Appendix 2.E, we know that $a_1 > 0$ if

$$0 < \mu_{\max} < \mu_o, \quad t < \frac{\nu}{\delta + \rho \delta_p}, \quad \rho > \frac{\delta}{\delta_p} \tag{2.228}$$

70

Under these conditions we can rewrite (2.225) as

$$\left(\|\widetilde{w}\| - \frac{b}{a_1} \right)^2 \leq \frac{\eta}{a_1} + \frac{b^2}{a_1^2}$$

$$\iff \quad \frac{b}{a_1} - \sqrt{\frac{\eta}{a_1} + \frac{b^2}{a_1^2}} \leq \|\widetilde{w}\| \leq \frac{b}{a_1} + \sqrt{\frac{\eta}{a_1} + \frac{b^2}{a_1^2}} \tag{2.229}$$

Noting that

$$\frac{b}{a_1} - \sqrt{\frac{\eta}{a_1} + \frac{b^2}{a_1^2}} = \frac{b}{a_1} - \frac{\sqrt{b^2 + a_1\eta}}{a_1} \leq 0 \tag{2.230}$$

we get

$$0 \leq \|\widetilde{w}\| \leq \frac{b}{a_1} + \sqrt{\frac{\eta}{a_1} + \frac{b^2}{a_1^2}} \tag{2.231}$$

Our goal is to study the bias performance for sufficiently small step-sizes, which can be examined from

$$\lim_{\mu_{\max} \to 0} \sup \frac{\|\widetilde{w}\|}{\mu_{\max}} \leq \lim_{\mu_{\max} \to 0} \frac{b}{a_1 \mu_{\max}} + \lim_{\mu_{\max} \to 0} \sqrt{\frac{\eta}{a_1 \mu_{\max}^2} + \frac{b^2}{a_1^2 \mu_{\max}^2}} \tag{2.232}$$

From (2.226) and (2.160) we have

$$\lim_{\mu_{\max} \to 0} \frac{b}{a_1 \mu_{\max}} = \lim_{\mu_{\max} \to 0} \frac{2\|F(w^\star)\|\sqrt{\mathcal{Y}}\mathcal{Z}}{2\nu' - \mu_{\max}(\delta^2 + \rho^2\delta_p^2 - 4t\nu''\rho\delta_p) + O(\mu_{\max}^2)}$$

$$= d_1(c_1\delta + c_2\rho\delta_p) \tag{2.233}$$

where we used the fact $\lim_{\mu_{\max} \to 0} \mathcal{Y} = 1$ and introduced

$$d_1 \triangleq \|F(w^\star)\|/\nu' \tag{2.234}$$

From the definition (2.227) we get

$$\lim_{\mu_{\max} \to 0} \frac{\eta}{a_1 \mu_{\max}^2} = \lim_{\mu_{\max} \to 0} \frac{\mu_{\max}(c_1\delta^2 + c_2\rho^2\delta_p^2) \cdot \|F(w^\star)\|^2}{2\nu' - \mu_{\max}(\delta^2 + \rho^2\delta_p^2 - 4t\nu''\rho\delta_p) + O(\mu_{\max}^2)}$$

$$= 0 \tag{2.235}$$

Consequently, we have

$$\lim_{\mu_{\max} \to 0} \sup \frac{\|\widetilde{w}\|}{\mu_{\max}} \leq 2d_1(c_1\delta + c_2\rho\delta_p) < 2d_1\rho\delta_p \tag{2.236}$$

where we used the condition $\rho > \delta/\delta_p$ and the fact $c_1 + c_2 = 1$.

71

CHAPTER 3

Learning by Networked Agents under Partial Information

Using the formulation developed in Chapter 2, we extend earlier contributions on adaptive networks, which generally assume that the agents work together for a common global objective or when they observe data that is generated by a common target model or parameter vector. In this chapter, this condition is relaxed since in many scenarios of interest, agents may only have access to partial information about an unknown model or target vector. Each agent may be sensing only a subset of the entries of a global target vector, and the number of these entries can be different across the agents. If each of the agents were to solve an inference task independently of the other agents, then they would not benefit from neighboring agents that may be sensing similar entries. This chapter develops cooperative distributed techniques that enable agents to cooperate even when their interactions are limited to exchanging estimates of select few entries. In the proposed strategies, agents are only required to share estimates of their common entries, which results in a significant reduction in communication overhead. Simulations show that the proposed approach improves both the performance of individual agents and the entire network through cooperation.

3.1 Introduction

In most prior works of adaptive networks, agents are assumed to have a common minimizer and cooperate to estimate it by using effective distributed strategies such as the consensus strategy (e.g., [1–4, 6–9]) or the diffusion strategy (e.g., [12, 15–17]). When the agents do not share a minimizer, it was shown in [15, 16, 81, 82] that the network converges to a Pareto

optimal solution. When it is desired instead that agents, or clusters of agents within the network, should converge to their respective models, rather than the Pareto solution, then multi-task diffusion strategies become useful and can be used to attain this objective [18, 20].

In these earlier contributions, it is generally assumed that the target vector for each agent has the same size and, moreover, that the agents sense data that is affected by all entries of their target vectors. In this chapter, we relax these conditions and consider a broader scenario where individual agents sense only a subset of the entries of the global target vector, and where different agents can sense subsets of different sizes. This formulation allows us to model the important situation in which agents may only have access to partial information about an unknown model or target vector. If each of the agents were to solve an inference task independently of the other agents, then they would not benefit from cooperation with neighboring agents that may be sensing common entries. This chapter develops cooperative distributed techniques that enable agents to cooperate even when their interactions are limited to exchanging estimates of select few entries. To attain this objective, we allow for some entries of the global target vector to be observable by more than one agent so that cooperation across the network is justified.

Our approach will be based on formulating a constrained optimization problem for recovering partial entries of the global target vector. However, rather than solve this problem directly, we will introduce a penalized version using a quadratic term to penalize the violation of the constraints. We will then develop a diffusion learning solution to solve the optimization problem in a distributed manner by relying on two incremental steps. In the adaptation step, agents descend along the negative direction of the gradients of their costs. And in the penalty step, they descend along the negative direction of the gradients of their penalties. When the exact gradient information is unavailable, the observed data is used to compute instantaneous approximations for the gradient vectors. In the penalty step, agents will only share the common entries of their estimates with neighbors to reduce the communication costs. The order of executing the two incremental steps results in the Adapt-then-Penalize (ATP) or Penalize-then-Adapt (PTA) diffusion strategies.

We remark that this chapter considers a more general scenario than the partial diffusion

formulation proposed in [90]. There, all agents sense data driven by the *same* target vector, cooperate to estimate this target vector, and exchange only part of their entries. In our formulation, each agent will be sensing data driven by different local target vectors and these can be of different sizes. In this way, agents are able to cooperate even if their target vectors are only partially common. We then show that sufficiently small step-sizes ensure mean and mean-square stability. We illustrate the results by means of computer simulations.

3.2   Problem Formulation

At each time instant $i \geq 0$, each agent k is assumed to have access to a scalar measurement $\boldsymbol{d}_k(i) \in \mathbb{R}$ and a regression vector $\boldsymbol{u}_{k,i} \in \mathbb{R}^{1 \times M_k}$ with covariance matrix $R_{u,k} = \mathbb{E}\boldsymbol{u}_{k,i}^\mathsf{T}\boldsymbol{u}_{k,i} > 0$. The regressors $\{\boldsymbol{u}_{k,i}\}$ are assumed to have zero mean and to be temporally white and spatially independent. The data $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$ are assumed to be related via the linear regression model:

$$\boldsymbol{d}_k(i) = \boldsymbol{u}_{k,i} w_k^o + \boldsymbol{v}_k(i) \tag{3.1}$$

where $w_k^o \in \mathbb{R}^{M_k \times 1}$ is the target vector to be estimated by agent k. The variable $\boldsymbol{v}_k(i) \in \mathbb{R}$ is a zero-mean white-noise process with variance $\mathbb{E}\boldsymbol{v}_k^2(i) = \sigma_{v,k}^2$ and assumed to be spatially independent. We further assume that the random processes $\boldsymbol{u}_{k,i}$ and $\boldsymbol{v}_\ell(j)$ are spatially and temporally independent for any k, ℓ, i, and j. We assume that each entry of w_k^o is determined by a grand target vector $w^o \in \mathbb{R}^{M \times 1}$, i.e., the relation between w_k^o and w^o can be described by

$$w_k^o = D_k w^o \tag{3.2}$$

where D_k is a matrix of size $M_k \times M$ and defined as

$$D_k(s, m) = \begin{cases} 1, & \text{if } w_k^o(s) \leftarrow w^o(m) \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

The notation $x \leftarrow y$ denotes that the value of y is assigned to x. We are therefore considering a distributed inference problem where each agent has partial information about a grand target vector, i.e., the data at each agent is influenced by only some entries of w^o. Observe

74

that the size M_k of the vector w_k^o is allowed to change with the node index, k, so that some agents may be influenced by more entries than other agents.

Now, given a network topology, two neighboring agents $\{k, \ell\}$ may share one or more common target entries, e.g., there can exist some index $m \in \{1, \ldots, M\}$ such that

$$w_k^o(s) \leftarrow w^o(m), \quad w_\ell^o(s') \leftarrow w^o(m), \quad \ell \in \mathcal{N}_k \setminus \{k\} \tag{3.4}$$

where $\mathcal{N}_k$ is the neighborhood of agent k. We are therefore motivated to consider the following constrained optimization problem for each agent k:

$$\min_{w_k} \quad J_k(w_k) \triangleq \frac{1}{2} \mathbb{E} |\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i} w_k|^2$$

$$\text{subject to} \quad w_k(s) = w_\ell(s'), \ \ell \in \mathcal{N}_k \setminus \{k\},$$

$$s \in \{1, \ldots, M_k\}, \ s' \in \{1, \ldots, M_\ell\} \tag{3.5}$$

The indices s and s' in (3.5) refer only to the common entries in w_k^o and w_ℓ^o. We provide an example in Fig. 3.1 to illustrate the setting defined in (3.5). For example, agents #1 and #2 share the common target entry $w^o(2)$; it is the leading element in the target vector for agent #2 and the trailing element in the target vector for agent #1. Observe that agents can share target entries even if they are not neighbors, as is the case with agents #1 and #3; they both share entry $w^o(1)$. However, the constraints for the common target entries can only exist between neighboring agents. For convenience and for later use, we collect the constraints for each agent k into a constraint set $\mathbb{S}_k$:

$$\mathbb{S}_k \triangleq \left\{ (\ell, s, s') \ \middle| \ \begin{array}{l} w_k(s) = w_\ell(s'), \ell \in \mathcal{N}_k \setminus \{k\} \\ s \in \{1, \ldots, M_k\}, \ s' \in \{1, \ldots, M_\ell\} \end{array} \right\} \tag{3.6}$$

3.3 Penalty-Based Learning

One way to solve problem (3.5) is to reformulate it using penalty functions. Specifically, instead of solving (3.5), we consider the penalized version:

$$\min_{w_k} \quad J_k^p(w_k) \triangleq \frac{1}{2} \mathbb{E} |\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i} w_k|^2 + p_k(w^k) \tag{3.7}$$
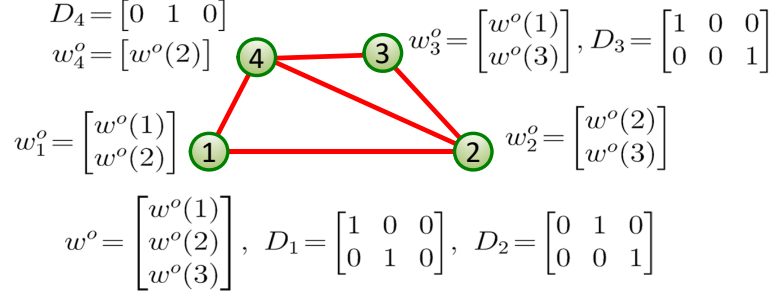
Figure 3.1: An example to illustrate distributed inference under partial information exchange.

where the notation $w^k \triangleq \mathrm{col}\{w_\ell; \ell \in \mathcal{N}_k\}$ aggregates all unknowns in the neighborhood $\mathcal{N}_k$ and $p_k(w^k)$ is a quadratic penalty function used to penalize agent k when any constraint $w_k(s) = w_\ell(s')$ is violated, i.e.,

$$p_k(w^k) \triangleq \sum_{(\ell,s,s') \in \mathbb{S}_k} \rho_k(\ell, s, s') \cdot [w_k(s) - w_\ell(s')]^2 \tag{3.8}$$

where $\rho_k(\ell, s, s')$ is a positive penalty parameter used to control the punishment level of violating the constraint $w_k(s) = w_\ell(s')$. Other choices for the penalty function are possible. It is sufficient for our purposes in this article to illustrate the main construction and results using (3.8).

3.3.1   Entry-Wise Diffusion Implementation

Following the approach from [54], the optimization problem (3.7) can be solved in two incremental steps: we first adapt with respect to $J_k(w_k)$ and then adapt with respect to $p_k(w^k)$. For this purpose, we start by noting that the gradient vector of $J_k(w_k)$ with respect to w_k is given by

$$\nabla_{w_k^\mathsf{T}} J_k(w_k) = R_{u,k} w_k - r_{du,k} \tag{3.9}$$

where $r_{du,k} = \mathbb{E}\boldsymbol{d}_k(i)\boldsymbol{u}_{k,i}^\mathsf{T}$. When the gradient of $J_k(w_k)$ is unavailable, we can approximate it by using the instantaneous approximations $r_{du,k} \approx \boldsymbol{d}_k(i)\boldsymbol{u}_{k,i}^\mathsf{T}$ and $R_{u,k} \approx \boldsymbol{u}_{k,i}^\mathsf{T}\boldsymbol{u}_{k,i}$:

$$\widehat{\nabla_{w_k^\mathsf{T}} J_k}(w_k) = \boldsymbol{u}_{k,i}^\mathsf{T}[\boldsymbol{u}_{k,i} w_k - \boldsymbol{d}_k(i)] \tag{3.10}$$

76

By doing so, we arrive at the adapt-then-penalize (ATP) diffusion strategy:

$$\text{(ATP)} \begin{cases} \boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} + \mu_k \boldsymbol{u}_{k,i}^{\mathsf{T}}[\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{w}_{k,i-1}] & (3.11) \\[2ex] \boldsymbol{w}_{k,i} = \boldsymbol{\psi}_{k,i} - \mu_k \nabla_{w_k^{\mathsf{T}}} p_k(\boldsymbol{\psi}_i^k) & (3.12) \end{cases}$$

where $\boldsymbol{\psi}_i^k \triangleq \text{col}\{\boldsymbol{\psi}_{\ell,i}; \ell \in \mathcal{N}_k\}$. By differentiating the penalty function $p_k(\boldsymbol{\psi}_i^k)$ with respect to w_k, it can be verified that step (3.12) can be rewritten as

$$\boldsymbol{w}_{k,i} = \sum_{\ell \in \mathcal{N}_k} A_{\ell k}^{\mathsf{T}} \boldsymbol{\psi}_{\ell,i} \tag{3.13}$$

where $A_{\ell k}$ is the $M_\ell \times M_k$ matrix with entries defined by

$$A_{\ell k}(s', s) = \begin{cases} 2\mu_k \rho_k(\ell, s, s'), & \text{if } (\ell, s, s') \in \mathbb{S}_k \\[2ex] 0, & \text{otherwise} \end{cases} \tag{3.14}$$

for $k \neq \ell$, and

$$A_{kk} = \text{diag}\left\{ 1 - \sum_{(\ell,1,s') \in \mathbb{S}_k} 2\mu_k \rho_k(\ell, 1, s'), \cdots, 1 - \sum_{(\ell,M_k,s') \in \mathbb{S}_k} 2\mu_k \rho_k(\ell, M_k, s') \right\} \tag{3.15}$$

From (3.14) and (3.15), we get that for any s:

$$A_{kk}(s, s) + \sum_{\substack{\forall \ell, \forall s' \\ (\ell,s,s') \in \mathbb{S}_k}} A_{\ell k}(s', s) = 1 \tag{3.16}$$

which can be collected for all s into a compact expression:

$$\begin{bmatrix} A_{1k}^{\mathsf{T}} & \cdots & A_{Nk}^{\mathsf{T}} \end{bmatrix} \mathbb{1}_{M_k} = \mathbb{1}_{M_k} \tag{3.17}$$

where $\mathbb{1}_{M_k}$ denotes the vector of size $M_k \times 1$ with all one entries. Observe that the matrix $A_{\ell k}$ defines the combination weights between agent k and the common entries with agent ℓ from its neighborhood; we assume the step-sizes $\{\mu_k\}$ are sufficiently small such that the diagonal entries in $\{A_{kk}\}$ are nonnegative so that $A = [A_{\ell k}]$ is a left-stochastic matrix. It turns out that the particular forms (3.14) and (3.15) are not critical. It is sufficient to select an arbitrary left-stochastic matrix A as long as the zero-structure of its block components $\{A_{\ell k}\}$ and property (3.17) are satisfied. We can also switch the order of the incremental

steps in (3.11)–(3.12) and arrive at the penalize-then-adapt (PTA) diffusion strategy:

$$\text{(PTA)} \begin{cases} \boldsymbol{\psi}_{k,i} = \sum_{\ell \in \mathcal{N}_k} A_{\ell k}^\mathsf{T} \boldsymbol{w}_{\ell,i-1} & (3.18) \\[2mm] \boldsymbol{w}_{k,i} = \boldsymbol{\psi}_{k,i} + \mu_k \boldsymbol{u}_{k,i}^\mathsf{T}[\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{\psi}_{k,i}] & (3.19) \end{cases}$$

We remark that in the penalty steps (3.13) and (3.18), agents are only required to exchange a subset of the entries of their iterates with their neighboring agents, namely, those entries that define the constraints. This property reduces communication overhead, compared with traditional diffusion ATC and CTA algorithms in [16]. The ATP and PTA formulations (3.11)–(3.13) and (3.18)–(3.19) define a useful class of distributed strategies that include other important cases as special instances. For example, if for each agent k we set $D_k = I_M$ in (3.2) and extend the constraint set $\mathbb{S}_k$ to cover all possible entries s and s', we will arrive at the group diffusion LMS algorithm proposed in [91] for agents to assign different combination weights to different entries. If we further set $A_{\ell k} = a_{\ell k} I$, we will get the traditional diffusion LMS algorithms from [16].

3.4 Performance Analysis

We consider the ATP diffusion strategy (3.11)–(3.13). We can rewrite the adaptation step in (3.11) as

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} + \mu_k \boldsymbol{u}_{k,i}^\mathsf{T} \boldsymbol{u}_{k,i}(w_k^o - \boldsymbol{w}_{k,i-1}) + \mu_k \boldsymbol{u}_{k,i}^\mathsf{T} \boldsymbol{v}_k(i) \qquad (3.20)$$

where we used the linear model (3.1). Collecting the iterates $\{\boldsymbol{\psi}_{k,i}\}$ and $\{\boldsymbol{w}_{k,i}\}$ from the across the network at time i into the aggregate vectors $\boldsymbol{\psi}_i = \mathrm{col}\{\boldsymbol{\psi}_{1,i}, \ldots, \boldsymbol{\psi}_{N,i}\}$ and $\boldsymbol{w}_i = \mathrm{col}\{\boldsymbol{w}_{1,i}, \ldots, \boldsymbol{w}_{N,i}\}$, we find that these network vectors evolve according to the dynamics:

$$\boldsymbol{\psi}_i = \boldsymbol{w}_{i-1} + \mathcal{M}\mathcal{R}_i(w_\star^o - \boldsymbol{w}_{i-1}) + \mathcal{M}\boldsymbol{s}_i \qquad (3.21)$$

$$\boldsymbol{w}_i = \mathcal{A}^\mathsf{T} \boldsymbol{\psi}_i \qquad (3.22)$$

where

$$\mathcal{M} \triangleq \operatorname{diag}\{\mu_1 I_{M_1}, \ldots, \mu_N I_{M_N}\} \tag{3.23}$$

$$\boldsymbol{\mathcal{R}}_i \triangleq \operatorname{diag}\{\boldsymbol{u}_{1,i}^\mathsf{T}\boldsymbol{u}_{1,i}, \ldots, \boldsymbol{u}_{N,i}^\mathsf{T}\boldsymbol{u}_{N,i}\} \tag{3.24}$$

$$w_\star^o \triangleq \operatorname{col}\{w_1^o, \ldots, w_N^o\} = \mathcal{D}w^o \tag{3.25}$$

$$\mathcal{D} \triangleq \begin{bmatrix} D_1 \\ \vdots \\ D_N \end{bmatrix}, \quad \boldsymbol{s}_i \triangleq \begin{bmatrix} \boldsymbol{u}_{1,i}^\mathsf{T}\boldsymbol{v}_1(i) \\ \vdots \\ \boldsymbol{u}_{N,i}^\mathsf{T}\boldsymbol{v}_N(i) \end{bmatrix}, \quad \mathcal{A} \triangleq \begin{bmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{NN} \end{bmatrix} \tag{3.26}$$

Similarly, the network recursion for PTA diffusion is given by

$$\boldsymbol{\psi}_i = \mathcal{A}^\mathsf{T}\boldsymbol{w}_{i-1} \tag{3.27}$$

$$\boldsymbol{w}_i = \boldsymbol{\psi}_i + \mathcal{M}\boldsymbol{\mathcal{R}}_i(w_\star^o - \boldsymbol{\psi}_i) + \mathcal{M}\boldsymbol{s}_i \tag{3.28}$$

We can capture the dynamics of both algorithms in a single unified model by introducing intermediate iterates $\boldsymbol{\phi}_{k,i}$ and matrices $\{\mathcal{A}_1, \mathcal{A}_2\}$ and then writing:

$$\boldsymbol{\phi}_i = \mathcal{A}_1^\mathsf{T}\boldsymbol{w}_{i-1} \tag{3.29}$$

$$\boldsymbol{\psi}_i = \boldsymbol{\phi}_i + \mathcal{M}\boldsymbol{\mathcal{R}}_i(w_\star^o - \boldsymbol{\phi}_i) + \mathcal{M}\boldsymbol{s}_i \tag{3.30}$$

$$\boldsymbol{w}_i = \mathcal{A}_2^\mathsf{T}\boldsymbol{\psi}_i \tag{3.31}$$

The PTA case corresponds to the choice $\mathcal{A}_1 = \mathcal{A}$ and $\mathcal{A}_2 = I$ while the ATP case corresponds to $\mathcal{A}_1 = I$ and $\mathcal{A}_2 = \mathcal{A}$. To continue, we note the following useful properties:

$$\mathcal{A}_2\mathcal{A}_1 = \mathcal{A}, \quad \mathcal{A}^\mathsf{T}\mathbb{1} = \mathbb{1} \tag{3.32}$$

Furthermore, from (3.16) and (3.4) we have

$$\begin{aligned} w_k^o(s) &= \left[A_{kk}(s,s) + \sum_{\substack{\forall \ell, \forall s' \\ (\ell,s,s') \in \mathbb{S}_k}} A_{\ell k}(s',s) \right] w_k^o(s) \\ &= A_{kk}(s,s)w_k^o(s) + \sum_{\substack{\forall \ell, \forall s' \\ (\ell,s,s') \in \mathbb{S}_k}} A_{\ell k}(s',s)w_\ell^o(s') \end{aligned} \tag{3.33}$$

79

which means

$$w_k^o = \sum_{\ell \in \mathcal{N}_k} A_{\ell k}^{\mathsf{T}} w_\ell^o \tag{3.34}$$

and, therefore,

$$w_\star^o = \mathcal{A}^{\mathsf{T}} w_\star^o \tag{3.35}$$

Then, it is ready to check

$$w_\star^o = \mathcal{A}_1^{\mathsf{T}} w_\star^o, \quad w_\star^o = \mathcal{A}_2^{\mathsf{T}} w_\star^o \tag{3.36}$$

since $\mathcal{A}_1$ and $\mathcal{A}_1$ can only be I or $\mathcal{A}$. From (3.25) and (3.29) we can write

$$w_\star^o - \boldsymbol{\phi}_i = \mathcal{A}_1^{\mathsf{T}} (w_\star^o - \boldsymbol{w}_{i-1}) \tag{3.37}$$

and, similarly,

$$w_\star^o - \boldsymbol{w}_i = \mathcal{A}_2^{\mathsf{T}} (w_\star^o - \boldsymbol{\psi}_i) \tag{3.38}$$

Therefore, we arrive at the network error recursions:

$$\widetilde{\boldsymbol{\phi}}_i = \mathcal{A}_1^{\mathsf{T}} \widetilde{\boldsymbol{w}}_{i-1} \tag{3.39}$$

$$\widetilde{\boldsymbol{\psi}}_i = \widetilde{\boldsymbol{\phi}}_i - \mathcal{M}\mathcal{R}_i \widetilde{\boldsymbol{\phi}}_i - \mathcal{M}\boldsymbol{s}_i \tag{3.40}$$

$$\widetilde{\boldsymbol{w}}_i = \mathcal{A}_2^{\mathsf{T}} \widetilde{\boldsymbol{\psi}}_i \tag{3.41}$$

where $\widetilde{\boldsymbol{\phi}}_i \triangleq w_\star^o - \boldsymbol{\phi}_i$ and similarly for the other error vectors. Combining (3.39)–(3.41) we conclude that the error vector $\widetilde{\boldsymbol{w}}_i$ evolves according to the recursion:

$$\widetilde{\boldsymbol{w}}_i = \boldsymbol{\mathcal{B}}_i \widetilde{\boldsymbol{w}}_{i-1} - \mathcal{G}\boldsymbol{s}_i \tag{3.42}$$

where

$$\boldsymbol{\mathcal{B}}_i \triangleq \mathcal{A}_2^{\mathsf{T}} (I - \mathcal{M}\mathcal{R}_i) \mathcal{A}_1^{\mathsf{T}} \tag{3.43}$$

$$\mathcal{G} \triangleq \mathcal{A}_2^{\mathsf{T}} \mathcal{M} \tag{3.44}$$

It was shown in [17] that the traditional ATC and CTA diffusion algorithms have a network error recursion of the same general form as (3.42), except that now we have one critical

difference. Expression (3.42) is more general and allows agents to have *different* sizes for their target vectors $\{w_k^o\}$. Furthermore, the matrices $\mathcal{A}_1$ and $\mathcal{A}_2$ now reflect refined connections: two connected agents only share a subset of their entries, which can be a single entry in the extreme case. Therefore, cooperation between agents is limited to entry-wise exchanges, as opposed to full vector exchanges in traditional implementations. Following similar arguments to those in [17], we can derive conditions on the step-size parameters to ensure mean-square convergence and stability. For any nonnegative symmetric matrix Σ, we let $\sigma = \text{vec}(\Sigma)$ and use the compact notation $\|x\|_\sigma^2$ to refer to the squared weighted quantity $x^\mathsf{T} \Sigma x$.

Theorem 3.1. *(**Mean-square-error stability**) For sufficiently small step step-sizes, i.e., for $\mu_k < \mu_o$ for some small enough μ_o, it holds that*

$$\rho(\mathcal{F}) < 1 \tag{3.45}$$

where $\rho(\cdot)$ is the spectral radius, $\otimes$ is the Kronecker product operator, and

$$\mathcal{F} \triangleq \mathbb{E}(\boldsymbol{\mathcal{B}}_i^T \otimes \boldsymbol{\mathcal{B}}_i^T) \tag{3.46}$$

Condition (3.45) ensures that the estimates are asymptotically unbiased, i.e., $\mathbb{E}\widetilde{\boldsymbol{w}}_i \to 0$, and moreover, the weighted error variance satisfies the relation:

$$\mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|_\sigma^2 = \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}\|_{\mathcal{F}\sigma}^2 + \left[vec\left(\mathcal{Y}^T\right)\right]^\mathsf{T} \sigma \tag{3.47}$$

where $\mathcal{Y} \triangleq \mathcal{G}\mathcal{S}\mathcal{G}^T$, and

$$\mathcal{S} \triangleq diag\{\sigma_{v,1}^2 R_{u,1}, ..., \sigma_{v,1}^2 R_{u,N}\} \tag{3.48}$$

$\square$

Proof. See Appendix 3.A. $\square$

The relation (3.47) can also be used to evaluate the steady-state mean-square-error performance. If we assume that condition (3.45) holds such that the network is stable, taking the limit of (3.47) as $i \to \infty$ we get

$$\lim_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|_\sigma^2 = \lim_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}\|_{\mathcal{F}\sigma}^2 + \left[\text{vec}\left(\mathcal{Y}^\mathsf{T}\right)\right]^\mathsf{T} \sigma \tag{3.49}$$

Therefore, we obtain

$$\lim_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|^2_{(I-\mathcal{F})\sigma} = \left[\text{vec}\left(\mathcal{Y}^{\mathsf{T}}\right)\right]^{\mathsf{T}} \sigma \tag{3.50}$$

Note that the matrix Σ with $\sigma = \text{vec}(\Sigma)$ can be any Hermitian nonnegative-definite matrix. One useful metric for mean-square-error performance is the mean-square deviation (MSD) measure, which can be used to assess how far the estimate $\boldsymbol{w}_{k,i}$ of each agent k is from w_k^o. The MSD of each agent k is defined as follows:

$$\text{MSD}_k \triangleq \lim_{i\to\infty} \mathbb{E}\|w_k^o - \boldsymbol{w}_{k,i}\|^2 \tag{3.51}$$

where $\|\cdot\|$ denotes the Euclidean norm for vectors. The network MSD is defined as the average MSD across the network, i.e.,

$$\text{MSD} \triangleq \frac{1}{N} \sum_{k=1}^{N} \text{MSD}_k \tag{3.52}$$

Now, we note that condition (3.45) implies that the matrix $I - \mathcal{F}$ is invertible. Therefore, by selecting

$$\sigma = (I - \mathcal{F})^{-1} \text{vec}\left[(e_k e_k^{\mathsf{T}}) \otimes I_{M_k}\right] \tag{3.53}$$

where e_k denotes the kth column of the identity matrix I_N, we arrive at the following expression for the MSD of agent k:

$$\text{MSD}_k = \left[\text{vec}(\mathcal{Y}^{\mathsf{T}})\right]^{\mathsf{T}} (I - \mathcal{F})^{-1} \text{vec}\left[(e_k e_k^{\mathsf{T}}) \otimes I_{M_k}\right] \tag{3.54}$$

Similarly, by selecting

$$\sigma = (I - \mathcal{F})^{-1} \text{vec}\left[(e_k e_k^{\mathsf{T}}) \otimes I_{M_k}\right] \tag{3.55}$$

we obtain the network MSD as

$$\text{MSD} = \frac{1}{N} \left[\text{vec}(\mathcal{Y}^{\mathsf{T}})\right]^{\mathsf{T}} (I - \mathcal{F})^{-1} \text{vec}(I_L) \tag{3.56}$$

where

$$L = \sum_{k=1}^{N} M_k \tag{3.57}$$

82

3.5   Simulation Results

We consider a network with $N = 10$ agents. Each agent k is estimating a target vector w_k^o, which is a subvector of the grand target vector w^o of size $M = 10$. For each agent k, we assume that $R_{u,k}$ is diagonal and its diagonal entries are determined by a grand diagonal covariance matrix R_u, i.e.,

$$R_{u,k} = D_k R_u D_k^\mathsf{T} \tag{3.58}$$

Figure 3.2 shows the entries of the grand target vector w^o, the diagonal entries of the grand covariance matrix R_u, and the noise variance $\{\sigma_{v,k}^2\}$ at the agents. The network topology and the relations between $\{w_k^o\}$ and w^o are shown in Fig. 3.3. We set the step-size to $\mu_k = \mu = 0.005$ and the penalty parameter to $\rho_k(\ell, s, s') = \rho = 30$.

In Fig. 3.4, we simulate the learning curves of instantaneous network MSD, which is defined as

$$\mathrm{MSD}_i \triangleq \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}\|w_k^o - \boldsymbol{w}_{k,i}\|^2 \tag{3.59}$$

We observe that both diffusion ATP and PTA algorithms exhibit better steady-state MSD performance than the non-cooperative case without imposing constraints and penalties. To examine the individual performance, we compare the steady-state individual MSD for each agent in Fig. 3.5. It is seen that all agents benefit from exchange of information with neighbors. The difference between diffusion ATP and PTA algorithms is negligible in the figures.

3.6   Concluding Remarks

In this chapter, we consider a broader scenario where individual agents may only have access to partial information about the global target vector, i.e., each agent may be sensing only a subset of the entries of the global target. We develop cooperative distributed techniques where agents are only required to share estimates of their common entries and still can benefit from neighboring agents. It is observed in the simulations that all agents benefit
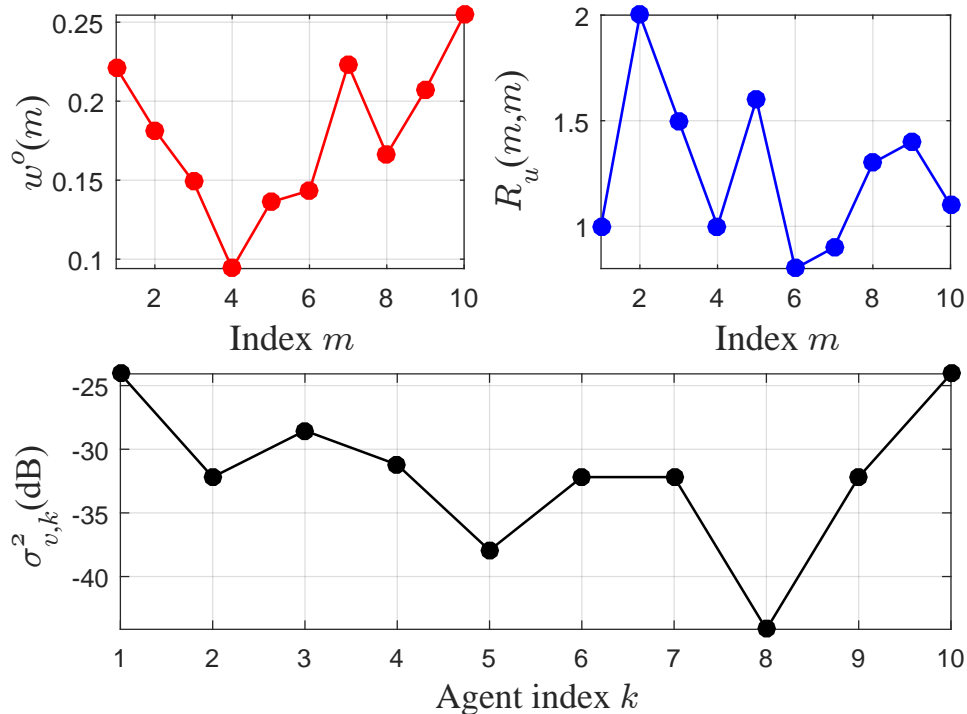
Figure 3.2: Entries of w^o, R_u, and noise variance $\{\sigma_{v,k}^2\}$ used in the simulations.

w_1^o	w_2^o	w_3^o	w_4^o	w_5^o	w_6^o	w_7^o	w_8^o	w_9^o	w_{10}^o
$\begin{bmatrix} w^o(1) \\ w^o(2) \\ w^o(7) \end{bmatrix}$	$\begin{bmatrix} w^o(2) \\ w^o(3) \end{bmatrix}$	$\begin{bmatrix} w^o(3) \\ w^o(4) \\ w^o(8) \end{bmatrix}$	$\begin{bmatrix} w^o(4) \\ w^o(5) \end{bmatrix}$	$\begin{bmatrix} w^o(5) \\ w^o(6) \\ w^o(9) \end{bmatrix}$	$\begin{bmatrix} w^o(4) \\ w^o(6) \\ w^o(7) \end{bmatrix}$	$\begin{bmatrix} w^o(3) \\ w^o(7) \\ w^o(8) \end{bmatrix}$	$\begin{bmatrix} w^o(8) \\ w^o(9) \end{bmatrix}$	$\begin{bmatrix} w^o(5) \\ w^o(9) \\ w^o(10) \end{bmatrix}$	$\begin{bmatrix} w^o(5) \\ w^o(10) \end{bmatrix}$
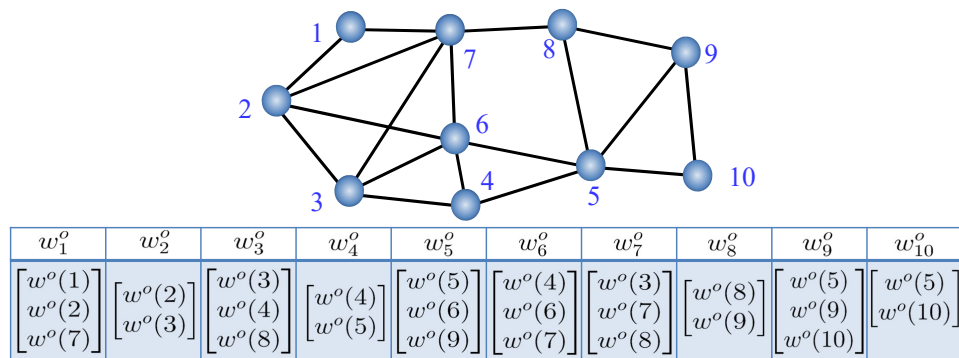
Figure 3.3: Network topology and target vectors $\{w_k^o\}$.

from exchange of information with neighbors. We remark that the traditional diffusion LMS algorithms can be obtained from our entry-wise diffusion strategies in the case that each agent has access to the full model and replace its combination matrices to some nonnegative scalars for the convex combination of neighbors' estimates.
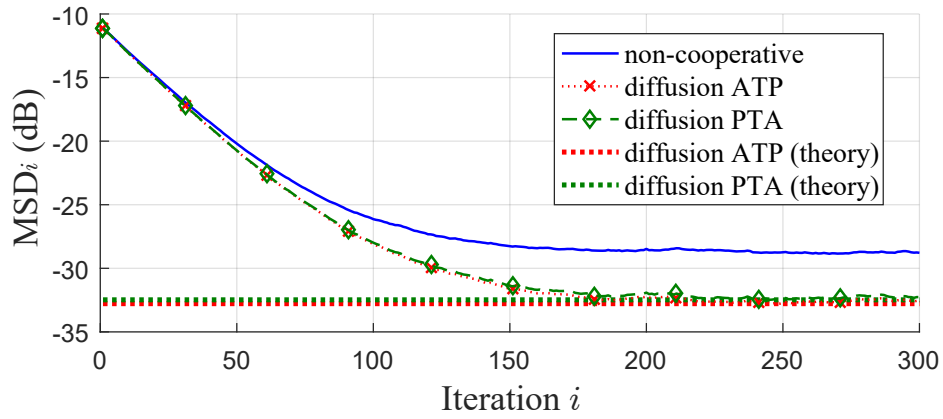
Figure 3.4: Learning curves for the network mean-square-error deviation (MSD).
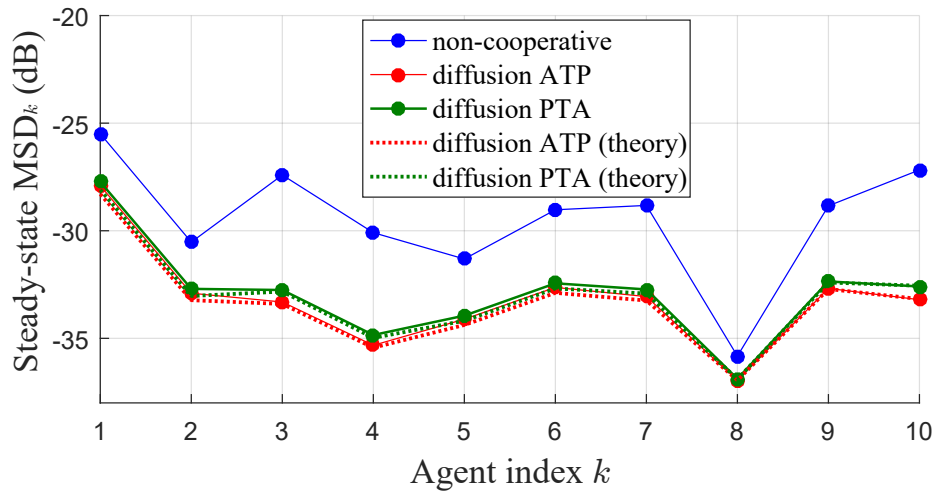
Figure 3.5: Steady-state MSD for individual agents.

APPENDICES

3.A   Proof of Theorem 3.1

Taking expectation of both sides of (3.42), we find that

$$\mathbb{E}\widetilde{\boldsymbol{w}}_i = \mathcal{B} \cdot \widetilde{\boldsymbol{w}}_{i-1} - \mathcal{G}\boldsymbol{s}_i \tag{3.60}$$

where

$$\mathcal{B} \triangleq \mathbb{E}\boldsymbol{\mathcal{B}}_i = \mathcal{A}_2^{\mathsf{T}}(I - \mathcal{M}\mathcal{R})\mathcal{A}_1^{\mathsf{T}} \tag{3.61}$$

and

$$\mathcal{R} \triangleq \mathbb{E}\boldsymbol{\mathcal{R}}_i = \mathrm{diag}\{R_{u,1}, ..., R_{u_N}\} \tag{3.62}$$

The necessary and sufficient condition for $\mathbb{E}\widetilde{\boldsymbol{w}}_i \to 0$ as $i \to \infty$ is to select the step-sizes μ_k such that the spectral radius of $\mathcal{B}$ satisfies

$$\rho(\mathcal{B}) < 1 \tag{3.63}$$

Therefore, we can conclude that the network mean stability is ensured for sufficiently small step-sizes $\{\mu_k\}$ satisfying

$$\mu_k < \frac{2}{\lambda_{\max}(R_u, k)} \tag{3.64}$$

For the mean-square-error stability, from (3.42) we get the following expression:

$$\begin{aligned}
\mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|_{\Sigma}^2 &= \mathbb{E}\|\boldsymbol{\mathcal{B}}_i\widetilde{\boldsymbol{w}}_{i-1} - \mathcal{G}\boldsymbol{s}_i\|_{\Sigma}^2 \\
&= \mathbb{E}\left(\widetilde{\boldsymbol{w}}_{i-1}^{\mathsf{T}}\boldsymbol{\mathcal{B}}_i^{\mathsf{T}}\Sigma\boldsymbol{\mathcal{B}}_i\widetilde{\boldsymbol{w}}_{i-1}\right) + \mathbb{E}\left(\boldsymbol{s}_i^{\mathsf{T}}\mathcal{G}\Sigma\mathcal{G}\boldsymbol{s}_i\right) \\
&\quad - \mathbb{E}\left(\widetilde{\boldsymbol{w}}_{i-1}^{\mathsf{T}}\boldsymbol{\mathcal{B}}_i^{\mathsf{T}}\Sigma\mathcal{G}\boldsymbol{s}_i\right) - \mathbb{E}\left(\boldsymbol{s}_i^{\mathsf{T}}\mathcal{G}\Sigma\boldsymbol{\mathcal{B}}_i\widetilde{\boldsymbol{w}}_{i-1}\right)
\end{aligned} \tag{3.65}$$

We note that the last two terms on the right-hand side are zero because the $\{\boldsymbol{u}_{k,i}, \boldsymbol{v}_k(i)\}$ are independent of each other and the $\{\boldsymbol{v}_k(i)\}$ are zero mean. The first term in (3.65) is given by

$$\begin{aligned}
\mathbb{E}\left(\widetilde{\boldsymbol{w}}_{i-1}^{\mathsf{T}}\boldsymbol{\mathcal{B}}_i^{\mathsf{T}}\Sigma\boldsymbol{\mathcal{B}}_i\widetilde{\boldsymbol{w}}_{i-1}\right) &= \mathbb{E}\left[\widetilde{\boldsymbol{w}}_{i-1}^{\mathsf{T}}\mathbb{E}\left(\boldsymbol{\mathcal{B}}_i^{\mathsf{T}}\Sigma\boldsymbol{\mathcal{B}}_i\right)\widetilde{\boldsymbol{w}}_{i-1}\right] \\
&= \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}\|_{\Sigma'}^2
\end{aligned} \tag{3.66}$$

where we used the fact that $\{\boldsymbol{\mathcal{B}}_i, \widetilde{\boldsymbol{w}}_{i-1}\}$ are independent of each other and introduced the nonnegative-definite matrix

$$\Sigma' \triangleq \mathbb{E}\boldsymbol{\mathcal{B}}_i^\mathsf{T} \Sigma \boldsymbol{\mathcal{B}}_i \tag{3.67}$$

The second term in (3.65) is given by

$$\mathbb{E}\left(\boldsymbol{s}_i^\mathsf{T} \mathcal{G}^\mathsf{T} \Sigma \mathcal{G} \boldsymbol{s}_i\right) = \mathrm{Tr}\left(\mathcal{G}^\mathsf{T} \Sigma \mathcal{G} \mathbb{E}\boldsymbol{s}_i \boldsymbol{s}_i^\mathsf{T}\right)$$
$$= \mathrm{Tr}\left(\Sigma \mathcal{G} \mathcal{S} \mathcal{G}^\mathsf{T}\right) \tag{3.68}$$

where

$$\mathcal{S} \triangleq \mathrm{diag}\{\sigma_{v,1}^2 R_{u,1}, ..., \sigma_{v,1}^2 R_{u,N}\} \tag{3.69}$$

We then use the following equalities for arbitrary matrices $\{U, W, \Sigma\}$ of compatible dimensions [17]:

$$\mathrm{vec}(U\Sigma W) = (W^\mathsf{T} \otimes U)\sigma \quad \text{and} \quad \mathrm{Tr}(\Sigma W) = \left[\mathrm{vec}(W^\mathsf{T})\right]^\mathsf{T} \sigma \tag{3.70}$$

to rewrite (3.65) in the following equivalent form:

$$\mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|_\sigma^2 = \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}\|_{\mathcal{F}\sigma}^2 + \left[\mathrm{vec}(\mathcal{Y}^\mathsf{T})\right]^\mathsf{T} \sigma \tag{3.71}$$

where $\mathcal{F} \triangleq \mathbb{E}\left(\boldsymbol{\mathcal{B}}_i^\mathsf{T} \otimes \boldsymbol{\mathcal{B}}_i^\mathsf{T}\right)$. We note that (3.71) is not an actual recursion. To continue, we follow the arguments pursued in [92] and express the characteristic polynomial of the matrix $\mathcal{F}$ of size $L^2 \times L^2$ as

$$p(x) = \det(xI_{L^2} - \mathcal{F}) = x^{L^2} + p_{L^2-1}x^{L^2-1} + ... + p_0 \tag{3.72}$$

where

$$L = \sum_{k=1}^N M_k \tag{3.73}$$

We have that $p(\mathcal{F}) = 0$ by the Cayley-Hamilton Theorem [93] and, therefore,

$$\mathcal{F}^L = -p_0 - p_1\mathcal{F} - \cdots - p_{L^2-1}\mathcal{F}^{L^2-1} \tag{3.74}$$

87

Using (3.74) we get

$$
\underbrace{\begin{bmatrix} \mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|_\sigma^2 \\ \mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|_{\mathcal{F}\sigma}^2 \\ \mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|_{\mathcal{F}^2\sigma}^2 \\ \vdots \\ \mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|_{\mathcal{F}^{L^2-1}\sigma}^2 \end{bmatrix}}_{\mathcal{W}_i} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \\ -p_0 & -p_1 & -p_2 & \cdots & -p_{L^2-1} \end{bmatrix}}_{\mathcal{H}} \underbrace{\begin{bmatrix} \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}\|_\sigma^2 \\ \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}\|_{\mathcal{F}\sigma}^2 \\ \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}\|_{\mathcal{F}^2\sigma}^2 \\ \vdots \\ \mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}\|_{\mathcal{F}^{L^2-1}\sigma}^2 \end{bmatrix}}_{\mathcal{W}_{i-1}} + \underbrace{\begin{bmatrix} y^\mathsf{T}\sigma \\ y^\mathsf{T}\mathcal{F}\sigma \\ y^\mathsf{T}\mathcal{F}^2\sigma \\ \vdots \\ y^\mathsf{T}\mathcal{F}^{L^2-1}\sigma \end{bmatrix}}_{\mathcal{Z}}
$$

(3.75)

where

$$
y = \mathrm{vec}(\mathcal{Y}^\mathsf{T}) \tag{3.76}
$$

If the recursion (3.75) is stable, then the relation (3.71) is stable. Now, since the matrix $\mathcal{H}$ is in companion form, its eigenvalues are the roots of $p(x)$ from (3.72) [93] and equal to the eigenvalues of $\mathcal{F}$. The mean-square stability of the network is guaranteed if, and only if, $\mathcal{H}$ is a stable matrix, which means

$$
\rho(\mathcal{F}) < 1 \tag{3.77}
$$

Let us consider sufficiently small step-sizes $\{\mu_k\}$ such that

$$
0 < \mu_k \cdot \rho(R_{u,k}) \ll 1 \quad \text{for all } k \tag{3.78}
$$

Then, a reasonable approximate expression for $\mathcal{F}$ for sufficiently small step-sizes is

$$
\begin{aligned}
\mathcal{F} &= \mathbb{E}\left[\mathcal{A}_1(I - \boldsymbol{\mathcal{R}}_i\mathcal{M})\mathcal{A}_2 \otimes \mathcal{A}_1(I - \boldsymbol{\mathcal{R}}_i\mathcal{M})\mathcal{A}_2\right] \\
&= (\mathcal{A}_1 \otimes \mathcal{A}_1)\left[I - (\mathcal{R}^\mathsf{T}\mathcal{M}) \otimes I - I \otimes (\mathcal{R}\mathcal{M}) + O(\mathcal{M}^2)\right](\mathcal{A}_2 \otimes \mathcal{A}_2) \\
&\approx \mathcal{B}^\mathsf{T} \otimes \mathcal{B}^\mathsf{T}
\end{aligned} \tag{3.79}
$$

where we used the following Kronecker product property for matrices $\{A, B, C, D\}$ of compatible dimensions [92]:

$$
(A \otimes B)(C \otimes D) = AC \otimes BD \tag{3.80}
$$

88

Therefore, from (3.77) we have

$$\rho(\mathcal{F}) \approx \rho(\mathcal{B}^{\mathsf{T}}) \cdot \rho(\mathcal{B}^{\mathsf{T}}) = [\rho(\mathcal{B})]^2 < 1 \tag{3.81}$$

which means that sufficiently small step-sizes ensure mean-square stability in the relation (3.71).

CHAPTER 4

Information-Sharing with Self-Interested Agents

In this chapter, we examine the behavior of multi-agent networks where information-sharing is subject to a positive communications cost over the edges linking the agents. We consider a general mean-square-error formulation where all agents are interested in estimating the same target vector. We first show that, in the absence of any incentives to cooperate, the optimal strategy for the agents is to behave in a selfish manner with each agent seeking the optimal solution independently of the other agents. Pareto inefficiency arises as a result of the fact that agents are not using historical data to predict the behavior of their neighbors and to know whether they will reciprocate and participate in sharing information. Motivated by this observation, we develop a reputation protocol to summarize the opponents past actions into a reputation score, which can then be used to form a belief about the opponents subsequent actions. The reputation protocol entices agents to cooperate and turns their optimal strategy into an action-choosing strategy that enhances the overall social benefit of the network. In particular, we show that when the communications cost becomes large, the expected social benefit of the proposed protocol outperforms the social benefit that is obtained by cooperative agents that always share data. We perform a detailed mean-square-error analysis of the evolution of the network over three domains: far field, near-field, and middle-field, and show that the network behavior is stable for sufficiently small step-sizes. The various theoretical results are illustrated by numerical simulations.

4.1 Introduction

In most prior works of adaptive networks, agents are assumed to be cooperative and designed to follow certain distributed rules like the consensus strategy or the diffusion strategy discussed in the previous chapters. These rules generally include a self-learning step to update the agents' estimates using their local data, and a social-learning step to fuse and combine the estimates shared by neighboring agents in order to satisfy the constraints. However, when agents are selfish, they would not obey the preset rules unless these strategies conform to their own interests, such as minimizing their own costs. In this chapter, we assume that the agents can behave selfishly and that they, therefore, have the freedom to decide whether or not they want to participate in sharing information with their neighbors at every point in time. Under these conditions, the global social benefit for the network can be degraded unless a policy is introduced to entice agents to participate in the collaborative process despite their individual interests. In this article, we will address this difficulty in the context of adaptive networks where agents are continually subjected to streaming data, and where they can predict in real-time, from their successive interactions, how reliable their neighbors are and whether they can be trusted to share information based on their past history. This formulation is different from the useful work in [94], which considered one particular form of selfish behavior in the context of a game-theoretic formulation. In that work, the focus is on activating the self-learning and social learning steps simultaneously, and agents simply decide whether to enter into a sleep mode (to save energy) or to continue acquiring and processing data. In the framework considered in our chapter, agents always remain active and are continually acquiring data; the main question instead is to entice agents to participate in the collaborative information-sharing process regardless of their self-centered evaluations.

More specifically, we study the behavior of multi-agent networks where information-sharing is subject to a positive communication cost over the edges linking the agents. This situation is common in applications, such as information sharing over cognitive networks [95], online learning under communication bandwidth and/or latency constraints [96], [97, Ch. 14], and over social learning networks when the delivery of opinions involves

some costs such as messaging fees [98–100]. In our network model, each agent is self-interested and seeks to minimize its own sharing cost *and* estimation error. Motivated by the practical scenario studied in [95], we formulate a general mean-square error estimation problem where all agents are interested in estimating the same target parameter vector. Agents are assumed to be foresighted and to have bounded rationality [101] in the manner defined further ahead in the article. Then, we show that if left unattended, the dominant strategy for all agents is for them not to participate in the sharing of information, which leads to networks operating under an inefficient Pareto condition. This situation arises because agents do not have enough information to tell beforehand if their paired neighbors will reciprocate their actions (i.e., if an agent shares data with a second agent, will the second agent reciprocate and share data back?) This prediction-deficiency problem follows from the fact that agents are not using historical data to predict other agents' actions.

One method to deal with this inefficient scenario is to assume that agents adapt to their opponents' strategies and improve returns by forming some regret measures. In [102], a decision maker determines its action using a regret measure to evaluate the utility loss from the chosen action to the optimal action in the previous stage game. For multi-agent networks, a regret-based algorithm was proposed in [94] and [103] for agents to update their actions based on a weighted loss of the utility functions from the previous stage games. However, these works assume myopic agents and formulate repeated games with fixed utility functions over each stage game, which is different from the scenario considered in this article where the benefit of sharing information over adaptive networks continually evolves over time. This is because, as the estimation accuracy improves and/or the communication cost becomes expensive, the return to continue cooperating for estimation purposes falls and thus the act of cooperating with other agents becomes unattractive and inefficient. In this case, the regret measures computed from the previous stage games may not provide an accurate reference to the current stage game.

A second useful method to deal with Pareto inefficient and non-cooperative scenarios is to employ reputation schemes (e.g., [104–107]). In this method, foresighted agents use reputation scores to assess the willingness of other agents to cooperate; the scores are also

used to punish non-cooperative behavior. For example, the works [105,106] rely on discrete-value reputation scores, say, on a scale 1-10, and these scores are regularly updated according to the agents' actions. Similar to the regret learning references mentioned before, in our problem the utilities or cost functions of stage games change over time and evolve based on agents' estimates. Conventional reputation designs do not address this time variation within the payoff of agents, which will be examined more closely in our study. Motivated by these considerations, in Sec. 4.4, we propose a dynamic/adaptive reputation protocol that is based on the belief measure of future actions with real-time benefit predictions.

In our formulation, we assume a general random-pairing model similar to [9], where agents are randomly paired at the beginning of each time interval. This situation could occur, for example, due to an exogenous matcher or the mobility of the agents. The paired agents are assumed to follow a diffusion strategy [15–17,83], which includes an adaptation step and a consultation step, to iteratively update their estimates. Different from conventional diffusion strategies, the consultation step here is influenced by the random-pairing environment and by cooperation uncertainty. The interactions among self-interested agents are formulated as successive stage games of two players using pure strategies. To motivate agents to cooperate with each other, we formulate an adaptive reputation protocol to help agents jointly assess the instantaneous benefit of depreciating information and the transmission cost of sharing information. The reputation score helps agents to form a belief of their opponent's subsequent actions. Based on this belief, we entice agents to cooperate and turn their best response strategy into an action choosing strategy that conforms to Pareto efficiency and enhances the overall social benefit of the network.

In the performance evaluation, we are interested in ensuring the mean-square-error stability of the network instead of examining equilibria as is common in the game theoretical literature since our emphasis is on adaptation under successive time-variant stage games. The performance analysis is challenging due to the adaptive behavior by the agents. For this reason, we pursue the mean-square-error analysis of the evolution of the network over three domains: far-field, near-field, and middle-field, and show that the network behavior is stable for sufficiently small step-sizes. We also show that when information sharing becomes

costly, the expected social benefit of the proposed reputation protocol outperforms the social benefit that is obtained by cooperative agents that always share data.

4.2 System Model

4.2.1 Distributed Optimization and Communication Cost

We consider a subset of the modeling in Chapter 3 that in a connected network consisting of N agents, each agent has access to entire information about the global target vector. Therefore, when agents act independently of each other, each agent k would seek to estimate the $M \times 1$ vector w^o that minimizes an individual estimation cost function, which is denoted by $J_k^{\text{est}}(w) : \mathbb{C}^M \to \mathbb{R}$ to emphasize this cost function measuring the estimation performance. We assume each of the costs $\{J_k^{\text{est}}(w)\}$ is strongly convex for $k = 1, 2, \ldots, N$, and that all agents have the same objective so that all costs are minimized at the common location $w^o \in \mathbb{C}^{M \times 1}$.

In this chapter, we are interested in scenarios where agents can be motivated to cooperate among themselves as permitted by the network topology. We associate an extended cost function with each agent k, and denote it by $J_k(w, a_k)$. In this new cost, the scalar a_k is a binary variable that is used to model whether agent k is willing to cooperate and share information with its neighbors. The value $a_k = 1$ means that agent k is willing to share information (e.g., its estimate of w^o) with its neighbors, while the value $a_k = 0$ means that agent k is not willing to share information. The reason why agents may or may not share information is because this decision will generally entail some cost. We consider the scenario where a positive transmission cost, $c_k > 0$, is required for each act by agent k involving sharing an estimate with any of its neighbors. By taking c_k into consideration, the extended cost $J_k(w, a)$ that is now associated with agent k will consist of the sum of two components:

the estimation cost and the communication cost[1]:

$$J_k(w, a_k) \triangleq J_k^{\text{est}}(w) + J_k^{\text{com}}(a_k) \tag{4.1}$$

where the latter component is modeled as

$$J_k^{\text{com}}(a_k) \triangleq a_k c_k \tag{4.2}$$

We express the communication expense in the form (4.2) because, as described further ahead, when an agent k decides to share information, it will be sharing the information with one neighbor at a time; the cost for this communication will be $a_k c_k$. With regards to the estimation cost, $J_k^{\text{est}}(w)$, this measure can be selected in many ways. One common choice is the mean-square-error (MSE) cost, which we adopt in this chapter.

At each time instant $i \geq 0$, each agent k is assumed to have access to a scalar measurement $\boldsymbol{d}_k(i) \in \mathbb{C}$ and a $1 \times M$ regression vector $\boldsymbol{u}_{k,i} \in \mathbb{C}^{1 \times M}$ with covariance matrix $R_{u,k} \triangleq \mathbb{E}\boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i} > 0$. The regressors $\{\boldsymbol{u}_{k,i}\}$ are assumed to have zero-mean and to be temporally white and spatially independent, i.e.,

$$\mathbb{E}\boldsymbol{u}_{k,i}^* \boldsymbol{u}_{\ell,j} = R_{u,k}\delta_{k\ell}\delta_{ij} \tag{4.3}$$

in terms of the Kronecker delta function. The data $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$ are assumed to be related via the linear regression model:

$$\boldsymbol{d}_k(i) = \boldsymbol{u}_{k,i}w^o + \boldsymbol{v}_k(i) \tag{4.4}$$

where w^o is the common target vector to be estimated by the agents. In (4.4), the variable $\boldsymbol{v}_k(i) \in \mathbb{C}$ is a zero-mean white-noise process with power $\sigma_{v,k}^2$ that is assumed to be spatially independent, i.e.,

$$\mathbb{E}\boldsymbol{v}_k^*(i)\boldsymbol{v}_\ell(j) = \sigma_{v,k}^2\delta_{k\ell}\delta_{ij} \tag{4.5}$$

---

[1]We focus on the sum of the estimation cost and the communication cost due to its simplicity and meaningfulness in applications. Note that a possible generalization is to consider a penalty-based objective function $J_k^{\text{est}}(w) + p(J_k^{\text{com}}(a_k))$ for some penalty function $p(\cdot)$.

We further assume that the random processes $\boldsymbol{u}_{k,i}$ and $\boldsymbol{v}_\ell(i)$ are spatially and temporally independent for any k, ℓ, i, and j. Models of the form (4.4) are common in many applications, e.g., channel estimation, model fitting, target tracking, etc (see, e.g., [17]).

Let $\boldsymbol{w}_{k,i-1}$ denote the estimator for w^o that will be available to agent k at time $i-1$. We will describe in the sequel how agents evaluate these estimates. The corresponding *a-priori* estimation error is defined by

$$\boldsymbol{e}_{a,k}(i) \triangleq \boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{w}_{k,i-1} \tag{4.6}$$

and it measures how close the weight estimate matches the measurements $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$ to each other. In view of model (4.4), we can also write

$$\boldsymbol{e}_{a,k}(i) = \boldsymbol{u}_{k,i}\widetilde{\boldsymbol{w}}_{k,i-1} + \boldsymbol{v}_k(i) \tag{4.7}$$

in terms of the estimation error vector

$$\widetilde{\boldsymbol{w}}_{k,i-1} \triangleq w^o - \boldsymbol{w}_{k,i-1} \tag{4.8}$$

Motivated by these expressions and model (4.4), the instantaneous MSE cost that is associated with agent k based on the estimate from time $i-1$ is given by

$$\begin{aligned}
J_k^{\text{est}}(\boldsymbol{w}_{k,i-1}) &\triangleq \mathbb{E}|\boldsymbol{e}_{a,k}(i)|^2 \\
&= \mathbb{E}|\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{w}_{k,i-1}|^2 \\
&= \mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i-1}\|_{R_{u,k}}^2 + \sigma_{v,k}^2
\end{aligned} \tag{4.9}$$

Note that this MSE cost conforms to the strong convexity of J_k^{est} as we mentioned before. Combined with the action by agent k, the extended instantaneous cost at agent k that is based on the prior estimate, $\boldsymbol{w}_{k,i-1}$, is then given by:

$$J_k(\boldsymbol{w}_{k,i-1}, a_k) = \mathbb{E}|\boldsymbol{e}_{a,k}(i)|^2 + a_k c_k \tag{4.10}$$

4.2.2 Random-Pairing Model

We denote by $\mathcal{N}_k$ the neighborhood of each agent k, including itself. We consider a random pairing protocol for agents to share information at the beginning of every iteration cycle. The

pairing procedure can be executed either in a centralized or distributed manner. Centralized pairing schemes can be used when an online server randomly assigns its clients into pairs as in crowdsourcing applications [105, 106], or when a base-station makes pairing decisions for its mobile nodes for packet relaying [108]. Distributed paring schemes arise more naturally in the context of economic and market transactions [109]. In our formulation, we adopt a distributed pairing structure that takes neighborhoods into account when selecting pairs, as explained next.

We assume each agent k has bi-directional links to other agents in $\mathcal{N}_k$ and that agent k has a positive probability to be paired with any of its neighbors. Once two agents are paired, they can decide on whether to share or not their instantaneous estimates for w^o. We therefore model the result of the random-pairing process between each pair of agents k and $\ell \in \mathcal{N}_k \setminus \{k\}$ as temporally-independent Bernoulli random processes defined as:

$$\mathbf{1}_{k\ell}(i) = \mathbf{1}_{\ell k}(i) = \begin{cases} 1, & \text{with probability } p_{k\ell} = p_{\ell k} \\ 0, & \text{otherwise} \end{cases} \tag{4.11}$$

where $\mathbf{1}_{k\ell}(i) = 1$ indicates that agents k and ℓ are paired at time i and $\mathbf{1}_{k\ell}(i) = 0$ indicates that they are not paired. We are setting $\mathbf{1}_{k\ell}(i) = \mathbf{1}_{\ell k}(i)$ because these variables represent the same event: whether agents k and ℓ are paired, which results in $p_{k\ell} = p_{\ell k}$. For $\ell \notin \mathcal{N}_k$, we have $\mathbf{1}_{k\ell}(i) = 0$ since such pairs will never occur. For convenience, we use $\mathbf{1}_{kk}(i)$ to indicate the event that agent k is not paired with any agent $\ell \in \mathcal{N}_k \setminus \{k\}$ at time i, which happens with probability p_{kk}. Since each agent will pair itself with at most one agent at a time from its neighborhood, the following properties are directly obtained from the random-pairing procedure:

$$\sum_{\ell \in \mathcal{N}_k} \mathbf{1}_{k\ell}(i) = 1, \quad \sum_{\ell \in \mathcal{N}_k} p_{k\ell} = 1 \tag{4.12}$$

$$\mathbf{1}_{k\ell}(i)\mathbf{1}_{kq}(i) = 0, \quad \text{for } \ell \neq q \tag{4.13}$$

We assume that the random pairing indicators $\{\mathbf{1}_{k\ell}(i)\}$ for all k and ℓ are independent of the random variables $\{\boldsymbol{u}_{k,t}\}$ and $\{\boldsymbol{v}_k(t)\}$ for any time i and t. For example, a widely used setting in the literature is the fully-pairing network, which assumes a fully-connected

network topology [106, 110], i.e., $\mathcal{N}_k = \mathcal{N}$ for every agent k, where $\mathcal{N}$ denotes the set of all agents. The size $N = |\mathcal{N}|$ is assumed to be even and every agent is uniformly paired with exactly one agent in the network. Therefore, we have $N/2$ pairs at each time instant and the random-pairing probability becomes

$$p_{k\ell} = \begin{cases} \frac{1}{N-1}, & \text{for } \ell \neq k \\ 0, & \text{for } \ell = k \end{cases} \tag{4.14}$$

We will not be assuming fully-connected networks or fully-paired protocols and will deal more generally with networks that can be sparsely connected. Later in Sec. IV we will demonstrate a simple random-pairing protocol which can be implemented in a fully distributed manner.

4.2.3 Diffusion Strategy

Conventional diffusion strategies assume that the agents are cooperative (or obedient) and continuously share information with their neighbors as necessary [15, 17, 83]. With a common target w^o under full information, the ATP strategy (3.11) considered in Chapter 3 becomes the adapt-then-combine (ATC) version of traditional diffusion adaptation. In this strategy, each agent k updates its estimate, $\boldsymbol{w}_{k,i}$, according to the following relations:

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} + \mu \boldsymbol{u}_{k,i}^*[\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{w}_{k,i-1}] \tag{4.15}$$

$$\boldsymbol{w}_{k,i} = \sum_{\ell \in \mathcal{N}_k} \alpha_{\ell k} \boldsymbol{\psi}_{\ell,i} \tag{4.16}$$

where $\mu > 0$ is the step-size parameter of agent k, and the $\{\alpha_{\ell k}, \ell \in \mathcal{N}_k\}$ are nonnegative combination coefficients that add up to one. In implementation (4.15)–(4.16), each agent k computes an intermediate estimate $\boldsymbol{\psi}_{k,i}$ using its local data, and subsequently fuses the intermediate estimates from its neighbors. For the combination step (4.16), since agent k is allowed to interact with only one of its neighbors, then we rewrite (4.16) in terms of a single coefficient $0 \leq \alpha_k \leq 1$ as follows:

$$\boldsymbol{w}_{k,i} = \begin{cases} \alpha_k \boldsymbol{\psi}_{k,i} + (1 - \alpha_k)\boldsymbol{\psi}_{\ell,i}, & \text{if } \mathbf{1}_{k\ell}(i) = 1 \text{ for some } \ell \neq k \\ \boldsymbol{\psi}_{k,i}, & \text{otherwise} \end{cases} \tag{4.17}$$

We can capture both situations in (4.17) in a single equation as follows:

$$\boldsymbol{w}_{k,i} = \alpha_k \boldsymbol{\psi}_{k,i} + (1 - \alpha_k) \sum_{\ell \in \mathcal{N}_k} \mathbf{1}_{k\ell}(i) \boldsymbol{\psi}_{\ell,i} \tag{4.18}$$

In formulation (4.15) and (4.18), it is assumed that once agents k and ℓ are paired, they share information according to (4.18).

Let us now incorporate an additional layer into the algorithm in order to model instances of selfish behavior. When agents behave in a selfish (strategic) manner, even when agents k and ℓ are paired, each one of them may still decide (independently) to refuse to share information with the other agent for selfish reasons (for example, agent k may decide that this cooperation will cost more than the benefit it will reap for the estimation task). To capture this behavior, we use the specific notation $\boldsymbol{a}_{k\ell}(i)$, instead of $\boldsymbol{a}_k(i)$, to represent the action taken by agent k on agent ℓ at time i, and similarly for $\boldsymbol{a}_{\ell k}(i)$. Both agents will end up sharing information with each other only if $\boldsymbol{a}_{k\ell}(i) = \boldsymbol{a}_{\ell k}(i) = 1$, i.e., only when both agents are in favor of cooperating once they have been paired. We set $\boldsymbol{a}_{kk}(i) = 1$ for every time i. We can now rewrite the combination step (4.18) more generally as:

$$\boldsymbol{w}_{k,i} = \alpha_k \boldsymbol{\psi}_{k,i} + (1 - \alpha_k) \sum_{\ell \in \mathcal{N}_k} \mathbf{1}_{k\ell}(i) [\boldsymbol{a}_{\ell k}(i) \boldsymbol{\psi}_{\ell,i} + (1 - \boldsymbol{a}_{\ell k}(i)) \boldsymbol{\psi}_{k,i}] \tag{4.19}$$

From (4.19), when agent k is not paired with any agent at time i ($\mathbf{1}_{kk}(i) = 1$), we get $\boldsymbol{w}_{k,i} = \boldsymbol{\psi}_{k,i}$. On the other hand, when agent k is paired with some neighboring agent ℓ, which means $\mathbf{1}_{k\ell}(i) = 1$, we get

$$\boldsymbol{w}_{k,i} = \alpha_k \boldsymbol{\psi}_{k,i} + (1 - \alpha_k) [\boldsymbol{a}_{\ell k}(i) \boldsymbol{\psi}_{\ell,i} + (1 - \boldsymbol{a}_{\ell k}(i)) \boldsymbol{\psi}_{k,i}] \tag{4.20}$$

It is then clear that $\boldsymbol{a}_{\ell k}(i) = 0$ results in $\boldsymbol{w}_{k,i} = \boldsymbol{\psi}_{k,i}$, while $\boldsymbol{a}_{\ell k}(i) = 1$ results in a combination of the estimates of agents k and ℓ. In other words, when $\mathbf{1}_{k\ell}(i) = 1$:

$$\boldsymbol{w}_{k,i} = \begin{cases} \boldsymbol{\psi}_{k,i}, & \text{if } \boldsymbol{a}_{\ell k}(i) = 0 \\ \alpha_k \boldsymbol{\psi}_{k,i} + (1 - \alpha_k) \boldsymbol{\psi}_{\ell,i}, & \text{if } \boldsymbol{a}_{\ell k}(i) = 1 \end{cases} \tag{4.21}$$

In the sequel, we assume that agents update and combine their estimates using (4.15) and (4.19). One important question to address is how the agents determine their actions $\{\boldsymbol{a}_{k\ell}(i)\}$.

4.3 Agent Interactions

When an arbitrary agent k needs to decide on whether to set its action to $\boldsymbol{a}_{k\ell}(i) = 1$ (i.e., to cooperate) or $\boldsymbol{a}_{k\ell}(i) = 0$ (i.e., not to cooperate), it generally cannot tell beforehand whether agent ℓ will reciprocate. In this section, we first show that when self-interested agents are boundedly rational and incapable of transforming the past actions of neighbors into a prediction of their future actions, then the dominant strategy for each agent will be to choose noncooperation. Consequently, the entire network becomes noncooperative. Later, in Sec. 4.4, we explain how to address this inefficient scenario by proposing a protocol that will encourage cooperation.

4.3.1 Long-Term Discounted Cost Function

To begin with, let us examine the interaction between a pair of agents, such as k and ℓ, at some time instant i ($\boldsymbol{1}_{k\ell}(i) = 1$). We assume that agents k and ℓ simultaneously select their actions $\boldsymbol{a}_{k\ell}(i)$ and $\boldsymbol{a}_{\ell k}(i)$ by using some pure strategies (i.e., agents set their action variables by using data or realizations that are available to them, such as the estimates $\{\boldsymbol{w}_{k,i-1}, \boldsymbol{w}_{\ell,i-1}\}$, rather than select their actions according to some probability distributions)[2]. The criterion for setting $\boldsymbol{a}_{k\ell}(i)$ by agent k is to optimize agent k's payoff, which incorporates both the estimation cost, affected by agent ℓ's own action $\boldsymbol{a}_{\ell k}(i)$, and the communication cost, determined by agent k's action $\boldsymbol{a}_{k\ell}(i)$. Therefore, the instantaneous cost incurred by agent k is a mapping function from the action space $(\boldsymbol{a}_{k\ell}(i), \boldsymbol{a}_{\ell k}(i))$ to a real value. In order to account for selfish behavior, we need to modify the notation used in (4.1) to incorporate the actions of both agents k and ℓ. In this way, we need to denote the value of the cost incurred by agent k at time i, after $\boldsymbol{w}_{k,i-1}$ is updated to $\boldsymbol{w}_{k,i}$, more explicitly by $J_k(\boldsymbol{a}_{k\ell}(i), \boldsymbol{a}_{\ell k}(i))$ and

---

[2]In our scenario, the discrete action set $\boldsymbol{a}_{k\ell}(i) \in \{0,1\}$ will be shown to lead to threshold-based pure strategies — see Sec. 4.4.2.

it is given by:

$$J_k(\boldsymbol{a}_{k\ell}(i), \boldsymbol{a}_{\ell k}(i)) = \begin{cases} J_k^{\text{est}}(\boldsymbol{w}_{k,i} = \boldsymbol{\psi}_{k,i}), & \text{if } (0,0) \\ J_k^{\text{est}}(\boldsymbol{w}_{k,i} = \alpha_k \boldsymbol{\psi}_{k,i} + (1-\alpha_k)\boldsymbol{\psi}_{\ell,i}), & \text{if } (0,1) \\ J_k^{\text{est}}(\boldsymbol{w}_{k,i} = \boldsymbol{\psi}_{k,i}) + c_k, & \text{if } (1,0) \\ J_k^{\text{est}}(\boldsymbol{w}_{k,i} = \alpha_k \boldsymbol{\psi}_{k,i} + (1-\alpha_k)\boldsymbol{\psi}_{\ell,i}) + c_k, & \text{if } (1,1) \end{cases}$$

(4.22)

For example, the first line on the right-hand side of (4.22) corresponds to the situation in which none of the agents decides to cooperate. In that case, agent k can only rely on its intermediate estimate, $\boldsymbol{\psi}_{k,i}$, to improve its estimation accuracy. In comparison, the second line in (4.22) corresponds to the situation in which agent ℓ is willing to share its estimate but not agent k. In this case, agent k is able to perform the second combination step in (4.21) and enhance its estimation accuracy. In the third line in (4.22), agent ℓ does not cooperate while agent k does. In this case, agent k incurs a communication cost, c_k. Similarly, for the last line in (4.22), both agents cooperate. In this case, agent k is able to perform the second step in (4.21) while incurring a cost c_k.

We can write (4.22) more compactly as follows:

$$J_k(\boldsymbol{a}_{k\ell}(i), \boldsymbol{a}_{\ell k}(i)) = J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(i)) + \boldsymbol{a}_{k\ell}(i)c_k$$

(4.23)

where we introduced

$$J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(i)) \triangleq \begin{cases} J_k^{\text{est}}(\boldsymbol{w}_{k,i} = \boldsymbol{\psi}_{k,i}), & \text{if } \boldsymbol{a}_{\ell k}(i) = 0 \\ J_k^{\text{est}}(\boldsymbol{w}_{k,i} = \alpha_k \boldsymbol{\psi}_{k,i} + (1-\alpha_k)\boldsymbol{\psi}_{\ell,i}), & \text{if } \boldsymbol{a}_{\ell k}(i) = 1 \end{cases}$$

(4.24)

The function $J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(i))$ helps make explicit the influence of the action by agent ℓ on the estimation accuracy that is ultimately attained by agent k.

Now, the random-pairing process occurs repeatedly over time and, moreover, agents may leave the network. For this reason, rather than rely on the instantaneous cost function in (4.22), agent k will determine its action at time i by instead minimizing an expected

101

Table 4.1: The expected long-term cost functions $J_{k,i}^1$ and $J_{\ell,i}^1$.

Agent k / Agent ℓ	$a_{k\ell}(i) = 0$	$a_{k\ell}(i) = 1$
$a_{\ell k}(i) = 0$	$\mathbb{E}[J_\ell^{\mathrm{act}}(a_{k\ell}(i) = 0)\|\boldsymbol{w}_{\ell,i-1}]$ $\mathbb{E}[J_k^{\mathrm{act}}(a_{\ell k}(i) = 0)\|\boldsymbol{w}_{k,i-1}]$	$\mathbb{E}[J_\ell^{\mathrm{act}}(a_{k\ell}(i) = 1)\|\boldsymbol{w}_{\ell,i-1}]$ $\mathbb{E}[J_k^{\mathrm{act}}(a_{\ell k}(i) = 0)\|\boldsymbol{w}_{k,i-1}] + c_k$
$a_{\ell k}(i) = 1$	$\mathbb{E}[J_\ell^{\mathrm{act}}(a_{k\ell}(i) = 0)\|\boldsymbol{w}_{\ell,i-1}] + c_\ell$ $\mathbb{E}[J_k^{\mathrm{act}}(a_{\ell k}(i) = 1)\|\boldsymbol{w}_{k,i-1}]$	$\mathbb{E}[J_\ell^{\mathrm{act}}(a_{k\ell}(i) = 1)\|\boldsymbol{w}_{\ell,i-1}] + c_\ell$ $\mathbb{E}[J_k^{\mathrm{act}}(a_{\ell k}(i) = 1)\|\boldsymbol{w}_{k,i-1}] + c_k$

long-term discounted cost function of the form defined by

$$
J_{k,i}^\infty [a_{k\ell}(i), a_{\ell k}(i)|\boldsymbol{w}_{k,i-1}]
$$
$$
\triangleq \sum_{t=i}^{\infty} \delta_k^{t-i} \mathbb{E}\Big[J_k(\boldsymbol{a}_{k\ell}(t), \boldsymbol{a}_{\ell k}(t))\Big|\boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i) = a_{k\ell}(i), \boldsymbol{a}_{\ell k}(i) = a_{\ell k}(i)\Big]
$$
$$
= \sum_{t=i}^{\infty} \delta_k^{t-i} \mathbb{E}\Big[J_k^{\mathrm{act}}(\boldsymbol{a}_{\ell k}(t)) + \boldsymbol{a}_{k\ell}(t)c_k\Big|\boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i) = a_{k\ell}(i), \boldsymbol{a}_{\ell k}(i) = a_{\ell k}(i)\Big] \qquad (4.25)
$$

where $\delta_k \in (0,1)$ is a discount factor to model future network uncertainties and the fore-sightedness level of agent k. The expectation is taken over all randomness for $t \geq i$ and is conditioned on the estimate $\boldsymbol{w}_{k,i-1}$ when the actions $a_{k\ell}(i)$ and $a_{\ell k}(i)$ are selected. Formulation (4.25) is meant to assess the influence of the action selected at time i by agent k on its cumulative (but discounted) future costs. More specifically, whenever $\mathbf{1}_{k\ell}(i) = 1$, agent k selects its action $a_{k\ell}(i)$ at time i to minimize the expected long-term discounted cost given $\boldsymbol{w}_{k,i-1}$:

$$
\min_{a_{k\ell}(i)\in\{0,1\}} J_{k,i}^\infty [a_{k\ell}(i), a_{\ell k}(i)|\boldsymbol{w}_{k,i-1}] \qquad (4.26)
$$

Based on the payoff function in (4.25), we can formally regard the interaction between agents as consisting of stage games with recurrent random pairing. The stage information-sharing game for $\mathbf{1}_{k\ell}(i) = 1$ is a tuple $(\mathbb{N}, \mathbb{A}, \mathbb{J})$, where $\mathbb{N} \triangleq \{k, \ell\}$ is the set of players, and $\mathbb{A} \triangleq \mathbb{A}_k \times \mathbb{A}_\ell$ is the Cartesian product of binary sets $\mathbb{A}_k = \mathbb{A}_\ell \triangleq \{1, 0\}$ representing available actions for agents k and ℓ, respectively. The action profile is $a(i) \triangleq (a_{k\ell}(i), a_{\ell k}(i)) \in \mathbb{A}$.

Moreover, $\mathbb{J} = \{J_{k,i}^\infty, J_{\ell,i}^\infty\}$ is the set of real-valued long-term costs defined over $\mathbb{A} \to \mathbb{R}$ for agents k and ℓ, respectively. We remark that since $J_{k,i}^\infty$ depends on $\boldsymbol{w}_{k,i-1}$, its value generally varies from stage to stage. As a result, each agent k faces a dynamic game structure with repeated interactions in contrast to conventional repeated games as in [111, 112] where the game structure is fixed over time. Time variation is an essential feature that arises when we examine selfish behavior over adaptive networks.

Therefore, solving problem (4.26) involves the forecast of future game structures and future actions chosen by the opponent. These two factors are actually coupled and influence each other; this fact makes prediction under such conditions rather challenging. To continue with the analysis, we adopt a common assumption from the literature that agents have computational constraints. In particular, we assume the agents have bounded rationality [101, 113, 114]. In our context, this means that the agents have limited capability to forecast future game structures and are therefore obliged to assume that future parameters remain unchanged at current values. We will show how this assumption enables each agent k to evaluate $J_{k,i}^\infty$ in later discussions.

Assumption 4.1 (Bounded rationality). *Every agent k solves the optimization problem (4.26) under the assumptions:*

$$\boldsymbol{w}_{k,t} = \boldsymbol{w}_{k,i-1}, \quad \mathbf{1}_{k\ell}(t) = \mathbf{1}_{k\ell}(i), \quad \text{for } t \geq i \tag{4.27}$$

$\square$

We note that the above assumption is only made by the agent at time i while solving problem (4.26); the actual estimates $\boldsymbol{w}_{k,t}$ and pairing choices $\mathbf{1}_{k\ell}(t)$ will continue to evolve over time. We further assume that the bounded rationality assumption is common knowledge to all agents in the network[3].

---

[3]Common knowledge of p means that each agent knows p, each agent knows that all other agents know p, each agent knows that all other agents know that all the agents know p, and so on [115].

4.3.2 Pareto Inefficiency

In this section, we show that if no further measures are taken, then Pareto inefficiency may occur. Thus, assume that the agents are unable to store the history of their actions and the actions of their neighbors. Each agent k only has access to its immediate estimate $\boldsymbol{w}_{k,i-1}$, which can be interpreted as a state variable at time $i-1$ for agent k. In this case, each agent k will need to solve (4.26) under Assumption 4.1. It then follows that agent k will predict the same action for future time instants:

$$\boldsymbol{a}_{k\ell}(t) = a_{k\ell}(i), \quad \text{for } t > i \tag{4.28}$$

Furthermore, since the bounded rationality condition is common knowledge, agent k knows that the same future actions are used by agent ℓ, i.e.,

$$\boldsymbol{a}_{\ell k}(t) = a_{\ell k}(i), \quad \text{for } t > i \tag{4.29}$$

Using (4.28) and (4.29), agent k obtains

$$
\begin{aligned}
J_{k,i}^{\infty}\left[a_{k\ell}(i), a_{\ell k}(i)|\boldsymbol{w}_{k,i-1}\right] &= \sum_{t=i}^{\infty} \delta_k^{t-i} \mathbb{E}\left[J_k(a_{k\ell}(i), a_{\ell k}(i))\Big|\boldsymbol{w}_{k,i-1}\right] \\
&= \frac{1}{1-\delta_k} \cdot \mathbb{E}\left[J_k(a_{k\ell}(i), a_{\ell k}(i))\Big|\boldsymbol{w}_{k,i-1}\right] \\
&= \frac{1}{1-\delta_k}\left(\mathbb{E}[J_k^{\text{act}}(a_{\ell k}(i))|\boldsymbol{w}_{k,i-1}] + a_{k\ell}(i)c_k\right)
\end{aligned}
\tag{4.30}
$$

Therefore, the optimization problem (4.26) reduces to the following minimization problem:

$$\min_{a_{k\ell}(i)\in\{0,1\}} J_{k,i}^1\left(a_{k\ell}(i), a_{\ell k}(i)\right) \tag{4.31}$$

where

$$J_{k,i}^1\left(a_{k\ell}(i), a_{\ell k}(i)\right) \triangleq \mathbb{E}[J_k^{\text{act}}(a_{\ell k}(i))|\boldsymbol{w}_{k,i-1}] + a_{k\ell}(i)c_k \tag{4.32}$$

is the expected cost of agent k given $\boldsymbol{w}_{k,i-1}$ — compare with (4.23). Table 4.1 summarizes the values of $J_{k,i}^1$ and $J_{\ell,i}^1$ for both agents under their respective actions. From the entries in the table, we conclude that choosing action $a_{k\ell}(i) = 0$ is the dominant strategy for agent k regardless of the action chosen by agent ℓ because its cost will be the smallest it can be

104

(a) $\gamma_{k,i} > 1$ and $\gamma_{\ell,i} > 1$

(b) $\gamma_{k,i} < 1$ and $\gamma_{\ell,i} < 1$
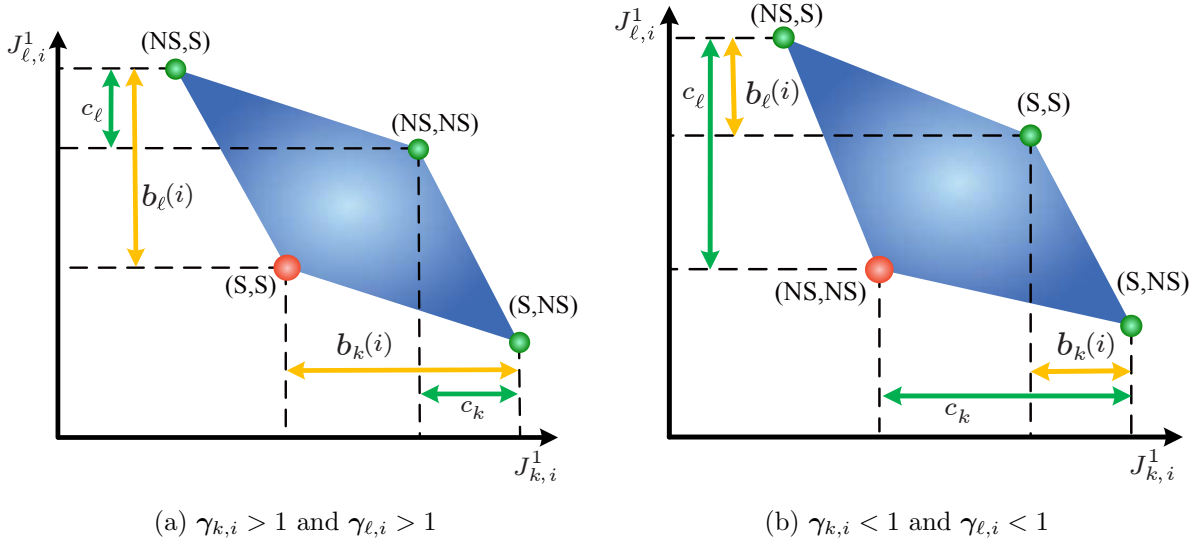
Figure 4.1: Illustration of the behavior of the payoffs in terms of the size of the benefit-cost ratios ("S" and "NS" refer to the actions "share" and "do not share", respectively).

in that situation. Likewise, the dominant strategy for agent ℓ is $a_{\ell k}(i) = 0$ regardless of the action chosen by agent k. Therefore, the action profile $(a_{k\ell}(i), a_{\ell k}(i)) = (0,0)$ is the unique outcome as a Nash and dominant strategy equilibrium for every stage game.

However, this resulting action profile will be Pareto inefficient for both agents if it can be verified that the alternative action profile $(1,1)$, where both agents cooperate, can lead to improved payoff values for both agents in comparison to the strategy $(0,0)$. To characterize when this is possible, let us denote the expected payoff for agent k when agent ℓ selects $a_{\ell k}(i) = 0$ by

$$s_{k,i}^0(a_{k\ell}(i)) \triangleq \mathbb{E}[J_k^{\text{act}}(a_{\ell k}(i) = 0)|\boldsymbol{w}_{k,i-1}] + a_{k\ell}(i)c_k \qquad (4.33)$$

Likewise, when $a_{\ell k}(i) = 1$, we denote the expected payoff for agent k by

$$s_{k,i}^1(a_{k\ell}(i)) \triangleq \mathbb{E}[J_k^{\text{act}}(a_{\ell k}(i) = 1)|\boldsymbol{w}_{k,i-1}] + a_{k\ell}(i)c_k \qquad (4.34)$$

The benefit for agent k from agent ℓ's sharing action, defined as the improvement from

$\boldsymbol{s}_{k,i}^0(a_{k\ell}(i))$ to $\boldsymbol{s}_{k,i}^1(a_{k\ell}(i))$, is seen to be independent of $a_{k\ell}(i)$:

$$
\begin{aligned}
\boldsymbol{b}_k(i) & \triangleq \boldsymbol{s}_{k,i}^0(a_{k\ell}(i)) - \boldsymbol{s}_{k,i}^1(a_{k\ell}(i)) \\
& = \mathbb{E}[J_k^{\text{act}}(a_{\ell k}(i)=0)|\boldsymbol{w}_{k,i-1}] - \mathbb{E}[J_k^{\text{act}}(a_{\ell k}(i)=1)|\boldsymbol{w}_{k,i-1}] \\
& = \mathbb{E}\left[J_k^{\text{est}}\left(\boldsymbol{w}_{k,i}=\boldsymbol{\psi}_{k,i}\right)|\boldsymbol{w}_{k,i-1}\right] \\
& \quad - \mathbb{E}\left[J_k^{\text{est}}\left(\boldsymbol{w}_{k,i}=\alpha_k\boldsymbol{\psi}_{k,i}+(1-\alpha_k)\boldsymbol{\psi}_{\ell,i}\right)|\boldsymbol{w}_{k,i-1}\right]
\end{aligned}
\tag{4.35}
$$

Now, note from definition (4.6) that

$$
\mathbb{E}\left[J_k^{\text{est}}(\boldsymbol{w}_{k,i})|\boldsymbol{w}_{k,i-1}\right] = \mathbb{E}\left[\left|\boldsymbol{d}_k(i+1)-\boldsymbol{u}_{k,i+1}\boldsymbol{w}_{k,i}\right|^2\Big|\boldsymbol{w}_{k,i-1}\right]
\tag{4.36}
$$

so that

$$
\begin{aligned}
\mathbb{E}[J_k^{\text{act}}(a_{\ell k}(i)=0)|\boldsymbol{w}_{k,i-1}] & = \mathbb{E}\left[J_k^{\text{est}}\left(\boldsymbol{w}_{k,i}=\boldsymbol{\psi}_{k,i}\right)|\boldsymbol{w}_{k,i-1}\right] \\
& = \mathbb{E}\left[\left|\boldsymbol{d}_k(i+1)-\boldsymbol{u}_{k,i+1}\boldsymbol{\psi}_{k,i}\right|^2\Big|\boldsymbol{w}_{k,i-1}\right] \\
& = \mathbb{E}\left[\left|\boldsymbol{u}_{k,i+1}\widetilde{\boldsymbol{\psi}}_{k,i}+\boldsymbol{v}_k(i+1)\right|^2\Big|\boldsymbol{w}_{k,i-1}\right] \\
& = \mathbb{E}\left[\|\widetilde{\boldsymbol{\psi}}_{k,i}\|_{R_{u,k}}^2\Big|\boldsymbol{w}_{k,i-1}\right] + \sigma_{v,k}^2
\end{aligned}
\tag{4.37}
$$

where $\widetilde{\boldsymbol{\psi}}_{k,i} \triangleq w^o - \boldsymbol{\psi}_{k,i}$ and, similarly,

$$
\begin{aligned}
\mathbb{E}[J_k^{\text{act}}(a_{\ell k}(i)=1)|\boldsymbol{w}_{k,i-1}] & = \mathbb{E}\left[J_k^{\text{est}}\left(\boldsymbol{w}_{k,i}=\alpha_k\boldsymbol{\psi}_{k,i}+(1-\alpha_k)\boldsymbol{\psi}_{\ell,i}\right)|\boldsymbol{w}_{k,i-1}\right] \\
& = \mathbb{E}\left[\|\alpha_k\widetilde{\boldsymbol{\psi}}_{k,i}+(1-\alpha_k)\widetilde{\boldsymbol{\psi}}_{\ell,i}\|_{R_{u,k}}^2\Big|\boldsymbol{w}_{k,i-1}\right] + \sigma_{v,k}^2
\end{aligned}
\tag{4.38}
$$

Then, the benefit $\boldsymbol{b}_k(i)$ becomes

$$
\boldsymbol{b}_k(i) = \mathbb{E}\left[\|\widetilde{\boldsymbol{\psi}}_{k,i}\|_{R_{u,k}}^2\Big|\boldsymbol{w}_{k,i-1}\right] - \mathbb{E}\left[\|\alpha_k\widetilde{\boldsymbol{\psi}}_{k,i}+(1-\alpha_k)\widetilde{\boldsymbol{\psi}}_{\ell,i}\|_{R_{u,k}}^2\Big|\boldsymbol{w}_{k,i-1}\right]
\tag{4.39}
$$

Note that $\boldsymbol{b}_k(i)$ is determined by the variable $\boldsymbol{w}_{k,i-1}$ and does not depend on the actions $a_{\ell k}(i)$ and $a_{k\ell}(i)$. We will explain how agents assess the information $\boldsymbol{b}_k(i)$ to choose actions further ahead in Sec. IV-C. Now, let us define the benefit-cost ratio as the ratio of the estimation benefit to the communication cost:

$$
\gamma_k(i) \triangleq \frac{\boldsymbol{b}_k(i)}{c_k}
\tag{4.40}
$$

106

Then, the action profile $(1,1)$ in the game defined in Table 4.1 is Pareto superior to the action profile $(0,0)$ when both of the following two conditions hold

$$\gamma_k(i) > 1 \;\; \text{and} \;\; \gamma_\ell(i) > 1 \;\; \iff \;\; \begin{cases} c_k < \boldsymbol{b}_k(i) \\ c_\ell < \boldsymbol{b}_\ell(i) \end{cases} \tag{4.41}$$

On the other hand, the action profile $(0,0)$ is Pareto superior to the action profile $(1,1)$ if, and only if,

$$\gamma_k(i) < 1 \;\; \text{and} \;\; \gamma_\ell(i) < 1 \tag{4.42}$$

In Fig. 4.1(a), we illustrate how the values of the payoffs compare to each other when (4.41) holds for the four possibilities of action profiles. It is seen from this figure that when $\gamma_k(i) > 1$ and $\gamma_\ell(i) > 1$, the action profile (S,S), i.e., $(1,1)$ in (4.32), is Pareto optimal and that the dominant strategy (NS,NS), i.e., $(0,0)$ in (4.32), is inefficient and leads to worse performance (which is a manifestation of the famous prisoner's dilemma problem [116]). On the other hand, if $\gamma_k(i) < 1$ and $\gamma_\ell(i) < 1$, then we are led to Fig. 4.1(b), where the action profile (NS,NS) becomes Pareto optimal and superior to (S,S). We remark that (NS,S) and (S,NS) are also Pareto optimal in both cases but not preferred in this chapter because they are only beneficial for one single agent.

4.4   Adaptive Reputation Protocol Design

As shown above, when both $\gamma_k(i) > 1$ and $\gamma_\ell(i) > 1$, the Pareto optimal strategies for agents k and ℓ correspond to cooperation; when both $\gamma_k(i) < 1$ and $\gamma_\ell(i) < 1$, the Pareto optimal strategies for agents k and ℓ reduce to non-cooperation. Since agents are self-interested and boundedly rational, we showed earlier that if left without incentives, their dominant strategy is to avoid sharing information because they cannot tell beforehand if their paired neighbor will reciprocate. This Pareto inefficiency therefore arises from the fact that agents are not using historical data to predict other agents' actions. We now propose a reputation protocol to summarize the opponent's past actions into a reputation score. The score will help agents to form a belief of their opponent's subsequent actions. Based on this belief, we will be able

to provide agents with a measure that entices them to cooperate. We will show, for example, that the best response rule for agents will be to cooperate whenever $\gamma_k(i)$ is large and not to cooperate whenever $\gamma_k(i)$ is small, in conformity with the Pareto-efficient design.

4.4.1 Reputation Protocol

Reputation scores have been used before in the literature as a mechanism to encourage cooperation [106, 117, 118]. Agents that cooperate are rewarded with higher scores; agents that do not cooperate are penalized with lower scores. For example, eBay uses a cumulative score mechanism, which simply sums the sellers feedback scores from all previous periods to provide buyers and sellers with trust evaluation [119]. Likewise, Amazon.com implements a reputation system by using an average score mechanism that averages the feedback scores from the previous periods [120]. However, as already explained in [118], cheating can occur over time in both cumulative and average score mechanisms because past scores carry a large weight in determining the current reputation. To overcome this problem, and in a manner similar to exponential weighting in adaptive filter designs [92], an exponentially-weighted moving average mechanism that gives higher weights to more recent actions is discussed in [118]. We follow a similar weighting formulation, with the main difference being that the reputation scores now need to be adapted in response to the evolution of the estimation task over the network. The construction can be described as follows.

When $\mathbf{1}_{k\ell}(i) = 1$, meaning that agent k is paired with agent ℓ, the reputation score $\boldsymbol{\theta}_{\ell k}(i) \in [0, 1]$ that is maintained by agent k for its neighbor ℓ is updated as:

$$\boldsymbol{\theta}_{\ell k}(i + 1) = r_k \boldsymbol{\theta}_{\ell k}(i) + (1 - r_k)\boldsymbol{a}_{\ell k}(i) \tag{4.43}$$

where $r_k \in (0, 1)$ is a smoothing factor for agent k to control the dynamics of the reputation updates. On the other hand, if $\mathbf{1}_{k\ell}(i) = 0$, the reputation score $\boldsymbol{\theta}_{\ell k}(i + 1)$ remains as $\boldsymbol{\theta}_{\ell k}(i)$. We can compactly describe the reputation rule as

$$\boldsymbol{\theta}_{\ell k}(i + 1) = \mathbf{1}_{k\ell}(i) \left[r_k \boldsymbol{\theta}_{\ell k}(i) + (1 - r_k)\boldsymbol{a}_{\ell k}(i) \right] + (1 - \mathbf{1}_{k\ell}(i))\boldsymbol{\theta}_{\ell k}(i) \tag{4.44}$$

Directly applying the above reputation formulation, however, can cause a loss in adaptation ability over the network. For example, the network would become permanently non-

cooperative when agent ℓ chooses $\boldsymbol{a}_{\ell k}(i) = 0$ for long consecutive iterations. That is because, in that case, the reputation score $\boldsymbol{\theta}_{\ell k}(i)$ will decay exponentially to zero, which keeps agent k from choosing $\boldsymbol{a}_{k\ell}(i) = 1$ in the future. In order to avoid this situation, we set a lowest value for the reputation score to a small positive threshold $0 < \varepsilon \ll 1$, i.e.,

$$\boldsymbol{\theta}_{\ell k}(i+1) = \mathbf{1}_{k\ell}(i) \cdot \max\{r_k \boldsymbol{\theta}_{\ell k}(i) + (1 - r_k)\boldsymbol{a}_{\ell k}(i), \varepsilon\} + (1 - \mathbf{1}_{k\ell}(i))\boldsymbol{\theta}_{\ell k}(i) \tag{4.45}$$

and thus $\boldsymbol{\theta}_{\ell k}(i) \in [\varepsilon, 1]$.

The reputation scores can now be utilized to evaluate the belief by agent k of subsequent actions by agent ℓ. To explain how this can be done, we argue that agent k would expect the probability of $\boldsymbol{a}_{\ell k}(t) = 1$, i.e., the probability that agent ℓ is willing to cooperate, to be an increasing function of both $\boldsymbol{\theta}_{\ell k}(t)$ and $\boldsymbol{\theta}_{k\ell}(t)$ for $t \geq i$. Specifically, if we denote this belief probability by $B(\boldsymbol{a}_{\ell k}(t) = 1)$, then it is expected to satisfy:

$$\frac{\partial B(\boldsymbol{a}_{\ell k}(t) = 1)}{\partial \boldsymbol{\theta}_{\ell k}(t)} \geq 0, \quad \frac{\partial B(\boldsymbol{a}_{\ell k}(t) = 1)}{\partial \boldsymbol{\theta}_{k\ell}(t)} \geq 0 \tag{4.46}$$

The first property is motivated by the fact that according to the history of actions, a higher value for $\boldsymbol{\theta}_{\ell k}(t)$ indicates that agent ℓ has higher willingness to share estimates. The second property is motivated by the fact that lower values for $\boldsymbol{\theta}_{k\ell}(t)$ mean that agent k has rarely shared estimates with agent ℓ in the recent past. Therefore, it can be expected that agent ℓ will have lower willingness to share information for lower values of $\boldsymbol{\theta}_{k\ell}(t)$. Based on this argument, we suggest a first-order construction for measuring belief with respect to both $\boldsymbol{\theta}_{\ell k}(t)$ and $\boldsymbol{\theta}_{k\ell}(t)$ as follows (other constructions are of course possible; our intent is to keep the complexity of the solution low while meeting the desired objectives):

$$B(\boldsymbol{a}_{\ell k}(t) = 1) = \boldsymbol{\theta}_{k\ell}(t) \cdot \boldsymbol{\theta}_{\ell k}(t), \quad t \geq i \tag{4.47}$$

which satisfies both properties in (4.46) and where $B(\boldsymbol{a}_{\ell k}(t) = 1) \in [\varepsilon^2, 1]$. Therefore, the reputation protocol implements (4.45) and (4.47) repeatedly. Each agent k will then employ the reference knowledge $\mathbb{K}_i \triangleq \{\boldsymbol{\theta}_{k\ell}(i), \boldsymbol{\theta}_{\ell k}(i), B(\boldsymbol{a}_{\ell k}(i) = 1)\}$ to select its action $a_{k\ell}(i)$ as described next.

4.4.2 Best Response Rule

The belief measure (4.47) provides agent k with additional information about agent ℓ's actions. That is, with (4.47), agent k can treat $\boldsymbol{a}_{\ell k}(t)$ as a random variable with distribution $B(\boldsymbol{a}_{\ell k}(t) = 1)$ for $t \geq i$. Then, the best response of agent k is obtained by solving the following optimization problem:

$$\min_{a_{k\ell}(i) \in \{0,1\}} J_{k,i}^{\infty'}[a_{k\ell}(i)|\boldsymbol{w}_{k,i-1}] \tag{4.48}$$

where $J_{k,i}^{\infty'}[a_{k\ell}(i)|\boldsymbol{w}_{k,i-1}]$ is defined by

$$
\begin{aligned}
J_{k,i}^{\infty'}&[a_{k\ell}(i)|\boldsymbol{w}_{k,i-1}] \\
&= \sum_{t=i}^{\infty} \delta_k^{t-i} \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t)) + \boldsymbol{a}_{k\ell}(t)c_k \Big| \boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i) = a_{k\ell}(i), \mathbb{K}_i\right] \\
&= \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(i)) \Big| \boldsymbol{w}_{k,i-1}, \mathbb{K}_i\right] + a_{k\ell}(i)c_k \\
&\quad + \sum_{t=i+1}^{\infty} \delta_k^{t-i} \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t)) + \boldsymbol{a}_{k\ell}(t)c_k \Big| \boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i) = a_{k\ell}(i), \mathbb{K}_i\right]
\end{aligned} \tag{4.49}
$$

and involves an additional expectation over the distribution of $\boldsymbol{a}_{\ell k}(t)$ — compare with (4.25). Similarly to Assumption 4.1, we assume the bounded rationality of the agents extends to the reputation scores $\boldsymbol{\theta}_{\ell k}(t)$ for $t \geq i$.

Assumption 4.2 (Extended bounded rationality). *We extend the assumption of bounded rationality from (4.27) to also include:*

$$\boldsymbol{\theta}_{\ell k}(t) = \boldsymbol{\theta}_{\ell k}(i), \quad \text{for } t \geq i \tag{4.50}$$

$\square$

Now, using pure strategies, the best response of agent k is to select the action $a_{k\ell}(i)$ such that

$$
a_{k\ell}(i) = \begin{cases} 1, & \text{if } J_{k,i}^{\infty'}[a_{k\ell}(i) = 1|\boldsymbol{w}_{k,i-1}] < J_{k,i}^{\infty'}[a_{k\ell}(i) = 0|\boldsymbol{w}_{k,i-1}] \\ 0, & \text{otherwise} \end{cases} \tag{4.51}
$$

The following lemma shows how the best response rule depends on the benefit-cost ratio $\boldsymbol{\gamma}_k(i)$ and the communication cost c_k:

Lemma 4.1. *With Assumptions 1 and 2, the best response rule $f_k(\cdot)$ becomes*

$$a_{k\ell}(i) = \begin{cases} 1, & \text{if } \boldsymbol{\gamma}_k(i) \triangleq \frac{\boldsymbol{b}_k(i)}{c_k} > \frac{\chi_k}{\boldsymbol{\theta}_{\ell k}(i)} \\ 0, & \text{otherwise} \end{cases} \tag{4.52}$$

where

$$\chi_k \triangleq \frac{1 - \delta_k r_k}{\delta_k(1 - r_k)} \tag{4.53}$$

Proof. See Appendix A. □

We note that the resulting rule aligns the agents to achieve the Pareto optimal strategy: to share information when $\boldsymbol{\gamma}_k(i)$ is sufficiently large and not to share information when $\boldsymbol{\gamma}_k(i)$ is small.

4.4.3 Benefit Prediction

To compute the benefit-cost ratio $\boldsymbol{\gamma}_k(i) = \boldsymbol{b}_k(i)/c_k$, the agent still needs to know $\boldsymbol{b}_k(i)$ defined by (4.35), which depends on the quantities $\mathbb{E}[J_k^{\mathrm{act}}(a_{\ell k}(i) = 0)|\boldsymbol{w}_{k,i-1}]$ and $\mathbb{E}[J_k^{\mathrm{act}}(a_{\ell k}(i) = 1)|\boldsymbol{w}_{k,i-1}]$. It is common in the literature, as in [103, 121], to assume that agents have complete information about the payoff functions like the ones shown in Table 4.1. However, in the context of adaptive networks where agents have only access to data realizations and not to their statistical distributions, the payoffs are unknown and need to be estimated or predicted. For example, in our case, the convex combination $\alpha_k \boldsymbol{\psi}_{k,i} + (1-\alpha_k)\boldsymbol{\psi}_{\ell,i}$ is unknown for agent k before agent ℓ shares $\boldsymbol{\psi}_{\ell,i}$ with it. We now describe one way by which agent k can predict $\boldsymbol{b}_k(i)$; other ways are possible depending on how much information is available to the agent. Let us assume a special type of agents, which are called risk-taking [122]: agent k chooses $a_{k\ell}(i) = 1$ as long as the largest achievable benefit, denoted by $\bar{\boldsymbol{b}}_k(i)$, exceeds the threshold:

$$a_{k\ell}(i) = \begin{cases} 1, & \text{if } \frac{\bar{\boldsymbol{b}}_k(i)}{c_k} > \frac{\chi_k}{\boldsymbol{\theta}_{\ell k}(i)} \\ 0, & \text{otherwise} \end{cases} \tag{4.54}$$

Using (4.39), the largest achievable benefit $\bar{\boldsymbol{b}}_k(i)$ can be found by solving the following optimization problem:

$$
\begin{aligned}
\bar{\boldsymbol{b}}_k(i) &\triangleq \max_{\widetilde{\boldsymbol{\psi}}_{\ell,i}} \boldsymbol{b}_k(i) \\
&= \max_{\widetilde{\boldsymbol{\psi}}_{\ell,i}} \left\{ \mathbb{E}\left[\|\widetilde{\boldsymbol{\psi}}_{k,i}\|^2_{R_{u,k}} \Big| \boldsymbol{w}_{k,i-1} \right] - \mathbb{E}\left[\|\alpha_k \widetilde{\boldsymbol{\psi}}_{k,i} + (1-\alpha_k)\widetilde{\boldsymbol{\psi}}_{\ell,i}\|^2_{R_{u,k}} \Big| \boldsymbol{w}_{k,i-1} \right] \right\} \\
&= \mathbb{E}\left[\|\widetilde{\boldsymbol{\psi}}_{k,i}\|^2_{R_{u,k}} \Big| \boldsymbol{w}_{k,i-1} \right]
\end{aligned}
\tag{4.55}
$$

since the maximum occurs when

$$
\widetilde{\boldsymbol{\psi}}_{\ell,i} = -\frac{\alpha_k}{1-\alpha_k}\widetilde{\boldsymbol{\psi}}_{k,i}
\tag{4.56}
$$

Let us express the adaptation step (4.15) in terms of the estimation error as

$$
\widetilde{\boldsymbol{\psi}}_{k,i} = (I - \mu \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i})\widetilde{\boldsymbol{w}}_{k,i-1} - \mu \boldsymbol{u}_{k,i}^* \boldsymbol{v}_k(i)
\tag{4.57}
$$

To continue, we assume that the step-size μ is sufficiently small. Then,

$$
\begin{aligned}
&\mathbb{E}\left[\|\widetilde{\boldsymbol{\psi}}_{k,i}\|^2_{R_{u,k}} \Big| \boldsymbol{w}_{k,i-1} \right] \\
&= \mathbb{E}\left[\|(I - \mu \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i})\widetilde{\boldsymbol{w}}_{k,i-1} - \mu \boldsymbol{u}_{k,i}^* \boldsymbol{v}_k(i)\|^2_{R_{u,k}} \Big| \boldsymbol{w}_{k,i-1} \right] \\
&= \mathbb{E}\left[\widetilde{\boldsymbol{w}}_{k,i-1}^*(I - \mu \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}) R_{u,k}(I - \mu \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}) \times \widetilde{\boldsymbol{w}}_{k,i-1} \Big| \boldsymbol{w}_{k,i-1} \right] + \mu^2 \text{Tr}(R_{u,k}^2)\sigma_{v,k}^2 \\
&= \widetilde{\boldsymbol{w}}_{k,i-1}^* \Omega_k \widetilde{\boldsymbol{w}}_{k,i-1} + O(\mu^2)
\end{aligned}
\tag{4.58}
$$

where we are collecting terms that are second-order in the step-size into the factor $O(\mu^2)$,[4] and where we introduced $\Omega_k \triangleq R_{u,k}(I - 2\mu R_{u,k})$. We note that for sufficiently small step-sizes:

$$
\Omega_k' \triangleq R_{u,k}(I - \mu R_{u,k})^2 \approx R_{u,k}(I - 2\mu R_{u,k}) = \Omega_k
\tag{4.59}
$$

Therefore, each agent k can approximate $\bar{\boldsymbol{b}}_k(i)$ as

$$
\begin{aligned}
\bar{\boldsymbol{b}}_k(i) &= \widetilde{\boldsymbol{w}}_{k,i-1}^* \Omega_k' \widetilde{\boldsymbol{w}}_{k,i-1} \\
&= \widetilde{\boldsymbol{w}}_{k,i-1}^* R_{u,k}(I - \mu R_{u,k})^2 \widetilde{\boldsymbol{w}}_{k,i-1} \\
&= \|(I - \mu R_{u,k})\widetilde{\boldsymbol{w}}_{k,i-1}\|^2_{R_{u,k}}
\end{aligned}
\tag{4.60}
$$

---

[4]This approximation simplifies the algorithm construction. However, when we study the network performance later in (4.74) we shall keep the second-order terms.

4.4.4 Real-Time Implementation

Expression (4.60) is still not useful for agents because it requires knowledge of both $R_{u,k}$ and $\widetilde{\boldsymbol{w}}_{k,i-1}$. With regards to $R_{u,k}$, we can use the instantaneous approximation $R_{u,k} \approx \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}$ to get

$$\bar{\boldsymbol{b}}_k(i) \approx \widetilde{\boldsymbol{w}}_{k,i-1}^* \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i} (I - \mu \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i})^2 \widetilde{\boldsymbol{w}}_{k,i-1}$$
$$= (1 - \mu \|\boldsymbol{u}_{k,i}\|^2)^2 \widetilde{\boldsymbol{w}}_{k,i-1}^* \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i} \widetilde{\boldsymbol{w}}_{k,i-1} \tag{4.61}$$

With regards to $\widetilde{\boldsymbol{w}}_{k,i-1}$, we assume that agents use a moving-average filter as in [123] to approximate w^o iteratively as follows:

$$\widehat{\boldsymbol{w}}_{k,i}^o = (1 - \nu)\widehat{\boldsymbol{w}}_{k,i-1}^o + \nu\boldsymbol{\psi}_{k,i} \tag{4.62}$$

$$\widetilde{\boldsymbol{w}}_{k,i-1} \approx \widehat{\boldsymbol{w}}_{k,i}^o - \boldsymbol{w}_{k,i-1} \tag{4.63}$$

where $\nu \in (0,1)$ is a positive forgetting factor close to 0 to give higher weights on recent results. We summarize the operation of the resulting algorithm in the following listing.

4.5 Stability Analysis and Limiting Behavior

In this section, we study the stability of Algorithm 1 and its limiting performance after sufficiently long iterations. In order to pursue a mathematically tractable analysis, we assume that the maximum benefit $\bar{\boldsymbol{b}}_k(i)$ is estimated rather accurately by each agent k. That is, instead of the real-time implementation (4.61)–(4.63), we consider (4.60) that

$$\bar{\boldsymbol{b}}_k(i) = \|(I - \mu R_{u,k})\widetilde{\boldsymbol{w}}_{k,i-1}\|_{R_{u,k}}^2 \tag{4.64}$$

This consideration is motivated by taking the expectation of expression (4.61) given $\widetilde{\boldsymbol{w}}_{k,i-1}$:

$$\mathbb{E}\left[(1 - \mu\|\boldsymbol{u}_{k,i}\|^2)^2 \widetilde{\boldsymbol{w}}_{k,i-1}^* \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i} \widetilde{\boldsymbol{w}}_{k,i-1} \Big| \widetilde{\boldsymbol{w}}_{k,i-1}\right]$$
$$= \widetilde{\boldsymbol{w}}_{k,i-1}^*\left[R_{u,k} - 2\mu\mathbb{E}[\boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i} \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}] + O(\mu^2)\right]\widetilde{\boldsymbol{w}}_{k,i-1} \tag{4.65}$$

By subtracting (4.64) from (4.65), we can see that the difference between (4.64) and (4.65) is at least in the order of μ:

$$\widetilde{\boldsymbol{w}}_{k,i-1}^*\left[2\mu(R_{u,k}^2 - \mathbb{E}[\boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i} \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}]) + O(\mu^2)\right]\widetilde{\boldsymbol{w}}_{k,i-1} \tag{4.66}$$

113

Algorithm 1 Diffusion Strategy with an Adaptive Reputation Scheme.

Let $\{\boldsymbol{w}_{k,-1} = 0\}$ and $\{\theta_{k\ell}(-1) = 1\}$ for all k and ℓ. Define $\chi_k \triangleq \frac{1-\delta_k r_k}{\delta_k(1-r_k)}$.

loop

    Generate $\{\mathbf{1}_{k\ell}(i)\}$ for all k and ℓ.

    % **Stage 1 (Adaptation)**: For each k:

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} + \mu\boldsymbol{u}_{k,i}^*[\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{w}_{k,i-1}]$$

$$\bar{\boldsymbol{b}}_k(i) \approx (1 - \mu\|\boldsymbol{u}_{k,i}\|^2)^2 \widetilde{\boldsymbol{w}}_{k,i-1}^* \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i} \widetilde{\boldsymbol{w}}_{k,i-1}$$

$$\widehat{\boldsymbol{w}}_{k,i}^o = (1 - \nu)\widehat{\boldsymbol{w}}_{k,i-1}^o + \nu\boldsymbol{\psi}_{k,i}$$

$$\widetilde{\boldsymbol{w}}_{k,i-1} \approx \widehat{\boldsymbol{w}}_{k,i}^o - \boldsymbol{w}_{k,i-1}$$

    % **Stage 2 (Action Selection)**: For all k and ℓ,

    if $\mathbf{1}_{k\ell}(i) = 1$ **then**

$$\boldsymbol{a}_{k\ell}(i) = \begin{cases} 1, & \text{if } \frac{\bar{\boldsymbol{b}}_k(i)}{c_k} > \frac{\chi_k}{\boldsymbol{\theta}_{\ell k}(i)} \\ \\ 0, & \text{otherwise} \end{cases}$$

    else $\boldsymbol{a}_{k\ell}(i) = 0$.

    end if

    % **Stage 3 (Reputation Update)**: For all k and ℓ,

$$\boldsymbol{\theta}_{\ell k}(i + 1) = \mathbf{1}_{k\ell}(i) \cdot \max\{r_k\boldsymbol{\theta}_{\ell k}(i) + (1 - r_k)\boldsymbol{a}_{\ell k}(i), \varepsilon\}$$
$$+ (1 - \mathbf{1}_{k\ell}(i))\boldsymbol{\theta}_{\ell k}(i)$$

    % **Stage 4 (Combination)**: For all k,

$$\boldsymbol{w}_{k,i} = \alpha_k\boldsymbol{\psi}_{k,i} + (1 - \alpha_k) \sum_{\ell \in \mathcal{N}_k} \mathbf{1}_{k\ell}(i) \left[\boldsymbol{a}_{\ell k}(i)\boldsymbol{\psi}_{\ell,i} + (1 - \boldsymbol{a}_{\ell k}(i)) \boldsymbol{\psi}_{k,i}\right]$$

end loop

Therefore, for small enough μ the expected value of the realization given by (4.61) approaches the true value (4.64). The performance degradation from the real-time implementation error will be illustrated by numerical simulations in Sec. 4.6.

Under this condition, and for $\mu \ll 1$, we shall argue that the operation of each agent is stable in terms of both the estimation cost *and* the communication cost. Specifically, for the estimation cost, we will provide a condition on the step-size to ensure that the mean-square estimation error of each agent is asymptotically bounded. Using this result, we will further show that the communication cost for each agent k, and which is denoted by J_k^{com}, is upper bounded by a constant value that is unrelated to the transmission cost c_k. This result will be in contrast to the case of always cooperative agents where J_k^{com} will be seen to increase proportionally with c_k. This is because in our case, the probability of cooperation, $\text{Prob}\{\boldsymbol{a}_{k\ell}(i) = 1\}$, will be shown to be upper bounded by the ratio c^o/c_k for some constant c^o independent of c_k. It will then follow that when the communication becomes expensive (large c_k), self-interested agents using the adaptive reputation scheme will become unwilling to cooperate.

4.5.1   Estimation Performance

In conventional stability analysis for diffusion strategies, the combination coefficients are either assumed to be fixed, as done in [11,14,81,82,124,125], or their expectations conditioned on the estimates $\boldsymbol{w}_{k,i-1}$ are assumed to be constant, as in [80]. These conditions are not applicable to our scenario. When self-interested agents employ the nonlinear threshold-based action policy (4.54), the ATC diffusion algorithm (4.15) and (4.19) ends up involving a combination matrix whose entries are *dependent* on the estimates $\boldsymbol{w}_{k,i-1}$ (or the errors $\widetilde{\boldsymbol{w}}_{k,i-1}$). This fact introduces a new challenging aspect into the analysis of the distributed strategy. In the sequel, and in order to facilitate the stability and mean-square analysis of the learning process, we shall examine the performance of the agents in the network in three operating regions: the far-field region, the near-field region, and a region in between. We will show that the evolution of the estimation errors in these operating regions can be

described by the same network error recursion given further ahead in (4.99). Following this conclusion, we will then be able to use (4.99) to provide general statements about stability and performance in the three regions.

To begin with, referring to the listing in Algorithm 1, we start by noting that we write down the following error recursions for each agent k:

$$\widetilde{\boldsymbol{\psi}}_{k,i} = (I - \mu \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}) \widetilde{\boldsymbol{w}}_{k,i-1} - \mu \boldsymbol{u}_{k,i}^* \boldsymbol{v}_k(i) \tag{4.67}$$

$$\widetilde{\boldsymbol{w}}_{k,i} = \sum_{\ell \in \mathcal{N}_k} \boldsymbol{g}_{\ell k}(i) \widetilde{\boldsymbol{\psi}}_{\ell,i} \tag{4.68}$$

where the combination coefficients $\{\boldsymbol{g}_{\ell k}(i),\ \ell \in \mathcal{N}_k\}$ used in (4.68) are defined as follows and add up to one:

$$\boldsymbol{g}_{\ell k}(i) \triangleq (1 - \alpha_k) \mathbf{1}_{\ell k}(i) \boldsymbol{a}_{\ell k}(i) \geq 0 \tag{4.69}$$

$$\boldsymbol{g}_{kk}(i) \triangleq 1 - \sum_{\ell \in \mathcal{N}_k \backslash \{k\}} \boldsymbol{g}_{\ell k}(i) \geq 0 \tag{4.70}$$

Note that, in view of the pairing process, at most two of the coefficients $\{\boldsymbol{g}_{\ell k}(i)\}$ in (4.68) are nonzero in each time instant. The subsequent performance analysis will depend on evaluating the squared weighted norm of $\widetilde{\boldsymbol{\psi}}_{k,i}$ in (4.67), which is seen to be:

$$\begin{aligned}
\|\widetilde{\boldsymbol{\psi}}_{k,i}\|_{R_{u,k}}^2 = {} & \|(I - \mu \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}) \widetilde{\boldsymbol{w}}_{k,i-1}\|_{R_{u,k}}^2 \\
& + \mu^2 \|\boldsymbol{v}_k(i)\|^2 \mathrm{Tr}(\boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i} R_{u,k}) \\
& - \mu \widetilde{\boldsymbol{w}}_{k,i-1}^* (I - \mu \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}) R_{u,k} \boldsymbol{u}_{k,i}^* \boldsymbol{v}_k(i) \\
& - \mu \boldsymbol{v}_k^*(i) \boldsymbol{u}_{k,i} R_{u,k} (I - \mu \boldsymbol{u}_{k,i}^* \boldsymbol{u}_{k,i}) \widetilde{\boldsymbol{w}}_{k,i-1}
\end{aligned} \tag{4.71}$$

Now, from (4.68) we can use Jensen's inequality and the convexity of the squared norm to write

$$\|\widetilde{\boldsymbol{w}}_{k,i}\|_{R_{u,k}}^2 \leq \sum_{\ell \in \mathcal{N}_k} \boldsymbol{g}_{\ell k}(i) \|\widetilde{\boldsymbol{\psi}}_{\ell,i}\|_{R_{u,\ell}}^2 \tag{4.72}$$

so that, under expectation,

$$\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|_{R_{u,k}}^2 \leq \sum_{\ell \in \mathcal{N}_k} \mathbb{E}\left[\boldsymbol{g}_{\ell k}(i) \|\widetilde{\boldsymbol{\psi}}_{\ell,i}\|_{R_{u,\ell}}^2\right] \tag{4.73}$$

116

We note that $\boldsymbol{g}_{\ell k}(i)$ is a function of the random variables $\{\mathbf{1}_{\ell k}(i), \boldsymbol{a}_{\ell k}(i)\}$. The random pairing indicator $\mathbf{1}_{\ell k}(i)$ is independent of $\boldsymbol{u}_{\ell,i}$ and $\boldsymbol{v}_\ell(i)$. As for $\boldsymbol{a}_{\ell k}(i)$, which is determined by $\bar{\boldsymbol{b}}_\ell(i)$ and $\boldsymbol{\theta}_{k\ell}(i)$, we can see from expressions (4.64) and (4.45) that both $\bar{\boldsymbol{b}}_\ell(i)$ and $\boldsymbol{\theta}_{k\ell}(i)$ only depend on the past data prior to time i and therefore are independent of $\boldsymbol{u}_{\ell,i}$ and $\boldsymbol{v}_\ell(i)$. Consequently, $\boldsymbol{g}_{\ell k}(i)$ is independent of $\boldsymbol{u}_{\ell,i}$ and $\boldsymbol{v}_\ell(i)$, and we get

$$
\mathbb{E}\left[\boldsymbol{g}_{\ell k}(i)\|\widetilde{\boldsymbol{\psi}}_{\ell,i}\|^2_{R_{u,\ell}}\right]
$$
$$
= \mathbb{E}\left[\boldsymbol{g}_{\ell k}(i)\Big(\|(I - \mu\boldsymbol{u}^*_{\ell,i}\boldsymbol{u}_{\ell,i})\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2_{R_{u,\ell}} + \mu^2\|\boldsymbol{v}_\ell(i)\|^2\mathrm{Tr}(\boldsymbol{u}^*_{\ell,i}\boldsymbol{u}_{\ell,i}R_{u,\ell})\right.
$$
$$
- \mu\widetilde{\boldsymbol{w}}^*_{\ell,i-1}(I - \mu\boldsymbol{u}^*_{\ell,i}\boldsymbol{u}_{\ell,i})R_{u,\ell}\boldsymbol{u}^*_{\ell,i}\boldsymbol{v}_\ell(i)
$$
$$
\left.- \mu\boldsymbol{v}^*_\ell(i)\boldsymbol{u}_{\ell,i}R_{u,\ell}(I - \mu\boldsymbol{u}^*_{\ell,i}\boldsymbol{u}_{\ell,i})\widetilde{\boldsymbol{w}}_{\ell,i-1}\Big)\right]
$$
$$
= \mathbb{E}\left[\boldsymbol{g}_{\ell k}(i)\|(I - \mu\boldsymbol{u}^*_{\ell,i}\boldsymbol{u}_{\ell,i})\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2_{R_{u,\ell}}\right] + \mu^2\mathbb{E}\boldsymbol{g}_{\ell k}(i)\mathrm{Tr}(R^2_{u,\ell})\sigma^2_{v,\ell} \qquad (4.74)
$$

Using the fact that $\boldsymbol{u}_{\ell,i}$ is independent of $\boldsymbol{g}_{\ell k}(i)$ and $\widetilde{\boldsymbol{w}}_{\ell,i-1}$, we get

$$
\mathbb{E}\left[\boldsymbol{g}_{\ell k}(i)\|(I - \mu\boldsymbol{u}^*_{\ell,i}\boldsymbol{u}_{\ell,i})\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2_{R_{u,\ell}}\right]
$$
$$
= \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{g}_{\ell k}(i)\|(I - \mu\boldsymbol{u}^*_{\ell,i}\boldsymbol{u}_{\ell,i})\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2_{R_{u,\ell}}\Big|\boldsymbol{g}_{\ell k}(i), \widetilde{\boldsymbol{w}}_{\ell,i-1}\right]\right]
$$
$$
= \mathbb{E}\left[\boldsymbol{g}_{\ell k}(i)\widetilde{\boldsymbol{w}}^*_{\ell,i-1}\Sigma_\ell\widetilde{\boldsymbol{w}}_{\ell,i-1}\right] \qquad (4.75)
$$

where

$$
\Sigma_\ell \triangleq \mathbb{E}(I - \mu\boldsymbol{u}^*_{\ell,i}\boldsymbol{u}_{\ell,i})R_{u,\ell}(I - \mu\boldsymbol{u}^*_{\ell,i}\boldsymbol{u}_{\ell,i})
$$
$$
= R_{u,\ell} - 2\mu R^2_{u,\ell} + \mu^2\mathbb{E}(\boldsymbol{u}^*_{\ell,i}\boldsymbol{u}_{\ell,i}R_{u,\ell}\boldsymbol{u}^*_{\ell,i}\boldsymbol{u}_{\ell,i}) \qquad (4.76)
$$

If the regression data happens to be circular Gaussian, then a closed-form expression exists for the last fourth-order moment term in (4.76) in terms of $R_{u,\ell}$ [92]. We will not assume Gaussian data. Instead, we will assume that the fourth-order moment is bounded and that the network is operating in the slow adaptation regime with a sufficiently small step-size so that terms that depend on higher-order powers of μ can be ignored in comparison to other terms. Under this assumption, we replace (4.75) by:

$$
\mathbb{E}\left[\boldsymbol{g}_{\ell k}(i)\|(I - \mu\boldsymbol{u}^*_{\ell,i}\boldsymbol{u}_{\ell,i})\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2_{R_{u,\ell}}\right] = \mathbb{E}\left[\boldsymbol{g}_{\ell k}(i)\widetilde{\boldsymbol{w}}^*_{\ell,i-1}\Omega'_\ell\widetilde{\boldsymbol{w}}_{\ell,i-1}\right] = \mathbb{E}\left[\boldsymbol{g}_{\ell k}(i)\bar{\boldsymbol{b}}_\ell(i)\right] \qquad (4.77)
$$

where $\Omega'_\ell = R_{u,\ell}(I - \mu R_{u,\ell})^2$ from (4.59) and $\bar{b}_\ell(i) = \widetilde{w}^*_{\ell,i-1}\Omega'_\ell\widetilde{w}_{\ell,i-1}$ from (4.64). Note that

$$\Sigma_\ell - \Omega'_\ell = O(\mu^2) \tag{4.78}$$

Therefore, expression (4.74) becomes

$$\mathbb{E}\left[\boldsymbol{g}_{\ell k}(i)\|\widetilde{\boldsymbol{\psi}}_{\ell,i}\|^2_{R_{u,\ell}}\right] = \mathbb{E}\left[\boldsymbol{g}_{\ell k}(i)\bar{\boldsymbol{b}}_\ell(i)\right] + \mu^2\mathrm{Tr}(R^2_{u,\ell})\sigma^2_{v,\ell}\mathbb{E}\boldsymbol{g}_{\ell k}(i) \tag{4.79}$$

To continue, we introduce the following lemma which provides useful bounds for $\bar{\boldsymbol{b}}_k(i)$.

Lemma 4.2 (Bounds on $\bar{\boldsymbol{b}}_k(i)$)**.** *The values of $\bar{\boldsymbol{b}}_k(i)$ defined by (4.64) are lower and upper bounded by:*

$$\rho^2_{min}\|\widetilde{\boldsymbol{w}}_{k,i-1}\|^2_{R_{u,k}} \leq \bar{\boldsymbol{b}}_k(i) \leq \rho^2_{max}\|\widetilde{\boldsymbol{w}}_{k,i-1}\|^2_{R_{u,k}} \tag{4.80}$$

where

$$\rho_{max} \triangleq \max_{1 \leq k \leq N} \lambda_{max}(I - \mu R_{u,k}) \tag{4.81}$$

$$\rho_{min} \triangleq \min_{1 \leq k \leq N} \lambda_{min}(I - \mu R_{u,k}) \tag{4.82}$$

Proof. We introduce the eigendecomposition of the covariance matrix, $R_{u,k} \triangleq U_k\Lambda_kU^*_k$, where U_k is a unitary matrix and $\Lambda_k \triangleq \mathrm{diag}\{\lambda_{1,k}, ..., \lambda_{M,k}\}$ is a diagonal matrix with positive entries. Then, $R_{u,k}$ can be factored as

$$R_{u,k} = R^{\frac{1}{2}}_{u,k}R^{\frac{1}{2}}_{u,k} \tag{4.83}$$

where

$$R^{\frac{1}{2}}_{u,k} \triangleq U_k\Lambda^{\frac{1}{2}}_kU^*_k, \quad \Lambda^{\frac{1}{2}}_k \triangleq \mathrm{diag}\{\sqrt{\lambda_{1,k}}, ..., \sqrt{\lambda_{M,k}}\} \tag{4.84}$$

It is easy to verify that $R^{\frac{1}{2}}_{u,k}$ and $I - \mu R_{u,k}$ are commutable. Using this property, we obtain the following inequality:

$$\begin{aligned}
\widetilde{\boldsymbol{w}}^*_{k,i-1}\Omega'_k\widetilde{\boldsymbol{w}}_{k,i-1} &= \widetilde{\boldsymbol{w}}^*_{k,i-1}(I - \mu R_{u,k})R^{\frac{1}{2}}_{u,k}R^{\frac{1}{2}}_{u,k}(I - \mu R_{u,k})\widetilde{\boldsymbol{w}}_{k,i-1} \\
&= \widetilde{\boldsymbol{w}}^*_{k,i-1}R^{\frac{1}{2}}_{u,k}(I - \mu R_{u,k})^2R^{\frac{1}{2}}_{u,k}\widetilde{\boldsymbol{w}}_{k,i-1} \\
&\leq \lambda_{max}((I - \mu R_{u,k})^2)\widetilde{\boldsymbol{w}}^*_{k,i-1}R_{u,k}\widetilde{\boldsymbol{w}}_{k,i-1} \\
&\leq \rho^2_{max}\|\widetilde{\boldsymbol{w}}_{k,i-1}\|^2_{R_{u,k}} \tag{4.85}
\end{aligned}$$

We can obtain the lower bound for $\bar{\boldsymbol{b}}_k(i)$ by similar arguments. $\qquad\square$

Using the upper bound from Lemma 4.2, we have

$$\mathbb{E}\Big[\boldsymbol{g}_{\ell k}(i)\bar{\boldsymbol{b}}_{\ell}(i)\Big] \leq \rho_{\max}^2 \mathbb{E}\Big[\boldsymbol{g}_{\ell k}(i)\|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|_{R_{u,\ell}}^2\Big] \tag{4.86}$$

Then, from (4.79) and (4.86) we deduce from (4.73) that

$$\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|_{R_{u,k}}^2 \leq \rho_{\max}^2 \sum_{\ell \in \mathcal{N}_k} \mathbb{E}\Big[\boldsymbol{g}_{\ell k}(i)\|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|_{R_{u,\ell}}^2\Big] + \mu^2 \sum_{\ell \in \mathcal{N}_k} \mathrm{Tr}(R_{u,\ell}^2)\sigma_{v,\ell}^2 \mathbb{E}\boldsymbol{g}_{\ell k}(i) \tag{4.87}$$

From (4.69)–(4.70), it is ready to check that $\{\mathbb{E}\boldsymbol{g}_{\ell k}(i)\}$ are nonnegative and add up to 1. Therefore, the second term on the right-hand side of (4.87) is a convex combination and has the following upper bound:

$$\mu^2 \sum_{\ell \in \mathcal{N}_k} \mathrm{Tr}(R_{u,\ell}^2)\sigma_{v,\ell}^2 \mathbb{E}\boldsymbol{g}_{\ell k}(i) \leq \mu^2\kappa \tag{4.88}$$

where

$$\kappa \triangleq \max_{1 \leq k \leq N} \mathrm{Tr}(R_{u,k}^2)\sigma_{v,k}^2 \tag{4.89}$$

Therefore, we have

$$\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|_{R_{u,k}}^2 \leq \rho_{\max}^2 \sum_{\ell \in \mathcal{N}_k} \mathbb{E}\Big[\boldsymbol{g}_{\ell k}(i)\|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|_{R_{u,\ell}}^2\Big] + \mu^2\kappa \tag{4.90}$$

Since the combination coefficients $\{\boldsymbol{g}_{\ell k}(i)\}$ and the estimation errors $\{\|\widetilde{\boldsymbol{w}}_{k,i-1}\|_{R_{u,k}}^2\}$ are related in a nonlinear manner (as revealed by (4.54, (4.64), and (4.69)–(4.70)), the analysis of Algorithm 1 becomes challenging. To continue, we examine the behavior of the agents in the three regions of operation mentioned before.

During the initial stage of adaptation, agents are generally away from the target vector w^o and therefore have large estimation errors. We refer to this domain as the far-field region of operation, and we will characterize it by the condition:

$$\text{Far-Field:} \quad \mathrm{Prob}\left\{\|\widetilde{\boldsymbol{w}}_{k,i-1}\|_{R_{u,k}}^2 > \frac{c_k\chi_k}{\rho_{\min}^2\varepsilon}\right\} > \phi \tag{4.91}$$

where χ_k and $\rho_{\min}$ are defined in (4.53) and (4.82), respectively, and the parameter ϕ is close to 1 and in the range of $1 \geq \phi \gg 0$. That is, in the far-field regime, the weighted squared norm of estimation error $\|\widetilde{\boldsymbol{w}}_{k,i-1}\|_{R_{u,k}}^2$ exceeds a threshold with high probability. We note

119

that the far-field condition (4.91) can be more easily achieved when c_k is small. We also note that when the event in (4.91) holds with high-probability, then agent k is enticed to cooperate since:

$$\|\widetilde{\boldsymbol{w}}_{k,i-1}\|^2_{R_{u,k}} > \frac{c_k \chi_k}{\rho^2_{\min} \varepsilon} \overset{(a)}{\Longrightarrow} \bar{\boldsymbol{b}}_k(i) \boldsymbol{\theta}_{\ell k}(i) > c_k \chi_k$$

$$\overset{(b)}{\Longrightarrow} \boldsymbol{a}_{k\ell}(i) = 1 \tag{4.92}$$

where step (a) is by (4.80) and the fact $\boldsymbol{\theta}_{\ell k}(i) \in [\varepsilon, 1]$, and step (b) is by (4.54). Consequently, in the far-field region it holds with high likelihood that

$$\text{Prob}\{\boldsymbol{a}_{k\ell}(i) = 1\} \approx 1 \tag{4.93}$$

This approximation explains the phenomenon that with large estimation errors, agents are willing to cooperate and share estimates. We then say that under the far-field condition (4.91), the combination coefficients in (4.69)–(4.70) can be expressed as

$$\boldsymbol{g}_{\ell k}(i) = (1 - \alpha_k)\mathbf{1}_{\ell k}(i) \tag{4.94}$$

$$\boldsymbol{g}_{kk}(i) = 1 - \sum_{\ell \in \mathcal{N}_k \backslash \{k\}} \boldsymbol{g}_{\ell k}(i) = \alpha_k + (1 - \alpha_k)\mathbf{1}_{kk}(i) \tag{4.95}$$

with expectation values as follows:

$$g_{\ell k} \triangleq \mathbb{E}\boldsymbol{g}_{\ell k}(i) = (1 - \alpha_k)p_{\ell k}$$

$$g_{kk} \triangleq \mathbb{E}\boldsymbol{g}_{kk}(i) = \alpha_k + (1 - \alpha_k)p_{kk} \tag{4.96}$$

In this case, expression (4.90) becomes

$$\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2_{R_{u,k}} \leq \rho^2_{\max} \sum_{\ell \in \mathcal{N}_k} g_{\ell k} \mathbb{E}\|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2_{R_{u,\ell}} + \mu^2 \kappa \tag{4.97}$$

where we used the independence between $\mathbf{1}_{\ell k}(i)$ and $\|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2_{R_{u,\ell}}$. If we stack $\{\|\widetilde{\boldsymbol{w}}_{k,i}\|^2_{R_{u,k}}\}$ into a vector $\boldsymbol{\mathcal{X}}_i \triangleq \text{col}\{\|\widetilde{\boldsymbol{w}}_{1,i}\|^2_{R_{u,1}}, ..., \|\widetilde{\boldsymbol{w}}_{N,i}\|^2_{R_{u,N}}\}$ and collect the combination coefficients $\{g_{\ell k}\}$ into a left-stochastic matrix G, then we obtain the network error recursion as:

$$\mathbb{E}\boldsymbol{\mathcal{X}}_i \preceq \rho^2_{\max} G^T (\mathbb{E}\boldsymbol{\mathcal{X}}_{i-1}) + \mu^2 \kappa e \tag{4.98}$$

120

where $e \in 1_N$ is the vector of all ones and the notation $x \preceq y$ denotes that the components of vector x are less than or equal to the corresponding components of vector y. Taking the maximum norm from both sides and using the left-stochastic matrix property $\|G^T\|_\infty = 1$, we obtain:

$$\max_{1 \leq k \leq N} \mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2_{R_{u,k}} \triangleq \|\mathbb{E}\boldsymbol{\mathcal{X}}_i\|_\infty < \rho^2_{\max}\|\mathbb{E}\boldsymbol{\mathcal{X}}_{i-1}\|_\infty + \mu^2\kappa \tag{4.99}$$

Let us consider the long-term scenario $i \gg 1$. The far-field regime (4.91) is more likely to occur when c_k is small. Let us now examine the situation in which the communication cost is expensive ($c_k \gg 0$). In this case, the agents will operate in a near-field regime, which we characterize by the condition:

$$\text{Near-Field:} \quad \text{Prob}\left\{\|\widetilde{\boldsymbol{w}}_{k,i-1}\|^2_{R_{u,k}} < \frac{c_k\chi_k}{\rho^2_{\max}}\right\} > \phi \tag{4.100}$$

where $\rho_{\max}$ is defined in (4.81). We note that

$$\|\widetilde{\boldsymbol{w}}_{k,i-1}\|^2_{R_{u,k}} < \frac{c_k\chi_k}{\rho^2_{\max}} \implies \bar{\boldsymbol{b}}_k(i)\boldsymbol{\theta}_{\ell k}(i) < c_k\chi_k$$

$$\implies \boldsymbol{a}_{k\ell}(i) = 0 \tag{4.101}$$

In this regime, it then holds with high likelihood that

$$\text{Prob}\{\boldsymbol{a}_{k\ell}(i) = 0\} \approx 1 \tag{4.102}$$

and the combination coefficients in (4.69)–(4.70) then become:

$$\boldsymbol{g}_{kk}(i) = 1, \ \boldsymbol{g}_{\ell k}(i) = 0 \tag{4.103}$$

This means that agents will now be operating non-cooperatively since the benefit of sharing estimates is small relative to the expensive communication cost $c_k \gg 0$. Using similar arguments to (4.97)–(4.99), we arrive at the same network recursion (4.99) for this regime.

However, there exists a third possibility that for moderate values of c_k, agents operate at a region that does not belong to neither the far-field nor the near-field regimes. In this region, agents will be choosing to cooperate or not depending on their local conditions. To facilitate the presentation, let us introduce the notation $I_\mathbb{E}$ for an indicator function over

event $\mathbb{E}$ where $I_{\mathbb{E}} = 1$ if event $\mathbb{E}$ occurs and $I_{\mathbb{E}} = 0$ otherwise. Then, the action policy (4.54) can be rewritten as:

$$\boldsymbol{a}_{k\ell}(i) = I_{\bar{\boldsymbol{b}}_k(i)\boldsymbol{\theta}_{\ell k}(i) > c_k \chi_k} \tag{4.104}$$

From Lemma 4.2, we know that $\bar{\boldsymbol{b}}_k(i)$ is sandwiched within an interval of width $(\rho_{\max}^2 - \rho_{\min}^2)\|\widetilde{\boldsymbol{w}}_{k,i-1}\|_{R_{u,k}}^2$. As the error $\|\widetilde{\boldsymbol{w}}_{k,i-1}\|_{R_{u,k}}^2$ becomes smaller after sufficient iterations, the feasible region of $\bar{\boldsymbol{b}}_k(i)$ shrinks. Therefore, it is reasonable to assume that $\bar{\boldsymbol{b}}_k(i)$ becomes concentrated around its mean $\bar{b}_k(i) \triangleq \mathbb{E}\bar{\boldsymbol{b}}_k(i)$ for $i \gg 0$. It follows that we can approximate $\boldsymbol{a}_{k\ell}(i)$ by replacing $\bar{\boldsymbol{b}}_k(i)$ by $\bar{b}_k(i)$:

$$\boldsymbol{a}_{k\ell}(i) \approx I_{\bar{b}_k(i)\boldsymbol{\theta}_{\ell k}(i) > c_k \chi_k}, \quad \text{as} \ \ i \gg 1 \tag{4.105}$$

To continue, we further assume that after long iterations, which means that repeated interactions between agents have occurred for many times, the reputation scores gradually become stationary, i.e., we can model $\{\boldsymbol{\theta}_{\ell k}(i)\}$ as

$$\boldsymbol{\theta}_{\ell k}(i) = \theta_{\ell k} + \boldsymbol{n}_{\ell k}(i), \quad \text{as} \ \ i \gg 1 \tag{4.106}$$

where $\theta_{\ell k} \triangleq \mathbb{E}\boldsymbol{\theta}_{\ell k}(i)$ and $\boldsymbol{n}_{\ell k}(i)$ is a random disturbance which is independent and identically distributed (i.i.d.) over time i and assumed to be independent of all other random variables. Under this modeling, we can obtain the independence between $\boldsymbol{n}_{\ell k}(i)$ and $\widetilde{\boldsymbol{w}}_{\ell,i-1}$ and write that for $k \neq \ell$

$$
\begin{aligned}
\mathbb{E}\Big[\boldsymbol{g}_{\ell k}(i)\|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|_{R_{u,\ell}}^2\Big] &= \mathbb{E}\Big[(1 - \alpha_k)\boldsymbol{1}_{\ell k}(i)\boldsymbol{a}_{\ell k}(i)\|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|_{R_{u,\ell}}^2\Big] \\
&\approx (1 - \alpha_k)p_{\ell k}\mathbb{E}\Big[I_{\bar{b}_\ell(i)(\theta_{\ell k} + \boldsymbol{n}_{\ell k}(i)) > c_k\chi_k} \cdot \|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|_{R_{u,\ell}}^2\Big] \\
&= g'_{\ell k}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|_{R_{u,\ell}}^2 \tag{4.107}
\end{aligned}
$$

where

$$g'_{\ell k} \triangleq (1 - \alpha_k)p_{\ell k}\mathbb{E}\big[I_{\bar{b}_k(i)(\theta_{\ell k} + \boldsymbol{n}_{\ell k}(i)) > c_k\chi_k}\big] \geq 0 \tag{4.108}$$

For $\ell = k$, we can use similar arguments to write

$$
\begin{aligned}
\mathbb{E}\Big[\boldsymbol{g}_{kk}(i)\|\widetilde{\boldsymbol{w}}_{k,i-1}\|_{R_{u,k}}^2\Big] &= \mathbb{E}\Big[\Big(1 - \sum_{\ell \in \mathcal{N}_k \backslash \{k\}} (1 - \alpha_k)\boldsymbol{1}_{\ell k}(i)\boldsymbol{a}_{\ell k}(i)\Big)\|\widetilde{\boldsymbol{w}}_{k,i-1}\|_{R_{u,k}}^2\Big] \\
&\approx g'_{\ell k}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i-1}\|_{R_{u,\ell}}^2 \tag{4.109}
\end{aligned}
$$

where

$$g'_{kk} \triangleq 1 - \sum_{\ell \in \mathcal{N}_k \setminus \{k\}} g'_{\ell k} \geq 0 \tag{4.110}$$

Therefore, expression (4.90) becomes

$$\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2_{R_{u,k}} \leq \rho^2_{\max} \sum_{\ell \in \mathcal{N}_k} g'_{\ell k} \mathbb{E}\|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2_{R_{u,\ell}} + \mu^2 \kappa \tag{4.111}$$

Following similar arguments to (4.97)–(4.99), we again arrive at the same network recursion (4.99).

We therefore conclude that after long iterations, the estimation performance can be approximated by recursion (4.99) for general values of c_k. Consequently, sufficiently small step-sizes that satisfy the following condition guarantees the stability of the network error $\|\mathbb{E}\boldsymbol{\mathcal{X}}_i\|_\infty$:

$$\rho^2_{\max} < 1 \iff \mu < \frac{2}{\max\limits_{1 \leq k \leq N} \lambda_{\max}(R_{u,k})} \tag{4.112}$$

which leads to

$$\limsup_{i \to \infty} \|\mathbb{E}\boldsymbol{\mathcal{X}}_i\|_\infty \leq \frac{\mu^2 \kappa}{1 - \rho^2_{\max}} \tag{4.113}$$

Recalling that we are assuming sufficiently small μ, we have

$$\begin{aligned}
\rho^2_{\max} &= \max_{1 \leq k \leq N} \max_{1 \leq m \leq M} (1 - \mu \lambda_m(R_{u,k}))^2 \\
&\approx \max_{1 \leq k \leq N} \max_{1 \leq m \leq M} 1 - 2\mu \lambda_m(R_{u,k}) \\
&\triangleq 1 - 2\mu\beta
\end{aligned} \tag{4.114}$$

where

$$\beta \triangleq \min_{1 \leq k \leq N} \lambda_{\min}(R_{u,k}) \tag{4.115}$$

It is straightforward to verify that the bound on the right-hand side of (4.113) is upper-bounded by $\mu\kappa/2\beta$, which is $O(\mu)$. Consequently, we conclude that

$$\limsup_{i \to \infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2_{R_{u,k}} = O(\mu) \tag{4.116}$$

which establishes the mean-square stability of the network under the assumed conditions and for small step-sizes.

4.5.2 Expected Individual and Public Cost

In this section we assess the probability that agents will opt for cooperation after sufficient time has elapsed and the network has become stable. From (4.116), we know that after sufficient iterations, the MSE cost at each agent k is bounded, say, as

$$\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2_{R_{u,k}} \leq \eta\mu, \quad \text{for} \quad i \gg 1 \tag{4.117}$$

for some constant η. Based on this result, we can upper bound the cooperation rate of every agent k, defined as the probability that agent k would select $\boldsymbol{a}_{k\ell}(i) = 1$ for every pairing agent ℓ.

Theorem 4.1. *(Upper bound on cooperation rate) After sufficient iterations, the cooperation rate for each agent k is upper bounded by:*

$$\text{Prob}\{\boldsymbol{a}_{k\ell}(i) = 1\} \leq \min\left\{\frac{c^o}{c_k}, 1\right\}, \quad \text{for any} \quad c_k < \infty \tag{4.118}$$

where c^o is independent of c_k and defined as

$$c^o \triangleq \frac{\eta\mu\rho^2_{\max}}{\chi_{\min}}, \quad \chi_{\min} \triangleq \min_{1 \leq k \leq N} \chi_k \tag{4.119}$$

Proof. From (4.54), the cooperation rate of agent k is bounded by:

$$\text{Prob}\{\boldsymbol{a}_{k\ell}(i) = 1\} = \text{Prob}\left\{\bar{\boldsymbol{b}}_k(i)\boldsymbol{\theta}_{\ell k}(i) > c_k\chi_k\right\}$$

$$\leq \text{Prob}\left\{\bar{\boldsymbol{b}}_k(i) > c_k\chi_k\right\}$$

$$\leq \text{Prob}\left\{\|\widetilde{\boldsymbol{w}}_{k,i-1}\|^2_{R_{u,k}} > \frac{c_k\chi_k}{\rho^2_{\max}}\right\} \tag{4.120}$$

where we used the fact that $\boldsymbol{\theta}_{k\ell}(i) \leq 1$ and the upper bound on $\bar{\boldsymbol{b}}_k(i)$ from (4.80). Since $\|\widetilde{\boldsymbol{w}}_{k,i-1}\|^2_{R_{u,k}}$ is a nonnegative random variable with $\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i-1}\|^2_{R_{u,k}} \leq \eta\mu$, we can use Markov's inequality [126] to write

$$\text{Prob}\left\{\|\widetilde{\boldsymbol{w}}_{k,i-1}\|^2_{R_{u,k}} > \frac{c_k\chi_k}{\rho^2_{\max}}\right\} \leq \frac{\eta\mu\rho^2_{\max}}{c_k\chi_k} \leq \frac{c^o}{c_k} \tag{4.121}$$

Combining (4.120) and (4.121), we obtain that the cooperation rate for $i \gg 1$ is upper bounded by c^o/c_k. Using the fact that $\text{Prob}\{\boldsymbol{a}_{k\ell}(i) = 1\} \leq 1$, we get (4.118). $\qquad\square$
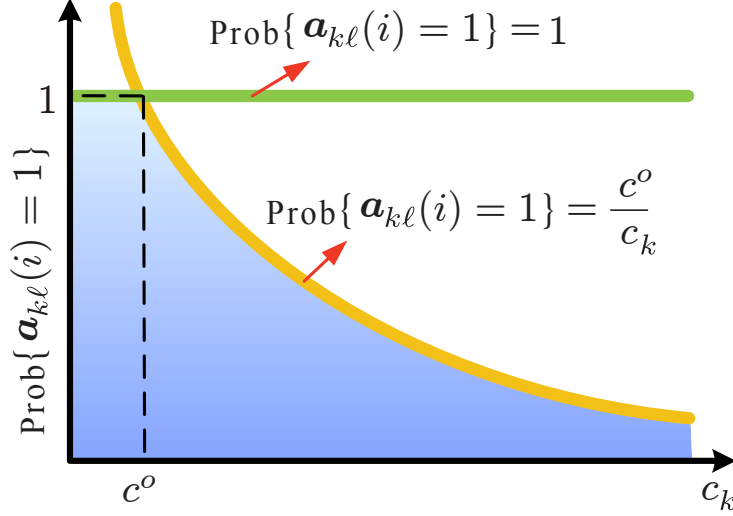
Figure 4.2: The feasible region of the probability of cooperation $\mathrm{Prob}\{\boldsymbol{a}_{k\ell}(i)=1\}$ for agent k.

As illustrated in Fig. 4.2, the feasible region of $\mathrm{Prob}\{\boldsymbol{a}_{k\ell}(i)=1\}$ is the intersection area of (4.118) and $0 \leq \mathrm{Prob}\{\boldsymbol{a}_{k\ell}(i)=1\} \leq 1$. We note that c^o has an order of μ. It describes the fact that when μ is small, the long term estimation errors reduce and agents have less willingness to cooperate and thus the cooperation rate $\mathrm{Prob}\{\boldsymbol{a}_{k\ell}(i)=1\}$ becomes low. Now, the expected communication cost for each agent k is

$$\mathbb{E}J_k^{\mathrm{com}} = \sum_{\ell \in \mathcal{N}_k \backslash \{k\}} \mathbb{E}[\mathbf{1}_{k\ell}(i)\boldsymbol{a}_{k\ell}(i)]c_k = \sum_{\ell \in \mathcal{N}_k \backslash \{k\}} p_{k\ell} \cdot \mathrm{Prob}\{\boldsymbol{a}_{k\ell}(i)=1\}c_k \tag{4.122}$$

where $\{k\}$ is excluded from the summation since there is no communication cost required for using own data. From Theorem 4.1, we know that when c_k is large, the expected communication cost has an upper bound which is independent of c_k, i.e., for $c_k \geq c^o$,

$$\mathbb{E}J_k^{\mathrm{com}} \leq \sum_{\ell \in \mathcal{N}_k \backslash \{k\}} p_{k\ell} \frac{c^o}{c_k} c_k = (1 - p_{kk})c^o \tag{4.123}$$

On the other hand, the expected estimation cost for each agent k for $i \gg 0$ is:

$$\mathbb{E}J_k^{\mathrm{est}} = \mathbb{E}|\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{w}_{k,i-1}|^2$$
$$= \mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i-1}\|_{R_{u,k}}^2 + \sigma_{v,k}^2$$
$$\leq \eta\mu + \sigma_{v,k}^2 \tag{4.124}$$

125

where we use (4.9) and the fact that $\widetilde{\boldsymbol{w}}_{k,i-1}$ is independent of $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$. It follows that for $i \gg 1$ the expected extended cost at each agent k is bounded by

$$\mathbb{E}\left[J_k(\boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i))\right] = \mathbb{E}J_k^{\text{est}} + \mathbb{E}J_k^{\text{com}} \le \eta\mu + \sigma_{v,k}^2 + (1 - p_{kk})c^o \tag{4.125}$$

If we now define the public cost as the accumulated expected extended cost over the network:

$$J^{\text{pub}} \triangleq \sum_{k=1}^{N} \mathbb{E}\left[J_k(\boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i))\right] \tag{4.126}$$

then

$$J^{\text{pub}} \le N\eta\mu + \sum_{k=1}^{N} \sigma_{v,k}^2 + c^o \sum_{k=1}^{N}(1 - p_{kk}) \tag{4.127}$$

which shows that J^{pub} is uniformly bounded by a constant value independent of c_k.

For comparison purposes, let us consider a uniform transmission cost $c_k = c$ and consider the case in (4.18) where the agents are controlled so that they always share estimates with their paired agents, i.e., $\boldsymbol{a}_{k\ell}(i) = 1$ for all k, ℓ, and i whenever $\boldsymbol{1}_{k\ell}(i) = 1$. Then, the random combination coefficients are the same as (4.94)–(4.95) defined in the far-field since agents always choose to cooperate. Let us denote the network mean-square-deviation (MSD) of cooperative agents by

$$\text{MSD}_{\text{coop}} = \lim_{i \to \infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \tag{4.128}$$

which has a closed-form expression provided in [80]. Therefore, we can characterize the performance of cooperative agents by MSD_{coop} and note that for $i \gg 0$, we have

$$\sum_{k=1}^{N} \mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|_{R_{u,k}}^2 \ge N\beta \cdot \text{MSD}_{\text{coop}} \tag{4.129}$$

where β is defined in (4.115). Consequently, after sufficient iterations, the expected public cost for cooperative agents becomes

$$J_{\text{coop}}^{\text{pub}} = \sum_{k=1}^{N} \left[\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|_{R_{u,k}}^2 + \sigma_{v,k}^2 + c \sum_{\ell \in \mathcal{N}_k \setminus \{k\}} \mathbb{E}\boldsymbol{1}_{k\ell}(i) \right]$$

$$\ge N\beta \cdot \text{MSD}_{\text{coop}} + \sum_{k=1}^{N} \sigma_{v,k}^2 + c \sum_{k=1}^{N}(1 - p_{kk}) \tag{4.130}$$
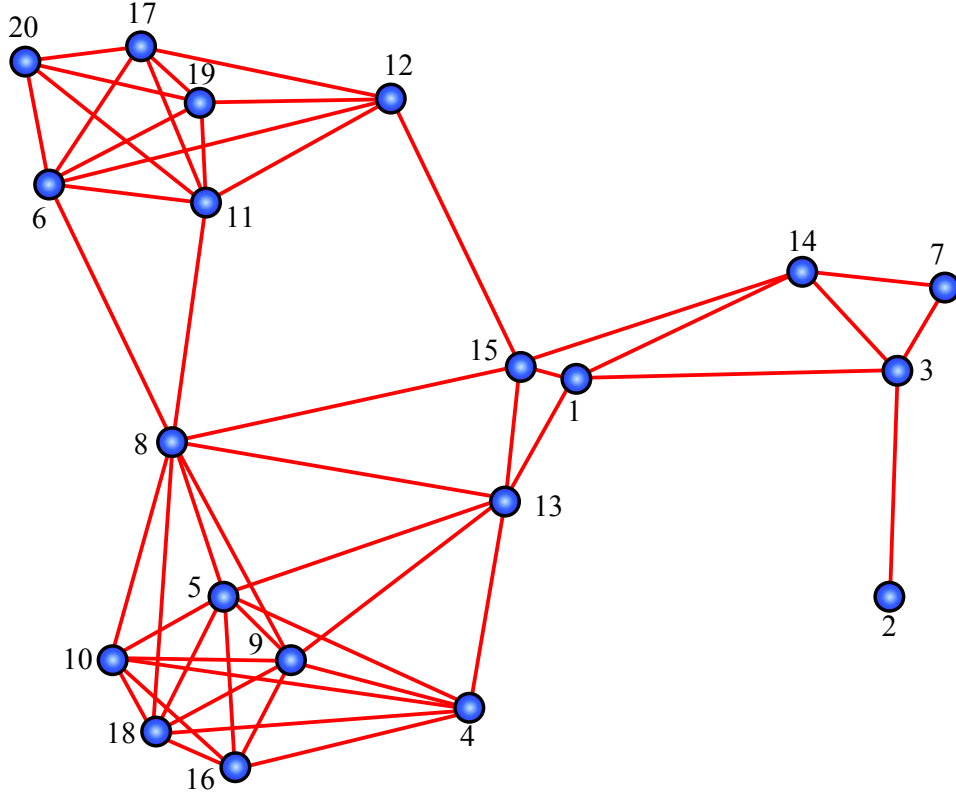
126

Figure 4.3: A network topology with $N = 20$ agents.

Comparing (4.127) with (4.130), we get $J^{\mathrm{pub}}_{\mathrm{coop}} \geq J^{\mathrm{pub}}$ whenever

$$N\beta \cdot \mathrm{MSD}_{\mathrm{coop}} + c \sum_{k=1}^{N}(1 - p_{kk}) \geq N\eta\mu + c^o \sum_{k=1}^{N}(1 - p_{kk})$$

$$\iff \quad c \geq c^o + \frac{N\left[\eta\mu - \beta \cdot \mathrm{MSD}_{\mathrm{coop}}\right]}{\sum\limits_{k=1}^{N}(1 - p_{kk})} \tag{4.131}$$

In other words, when the transmission cost c exceeds the above threshold, self-interested agents using the reputation protocol obtain a lower expected public cost than cooperative agents running the cooperative ATC strategy.
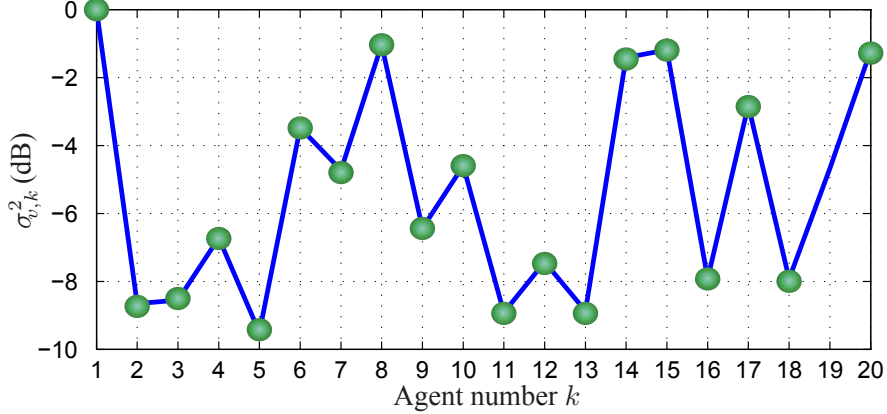
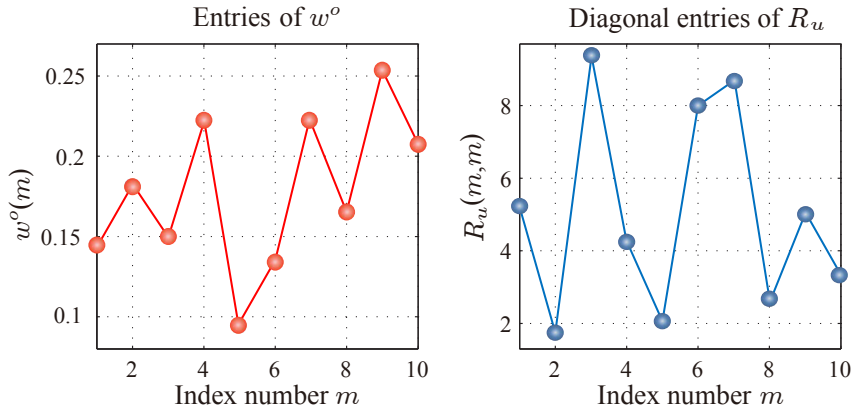Figure 4.4: The noise variance profile used in the simulations.

Figure 4.5: Entries of w^o and R_u used in the simulations.

4.6   Numerical Results

We consider a network with $N = 20$ agents. The network topology is shown in Fig. 4.3. The noise variance profile at the agents is shown in Fig. 4.4. In the simulations, we consider that the transmission cost $c_k = c$ is uniform and the matrix $R_{u,k} = R_u$ is uniform and diagonal. Figure 4.5 shows the entries of the target vector w^o of size $M = 10$ and the diagonal entries of R_u. We set the step-size at $\mu = 0.01$ and the combination weight at $\alpha_k = 0.5$ for all k. The parameters used in the reputation update rule are $\varepsilon = 0.1$ and the initial reputation scores $\boldsymbol{\theta}_{k,\ell}(0) = 1$ for all agents k and ℓ. The discount factor δ_k and the smoothing factor r_k

128
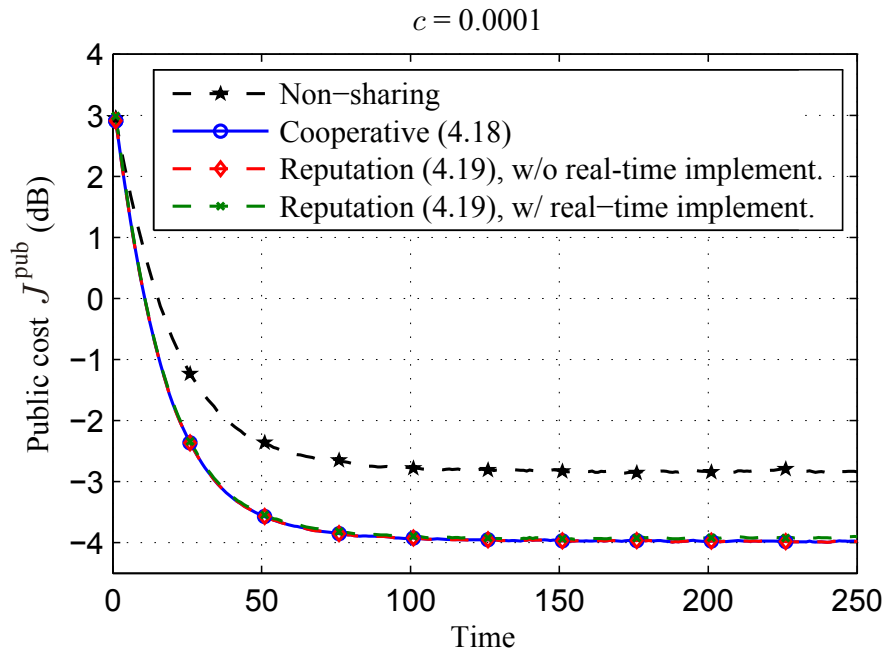
(a) Small communication cost $c = 0.0001$.
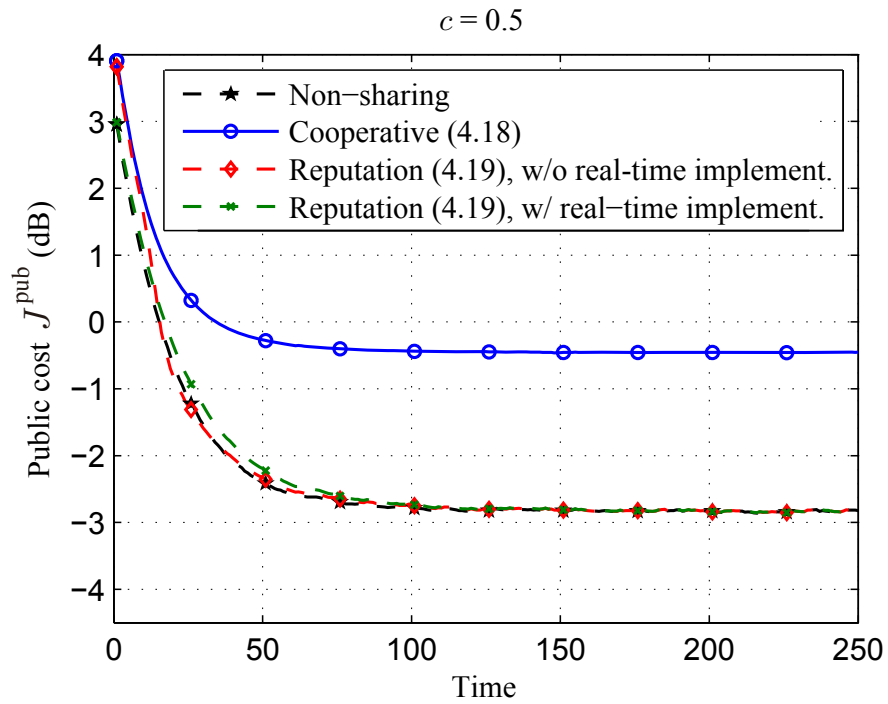
(b) Large communication cost $c = 0.5$.

Figure 4.6: Learning curve of public cost J^{pub} for small and large communication costs.
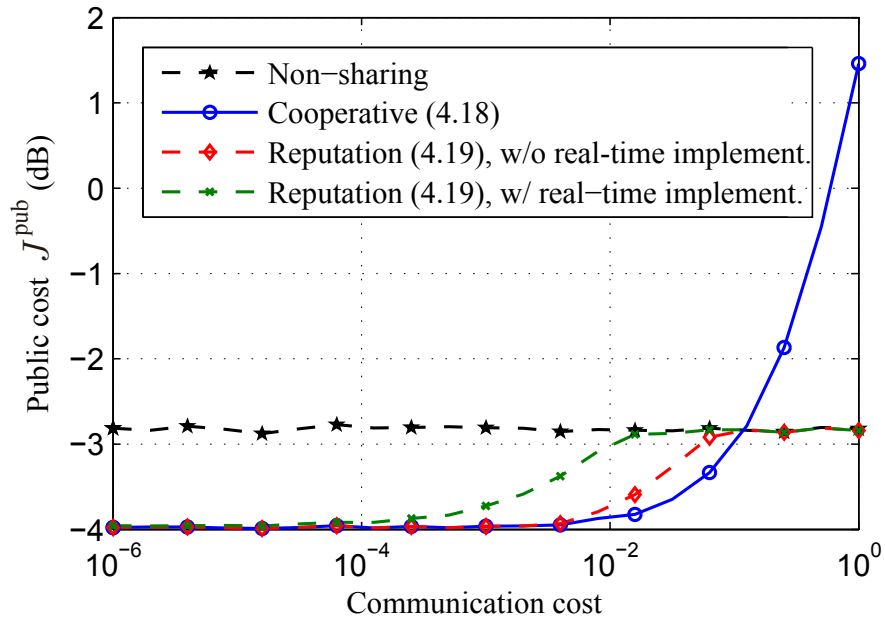
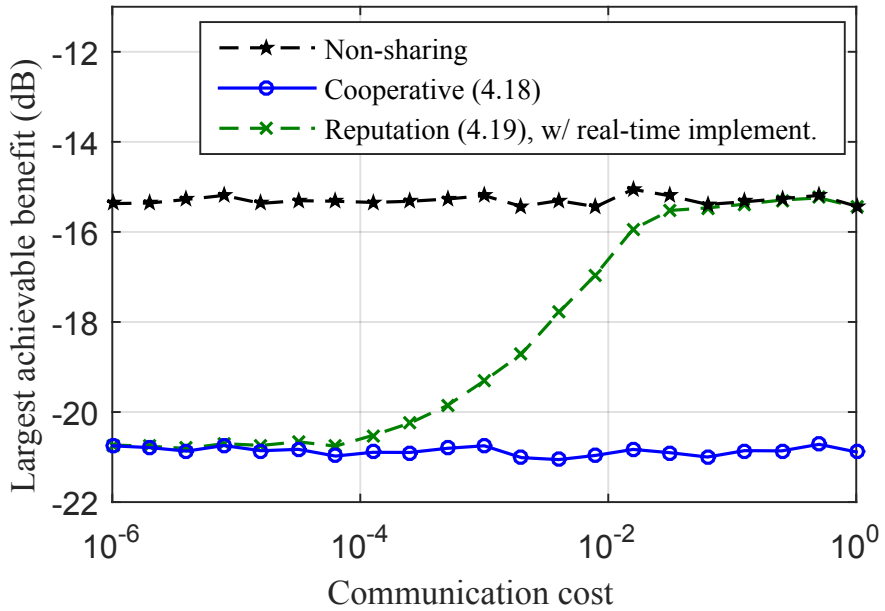Figure 4.7: Simulations of steady-state public costs J^{pub}.

Figure 4.8: Simulations of average largest achievable benefit $\bar{\boldsymbol{b}}_k(i)$ over agents.

for all k are set to 0.99 and 0.95, respectively, and the forgetting factor ν is set to 0.01.

In this simulation, we consider a distributed random-pairing mechanism as follows. At each time instant, each agent independently and uniformly generates a random continuous value from $[0, 1]$. Then, the agent holding the smallest value in the network, say agent k, is paired with the neighboring agent in $\mathcal{N}_k$ who has not been paired and holds the smallest value in $\mathcal{N}_k \setminus \{k\}$. Then, similar steps are followed by the agent who has not been paired and holds the second smallest value in the network. The random-pairing procedure continues until all agents complete these steps.

In Fig. 4.6, we simulate the learning curves of instantaneous public costs for small and large communication costs. It is seen that in both cases, using the proposed reputation protocol, the network of self-interested agents reaches the lowest public cost. Therefore, the network is efficient in terms of public cost. Furthermore, we note that in these two cases, there is only small difference between the performance of the reputation protocol using Algorithm 1 and the real-time implementation. To see the general effect of c, in Fig. 4.7 we simulate the public cost in steady-state versus the communication cost c. We observe that for large and small c, the reputation protocol performs as well as the non-sharing and the cooperative network, respectively. The only imperfection occurs around the switching region. Without real-time implementations, the reputation protocol has a small degradation in the range of $c \in [10^{-2}, 10^{-1}]$. While using real-time implementations (4.61)–(4.63), we can see that the degradation happens in a wider range of $c \in [10^{-4}, 10^{-1}]$. In Fig. 4.8, we simulate the network benefit defined as the largest achievable $\bar{\boldsymbol{b}}_k(i)$ averaged over all agents in steady state, i.e.,

$$b^{\mathrm{net}} \triangleq \lim_{i \to \infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}[\bar{\boldsymbol{b}}_k(i)] \tag{4.132}$$

where $\bar{\boldsymbol{b}}_k(i)$ follows the real-time implementation in (4.61)–(4.63). We note that the network benefits, b^{net}, for the non-sharing and cooperative cases are invariant for different communication costs since the behavior of agents is independent of c. Moreover, as expected in the form of (4.61), cooperative networks generally give smaller steady-state estimation errors and thus result in lower b^{net}. When the communication cost c increases, self-interested agents

following the proposed reputation protocol have less willingness to cooperate, and therefore have larger estimation errors and predict higher values of the benefit $\bar{\boldsymbol{b}}_k(i)$ in general.

4.7 Case Study: Distributed Spectrum Sensing

Cooperative spectrum sensing by secondary users (SUs) in a cognitive radio scenario can help avoid interference with transmissions by the primary user (PU) [127]. Spectrum sensing can be implemented either in a centralized manner [128] or decentralized manner [129] through coordination among the SUs. The latter approach exploits the spatial diversity of the SUs more fully and is scalable and robust, while the centralized approach is vulnerable to failure by the fusion center. The cooperative spectrum sensing problem generally involves a parameter estimation step. Various distributed strategies exist for the decentralized solution of estimation problems, most notably the consensus strategy [6, 7, 130, 131] and the diffusion strategy [10, 16, 132]. It has been shown in the prior work [21] that diffusion strategies have superior convergence, stability, and mean-square-error performance. For this reason, we shall employ diffusion adaptation to estimate the parameters of interest.

In collaborative spectrum sensing, it is not difficult to envision situations where some SUs may behave in a selfish manner and would participate in the sharing of information with other SUs only if this activity is beneficial to them. One example of such a scenario is studied in [133] where the SUs operate with the intention of maximizing their own transmission rates under the constraint of limited interference to the PUs. Other scenarios are studied in [114, 134, 135] using coalitional game formulations. In this paper, we examine the decentralized spectrum sensing problem in the presence of selfish SUs. We assume that the sharing of information among neighboring SUs entails some communication cost. In this way, each SU becomes interested in minimizing the error in estimating the parameter of interest to enable enhanced spectrum sensing (this objective favors cooperation) while reducing the cost of communicating with neighbors (this objective disfavors cooperation). We explained in [95] that under similar scenarios involving information-sharing games, the dominant strategy for each user is not to participate in the sharing of information. In order to address this inefficient

behavior, we embedded a reputation mechanism from [136] into the design of an adaptive collaborative process and developed a scheme that encourages users to cooperate. We show in this article how a similar design strategy can be developed for *online* cooperative spectrum sensing and leads to enhanced detection performance. In comparison to the framework in [95], here we to formulate a decentralized detection mechanism to guide the cooperation step.

4.7.1 System Model

We consider a network with N secondary users (SUs) and one primary user (PU). The frequency spectrum is divided into M sub-bands and the signal powers over these sub-bands are collected into a column vector w^o with nonnegative entries. The channels between the PU and the SUs are assumed to be frequency-selective and time-variant as follows. For each sub-band, and at each time instant i, the channel *power* gains from the PU to the $k-$th SU are represented by a $1 \times M$ vector $\boldsymbol{u}_{k,i} \in \mathbb{R}_+^{1 \times M}$ with non-negative entries. We assume that the channel information $u_{k,i}$, which is a realization for the random process $\boldsymbol{u}_{k,i}$ at time i, can be estimated through pilot signals during a training phase [137, 138]. During each i-th time interval, each SU k measures the received power that results from the aggregation of the signal powers in w^o multiplied by the channel power gains in $\boldsymbol{u}_{k,i}$. We denote the received power by $\boldsymbol{s}_k(i)$. This measurement is generally subject to noise and we write in a manner similar to [129]:

$$\boldsymbol{s}_k(i) = \boldsymbol{u}_{k,i} w^o + \boldsymbol{v}'_k(i) \tag{4.133}$$

where $\boldsymbol{v}'_k(i)$ combines the receiver and measurement noise sources and is assumed to have mean $\bar{v}_k$ and variance $\sigma^2_{v,k}$. We assume the random processes $\boldsymbol{u}_{k,i}$ and $\boldsymbol{v}'_\ell(j)$ are spatially and temporally independent over k, ℓ, i and j. To sense the spectrum, each SU solves a detection problem of the form:

$$\begin{cases} \mathcal{H}_0 : & w^o = 0 \\ \mathcal{H}_1 : & w^o = w^s \end{cases} \tag{4.134}$$

where $w^s \in \mathbb{R}_+^{M \times 1}$ represents the spectrum pattern that results from the presence of the PU. We assume that w^o varies slowly over time. We further assume that $\bar{v}_k$ is known by each SU, so that the data model can be centered as:

$$\boldsymbol{d}_k(i) \triangleq \boldsymbol{s}_k(i) - \bar{v}_k = \boldsymbol{u}_{k,i}w^o + \boldsymbol{v}_k(i) \tag{4.135}$$

where $\boldsymbol{v}_k(i) = \boldsymbol{v}'_k(i) - \bar{v}_k$ represents the centered zero-mean noise process.

We shall adopt a simple collaborative strategy for estimating w^o from the streaming data $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$. In the random-pairing model, when SUs k and ℓ are paired and SU ℓ agrees to collaborate with SU k, then SU k will update its estimate of the parameter vector w^o according to the strategy:

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} + \mu\boldsymbol{u}_{k,i}^T[\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i}\boldsymbol{w}_{k,i-1}] \tag{4.136}$$

$$\boldsymbol{w}_{k,i} = \alpha_k\boldsymbol{\psi}_{k,i} + (1 - \alpha_k)\boldsymbol{\psi}_{\ell,i} \tag{4.137}$$

where μ is a positive step-size factor, which is assumed to be sufficiently small to ensure mean-square stability. The second step (4.137) uses a coefficient $0 \leq \alpha_k \leq 1$ to combine the intermediate estimates of SUs k and ℓ. Using results from [16], it can be verified that a sufficiently small step-size μ ensures asymptotic mean stability of $\boldsymbol{w}_{k,i}$ in (4.136)–(4.137), i.e.,

$$\mathbb{E}\widetilde{\boldsymbol{w}}_{k,i} \to 0 \quad \text{as} \quad i \to \infty \tag{4.138}$$

in terms of the error vector $\widetilde{\boldsymbol{w}}_{k,i} \triangleq w^o - \boldsymbol{w}_{k,i}$. We could consider incorporating an additional projection step following (4.137) to ensure that all entries of $\boldsymbol{w}_{k,i}$ are non-negative. However, such a step generally leads to biased estimates for w^o. In this article, we continue with the unbiased solution that results from (4.136)–(4.137). The simulation results in the last section illustrate how this construction leads to good performance.

When SUs k and ℓ are paired together, we assume that they share the noise variances $\sigma_{v,k}^2$ and $\sigma_{v,\ell}^2$, and the channel realizations $u_{k,i}$ and $u_{\ell,i}$. Using this reference knowledge, SU ℓ will decide, according to the procedure described further ahead in (4.52)–(4.34), on whether to share its information $\boldsymbol{\psi}_{\ell,i}$ with SU k at time i, and vice-versa. The decision to

cooperate by either SU is based on each one of them evaluating a certain performance metric, described in the next section, and which reflects how well cooperation may enhance their detection accuracy against the communication cost. If SU ℓ decides not to share estimates, α_k in (4.137) is set to 1. For each SU ℓ, sharing the estimates $\boldsymbol{\psi}_{\ell,i}$ bears a known positive transmission cost c.

4.7.2 Performance Metric

4.7.2.1 Detection Performance

Let us denote by $\text{EMSE}_{k,i}$ the instantaneous excess-mean-square-error of SU k at time i conditioned on the known realization $u_{k,i}$. This quantity is defined as

$$\text{EMSE}_{k,i} \triangleq \mathbb{E}[|\boldsymbol{u}_{k,i}\widetilde{\boldsymbol{w}}_{k,i-1}|^2|\boldsymbol{u}_{k,i} = u_{k,i}] \tag{4.139}$$

which we rewrite as:

$$\text{EMSE}_{k,i} = \mathbb{E}|u_{k,i}\widetilde{\boldsymbol{w}}_{k,i-1}|^2 \geq 0 \tag{4.140}$$

Smaller values for $\text{EMSE}_{k,i}$ correspond to enhanced estimation accuracy. The analysis that follows explains how smaller values for $\text{EMSE}_{k,i}$ enhance the detection accuracy as well.

We reconsider the detection problem (4.134) by examining the statistics of the random variable $u_{k,i}\boldsymbol{w}_{k,i-1}$, which can be interpreted as an estimate for the received signal power. For small step-sizes and after sufficient iterations, the iterated $\boldsymbol{w}_{k,i-1}$ approaches w^o with a small mean-square error. We therefore approximate the mean of $u_{k,i}\boldsymbol{w}_{k,i-1}$ by

$$\mathbb{E}u_{k,i}\boldsymbol{w}_{k,i-1} \approx u_{k,i}w^o \tag{4.141}$$

Likewise, the variance of $u_{k,i}\boldsymbol{w}_{k,i-1}$ is approximated by:

$$\text{Var}(u_{k,i}\boldsymbol{w}_{k,i-1}) \triangleq \mathbb{E}|u_{k,i}\boldsymbol{w}_{k,i-1} - \mathbb{E}(u_{k,i}\boldsymbol{w}_{k,i-1})|^2 \approx \mathbb{E}|u_{k,i}\widetilde{\boldsymbol{w}}_{k,i-1}|^2 = \text{EMSE}_{k,i} \tag{4.142}$$

Thus, after a sufficient number iterations, we can replace the detection problem in (4.134) by

$$\begin{cases} \mathcal{H}_0 : & \mathbb{E}(u_{k,i}\boldsymbol{w}_{k,i-1}) \approx 0 \\ \mathcal{H}_1 : & \mathbb{E}(u_{k,i}\boldsymbol{w}_{k,i-1}) \approx u_{k,i}w^s \end{cases} \tag{4.143}$$

Now each SU k will decide on $\mathcal{H}_0$ or $\mathcal{H}_1$ by comparing the statistics $u_{k,i}\boldsymbol{w}_{k,i-1}$ with a threshold:

$$u_{k,i}\boldsymbol{w}_{k,i-1} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \eta_{k,i} \tag{4.144}$$

We consider the Neyman-Pearson test in which the threshold $\eta_{k,i}$ is chosen to maximize the detection probability under a constraint on the false-alarm probability, namely,

$$\begin{aligned}
\max_{\eta_{k,i}} \quad & P_{k,i}^{\mathrm{D}} \triangleq \Pr\{u_{k,i}\boldsymbol{w}_{k,i-1} \geq \eta_{k,i}; \mathcal{H}_1\} \\
\text{subject to} \quad & P_{k,i}^{\mathrm{FA}} \triangleq \Pr\{u_{k,i}\boldsymbol{w}_{k,i-1} \geq \eta_{k,i}; \mathcal{H}_0\} = \kappa
\end{aligned} \tag{4.145}$$

We note that other than the mean and variance, the statistics of $u_{k,i}\boldsymbol{w}_{k,i-1}$ are generally unknown. Therefore, the optimization problem (4.145) cannot be solved explicitly. To continue, we assume that the probability distribution of $u_{k,i}\boldsymbol{w}_{k,i-1}$ is symmetric around the mean under both $\mathcal{H}_0$ and $\mathcal{H}_1$. With this assumption, we can utilize Chebyshev's inequality to ensure that an upper bound on $P_{k,i}^{\mathrm{FA}}$ is smaller than κ. Thus, note that

$$\begin{aligned}
P_{k,i}^{\mathrm{FA}} &= \Pr\{u_{k,i}\boldsymbol{w}_{k,i-1} \geq \eta_{k,i}; \mathcal{H}_0\} \\
&= \frac{1}{2}\Pr\{|u_{k,i}\boldsymbol{w}_{k,i-1}| \geq \eta_{k,i}; \mathcal{H}_0\} \\
&\leq \frac{\mathrm{EMSE}_{k,i}}{2\eta_{k,i}^2}
\end{aligned} \tag{4.146}$$

Therefore, in order for (4.146) to be bounded by κ, the threshold should be selected to satisfy:

$$\eta_{k,i} \geq \sqrt{\frac{\mathrm{EMSE}_{k,i}}{2\kappa}} \tag{4.147}$$

Likewise, we maximize a lower bound on $P_{k,i}^{\mathrm{D}}$. Thus, note again that using the assumed symmetry of the distribution of $u_{k,i}\boldsymbol{w}_{k,i-1}$, we obtain

$$\begin{aligned}
\Pr\{u_{k,i}\boldsymbol{w}_{k,i-1} \leq \eta_{k,i}; \mathcal{H}_1\} &= \frac{1}{2}\Pr\{|u_{k,i}\boldsymbol{w}_{k,i-1} - u_{k,i}w^s| \geq u_{k,i}w^s - \eta_{k,i}; \mathcal{H}_1\} \\
&\leq \frac{\mathrm{EMSE}_{k,i}}{2(u_{k,i}w^s - \eta_{k,i})^2}
\end{aligned} \tag{4.148}$$

where we assume $u_{k,i}w^s > \eta_{k,i}$ in the second equality. This assumption is reasonable in most environments when the signal power $u_{k,i}w^s$ is sufficiently large, which means sufficiently high signal-to-noise ratio (SNR). Then,

$$
\begin{aligned}
P_{k,i}^{\mathrm{D}} &= \Pr\{u_{k,i}\boldsymbol{w}_{k,i} \geq \eta_{k,i}; \mathcal{H}_1\} \\
&= 1 - \Pr\{u_{k,i}\boldsymbol{w}_{k,i-1} \leq \eta_{k,i}; \mathcal{H}_1\} \\
&\geq 1 - \frac{\mathrm{EMSE}_{k,i}}{2(u_{k,i}w^s - \eta_{k,i})^2}
\end{aligned}
\tag{4.149}
$$

Therefore, the optimization problem (4.145) is approximated and replaced by

$$
\begin{aligned}
\max_{\eta_{k,i}} \quad & \bar{P}_{k,i}^{\mathrm{D}} \triangleq 1 - \frac{\mathrm{EMSE}_{k,i}}{2(u_{k,i}w^s - \eta_{k,i})^2} \\
\text{subject to} \quad & \eta_{k,i} \geq \sqrt{\frac{\mathrm{EMSE}_{k,i}}{2\kappa}}
\end{aligned}
\tag{4.150}
$$

Under the assumption $u_k w^s > \eta_{k,i}$, the objective function $\bar{P}_{k,i}^{\mathrm{D}}$ is monotonically decreasing with respect to $\eta_{k,i}$. Thus, the solution to (4.150) occurs at

$$
\eta_{k,i}^o = \sqrt{\frac{\mathrm{EMSE}_{k,i}}{2\kappa}}
\tag{4.151}
$$

and the resulting $\bar{P}_{k,i}^{\mathrm{D}}$ is

$$
\bar{P}_{k,i}^o = 1 - \frac{\mathrm{EMSE}_{k,i}}{2\left(u_{k,i}w^s - \sqrt{\frac{\mathrm{EMSE}_{k,i}}{2\kappa}}\right)^2}
\tag{4.152}
$$

It can be verified that $\bar{P}_{k,i}^o$ increases when $\mathrm{EMSE}_{k,i}$ decreases. It then follows that SUs should be motivated to cooperate in order to enhance the estimation accuracy and the detection probability. Therefore, given the current state estimate $\boldsymbol{w}_{k,i-1}$, the expression for $J_{k,i}$ in (4.22) is defined as follows:

$$
J_{k,i}(\boldsymbol{a}_k(i), \boldsymbol{a}_\ell(i)|\widetilde{\boldsymbol{w}}_{k,i-1}) \triangleq \mathbb{E}[\mathrm{EMSE}_{k,i+1}(\boldsymbol{a}_\ell(i)|\widetilde{\boldsymbol{w}}_{k,i-1})] + \boldsymbol{a}_k(i) \cdot c
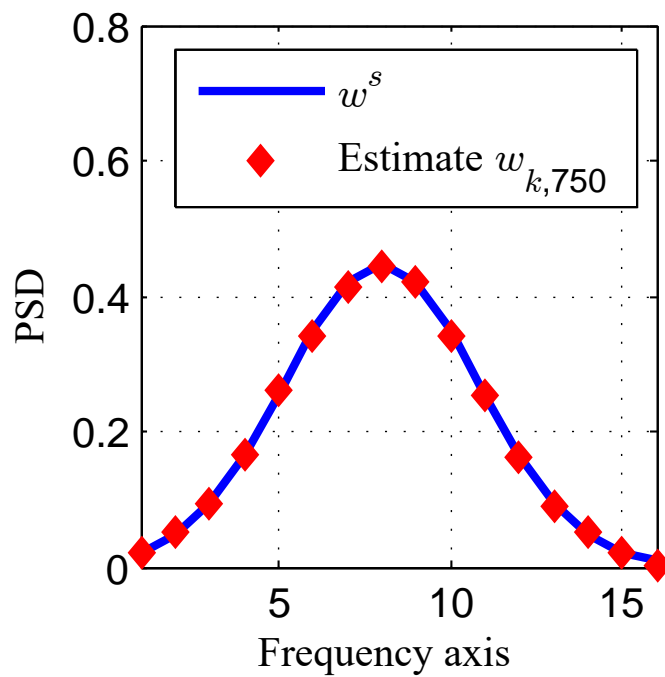\tag{4.153}
$$

4.7.3 Simulation Results

In the simulations, we assume there are $N = 15$ SUs. The locations of the PU and SUs are shown in Fig. 4.9(a). The PU is initially active at time $i = 0$, becomes inactive at $i = 1000$,
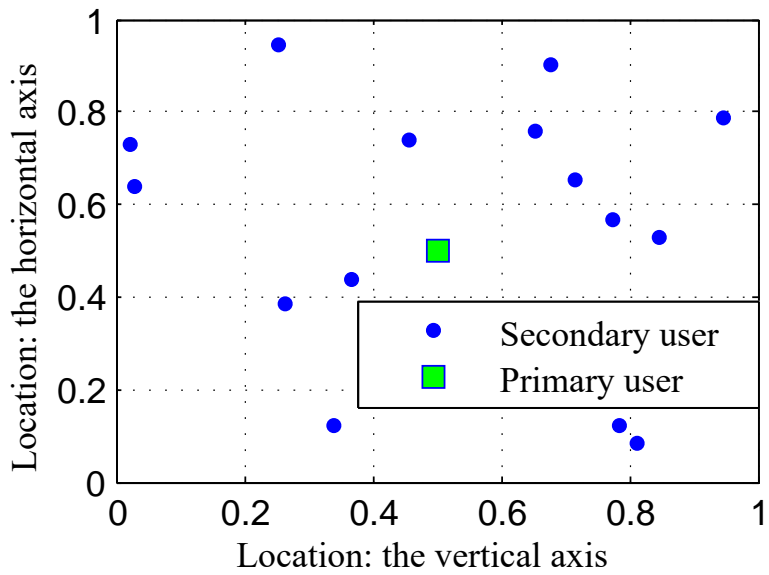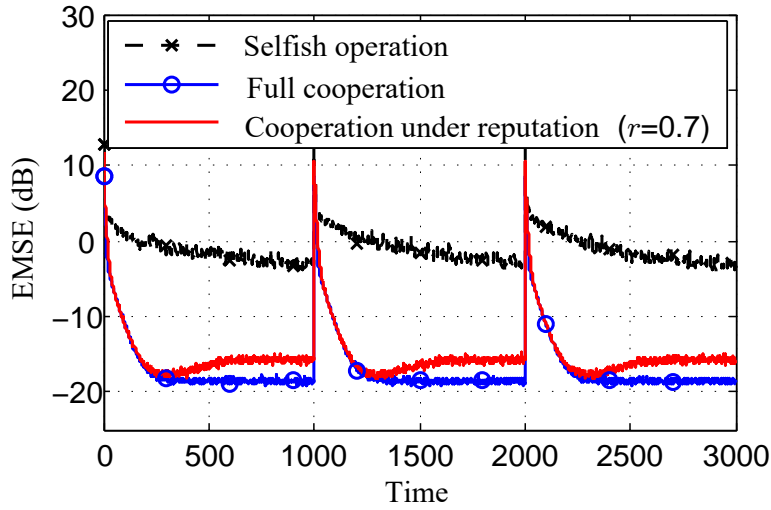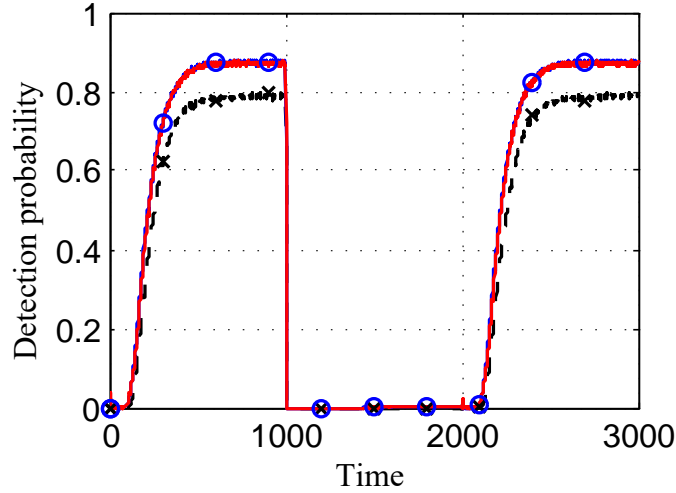
(a)

(b)

Figure 4.9: (a) Spatial distribution of the PU and SUs. (b) The spectrum pattern w^s and the average of estimates $\boldsymbol{w}_{k,750}$ over all SUs at $i = 750$.

138

(a) EMSE learning curve under different cooperative scenarios.

(b) Evolving probability of detection under different cooperative scenarios.

Figure 4.10: The PU is active during $i \in [0, 1000)$ and $i \geq 2000$.

and becomes active again at $i = 2000$. We assume that the SUs are randomly paired at each time instant. The spectrum pattern of w^s with $\|w^s\| = 1$ is represented by $M = 16$ samples and is illustrated in Fig. 4.9(b) along with the estimated $\boldsymbol{w}_{k,i}$ averaged across all SUs after sufficient iterations. The channel power gain $\boldsymbol{u}_{k,i}$ between the PU and each SU is assumed

to be a constant path loss gain with a random disturbance:

$$\boldsymbol{u}_{k,i} = g_{\mathrm{p},k}\mathbb{1} + \boldsymbol{g}_{k,i} \tag{4.154}$$

where the notation $\mathbb{1}_M$ denotes a vector with all its entries equal to one, $g_{\mathrm{p},k} = K_L \cdot (r_k/r_0)^{-2}$, $K_L = 0.1$ is a path-loss parameter, $r_0 = 1$ is a reference distance, and r_k is the distance between the PU and the k-th SU. The disturbance $\boldsymbol{g}_{k,i}$ is a zero-mean Gaussian random vector with covariance matrix $1.5I$. The measurement noise $\boldsymbol{v}'_k(i)$ is temporally white and spatially independent Gaussian distributed with mean $\bar{v}_k = 0.1$ and uniform variance $\sigma^2_{v,k} = \sigma^2_v = -10$ (dB). We set the step-size to $\mu = 0.005$, the transmission cost to $c = 10^{-6}$, the discounted parameter to $\delta = 0.99$, the minimum reputation $\epsilon = 0.1$, and the combination coefficients $\alpha_k = 1/2$ for all SUs when the shared estimates are available. All reputation scores are set to 1 at time $i = 0$ and discounted by $r = 0.7$.

In Fig. 4.10(a), the average EMSE over all SUs is simulated. Without the reputation scheme, the selfish SUs have no incentive to cooperate and their learning curve attains the worst EMSE performance. On the other hand, the reputation scheme encourages cooperation by selfish SUs and leads to better estimation performance. In Fig. 4.10(b), we simulate the detection performance in terms of the average $P^{\mathrm{D}}_{k,i}$ over all SUs. The threshold is determined by (4.151). The upper bound probability κ is 0.1.

4.8　Concluding Remarks

In this chapter, we studied the distributed information-sharing network in the presence of self-interested agents. We showed that without using any historical information to predict future actions, self-interested agents with bounded rationality become non-cooperative and refuse to share information. To entice them to cooperate, we developed an adaptive reputation protocol which turns the best response of self-interested agents into an action-choosing rule.

APPENDICES

4.A Proof of Lemma 4.1

We can represent the best response rule (4.51) by some mapping function, $f_k(\cdot)$, that maps the available realizations of $\boldsymbol{\theta}_{\ell k}(i)$ and $\boldsymbol{w}_{k,i-1}$ to $a_{k\ell}(i)$, i.e.,

$$a_{k\ell}(i) = f_k(\theta_{\ell k}(i), w_{k,i-1}) \tag{4.155}$$

We show the form of the resulting $f_k(\cdot)$ later in (4.52). For now, we note that construction (4.51) requires us to find the condition for

$$J_{k,i}^{\infty'}[a_{k\ell}(i) = 1 | \boldsymbol{w}_{k,i-1}] < J_{k,i}^{\infty'}[a_{k\ell}(i) = 0 | \boldsymbol{w}_{k,i-1}] \tag{4.156}$$

Using (4.49), this condition translates into requiring

$$J_{k,i}^{\infty'}[a_{k\ell}(i) = 1 | \boldsymbol{w}_{k,i-1}] - J_{k,i}^{\infty'}[a_{k\ell}(i) = 0 | \boldsymbol{w}_{k,i-1}]$$
$$= c_k + \sum_{t=i+1}^{\infty} \delta_k^{t-i} \triangle J_k^1(t) + \sum_{t=i+1}^{\infty} \delta_k^{t-i} \triangle J_k^2(t) < 0 \tag{4.157}$$

where, for simplicity, we introduced

$$\triangle J_k^1(t) \triangleq \mathbb{E}\left[\boldsymbol{a}_{k\ell}(t) c_k \Big| \boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i) = 1, \mathbb{K}_i\right]$$
$$- \mathbb{E}\left[\boldsymbol{a}_{k\ell}(t) c_k \Big| \boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i) = 0, \mathbb{K}_i\right] \tag{4.158}$$

$$\triangle J_k^2(t) \triangleq \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t)) \Big| \boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i) = 1, \mathbb{K}_i\right]$$
$$- \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t)) \Big| \boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i) = 0, \mathbb{K}_i\right] \tag{4.159}$$

Following similar argument to (4.28), we combine Assumptions 4.1 and 4.2 to conclude that

$$\boldsymbol{a}_{k\ell}(t) = a_{k\ell}(i), \quad \text{for } t \geq i \tag{4.160}$$

so that

$$\triangle J_k^1(t) = c_k \tag{4.161}$$

Similarly, using the assumption $\boldsymbol{w}_{k,t-1} = \boldsymbol{w}_{k,i-1}$ for $t \geq i$ from (4.27), we have

$$\mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t)) \middle| \boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i) = j, \mathbb{K}_i\right]$$

$$= \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t) = 1) \middle| \boldsymbol{w}_{k,t-1}, \boldsymbol{a}_{k\ell}(i) = j, \mathbb{K}_i\right] B(\boldsymbol{a}_{\ell k}(t) = 1)$$

$$+ \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t) = 0) \middle| \boldsymbol{w}_{k,t-1}, \boldsymbol{a}_{k\ell}(i) = j, \mathbb{K}_i\right] (1 - B(\boldsymbol{a}_{\ell k}(t) = 1)) \quad (4.162)$$

for $j = 0$ or 1. From (4.37) and (4.38), we can write

$$\mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t) = 0) \middle| \boldsymbol{w}_{k,t-1}, \boldsymbol{a}_{k\ell}(i) = j, \mathbb{K}_i\right] = \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t) = 0) \middle| \boldsymbol{w}_{k,t-1}\right] \quad (4.163)$$

$$\mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t) = 1) \middle| \boldsymbol{w}_{k,t-1}, \boldsymbol{a}_{k\ell}(i) = j, \mathbb{K}_i\right] = \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t) = 1) \middle| \boldsymbol{w}_{k,t-1}\right] \quad (4.164)$$

since these two conditional expectations depend only on $\boldsymbol{w}_{k,t-1}$. Therefore, expression (4.162) becomes:

$$\mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t)) \middle| \boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i) = j, \mathbb{K}_i\right]$$

$$= B(\boldsymbol{a}_{\ell k}(t) = 1) \cdot \left(\mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t) = 1) \middle| \boldsymbol{w}_{k,t-1}\right] - \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t) = 0) \middle| \boldsymbol{w}_{k,t-1}\right]\right)$$

$$+ \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t) = 0) \middle| \boldsymbol{w}_{k,t-1}\right]$$

$$= \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t) = 0) \middle| \boldsymbol{w}_{k,t-1}\right] - B(\boldsymbol{a}_{\ell k}(t) = 1)\boldsymbol{b}_k(t) \quad (4.165)$$

where we used the definition for $\boldsymbol{b}_k(t)$ from (4.35). We note that using (4.39) we have $\boldsymbol{b}_k(t) = \boldsymbol{b}_k(i)$ due to the assumption $\boldsymbol{w}_{k,t-1} = \boldsymbol{w}_{k,i-1}$. As a result, it follows that

$$\mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t)) \middle| \boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i) = j, \mathbb{K}_i\right]$$

$$= \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t) = 0) \middle| \boldsymbol{w}_{k,t-1}\right] - B(\boldsymbol{a}_{\ell k}(t) = 1)\boldsymbol{b}_k(i) \quad (4.166)$$

Let us denote by $\boldsymbol{\theta}_{k\ell}^j(t)$ the value of $\boldsymbol{\theta}_{k\ell}(t)$ at time t if $\boldsymbol{a}_{k\ell}(i) = j$ is selected at time i. We utilize the assumption $\boldsymbol{\theta}_{\ell k}(t) = \boldsymbol{\theta}_{\ell k}(i)$ to rewrite $B(\boldsymbol{a}_{\ell k}(t) = 1)$ as

$$B(\boldsymbol{a}_{\ell k}(t) = 1) = \boldsymbol{\theta}_{k\ell}^j(t)\boldsymbol{\theta}_{\ell k}(i) \quad (4.167)$$

It then follows that

$$\mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t)) \middle| \boldsymbol{w}_{k,i-1}, \boldsymbol{a}_{k\ell}(i) = j, \mathbb{K}_i\right]$$

$$= \mathbb{E}\left[J_k^{\text{act}}(\boldsymbol{a}_{\ell k}(t) = 0) \middle| \boldsymbol{w}_{k,t-1}\right] - \boldsymbol{\theta}_{k\ell}^j(t)\boldsymbol{\theta}_{\ell k}(i)\boldsymbol{b}_k(i) \quad (4.168)$$

142

Therefore, using (4.159) we get

$$\triangle J_k^2(t) = [\boldsymbol{\theta}_{k\ell}^0(t) - \boldsymbol{\theta}_{k\ell}^1(t)]\boldsymbol{\theta}_{\ell k}(i)\boldsymbol{b}_k(i) \tag{4.169}$$

Now, we recall that following Assumption 4.1 and the considered scenario $\mathbf{1}_{k\ell}(i) = 1$, we have $\mathbf{1}_{k\ell}(t) = \mathbf{1}_{k\ell}(i) = 1$ for $t \geq i$. As a result, the reputation update in (4.45) can be approximated by expression (4.43) for sufficiently small ε. Then, under (4.160), the future reputation scores $\boldsymbol{\theta}_{k\ell}^0(t)$ and $\boldsymbol{\theta}_{k\ell}^1(t)$ are given by:

$$\boldsymbol{\theta}_{k\ell}^0(t) = \boldsymbol{\theta}_{k\ell}(i)r_k^{t-i} \tag{4.170}$$

$$\boldsymbol{\theta}_{k\ell}^1(t) = \boldsymbol{\theta}_{k\ell}(i)r_k^{t-i} + (1 - r_k)\sum_{q=0}^{t-i-1} r_k^q = \boldsymbol{\theta}_{k\ell}(i)r_k^{t-i} + (1 - r_k^{t-i}) \tag{4.171}$$

Therefore, expression (4.169) becomes

$$\triangle J_k^2(t) = -(1 - r_k^{t-i})\boldsymbol{\theta}_{\ell k}(i)\boldsymbol{b}_k(i), \quad \text{for} \ \ t > i \tag{4.172}$$

Using (4.161) and (4.172), agent k then chooses $\boldsymbol{a}_{k\ell}(i) = 1$ if

$$c_k + \sum_{t=i+1}^{\infty} \delta_k^{t-i} c_k - \sum_{t=i+1}^{\infty} \delta_k^{t-i}(1 - r_k^{t-i})\boldsymbol{\theta}_{\ell k}(i)\boldsymbol{b}_k(i) < 0$$

$$\Longleftrightarrow \quad \sum_{t=i}^{\infty} \delta_k^{t-i} c_k < \sum_{t=i+1}^{\infty} \delta_k^{t-i}(1 - r_k^{t-i})\boldsymbol{\theta}_{\ell k}(i)\boldsymbol{b}_k(i)$$

$$\Longleftrightarrow \quad \frac{c_k}{1 - \delta_k} < \boldsymbol{\theta}_{\ell k}(i)\boldsymbol{b}_k(i)\delta_k \cdot \left(\frac{1}{1 - \delta_k} - \frac{r_k}{1 - \delta_k r_k} \right)$$

$$\Longleftrightarrow \quad \boldsymbol{\gamma}_k(i) \triangleq \frac{\boldsymbol{b}_k(i)}{c_k} > \frac{\chi_k}{\boldsymbol{\theta}_{\ell k}(i)} \tag{4.173}$$

where we introduced

$$\chi_k \triangleq \frac{1 - \delta_k r_k}{\delta_k(1 - r_k)} \tag{4.174}$$

The best response rule $f_k(\cdot)$ therefore becomes

$$a_{k\ell}(i) = \begin{cases} 1, & \text{if } \boldsymbol{\gamma}_k(i) \triangleq \frac{\boldsymbol{b}_k(i)}{c_k} > \frac{\chi_k}{\boldsymbol{\theta}_{\ell k}(i)} \\ 0, & \text{otherwise} \end{cases} \tag{4.175}$$

CHAPTER 5

Cluster Formation with Selfish Agents

As an alternative to the reputation protocols considered in Chapter 4, we go beyond interactions among single agents and develop a clustering technique to entice cooperation by clusters of agents. We divide the operation of the network into two stages: a cluster formation stage and an information sharing stage. During cluster formation, agents evaluate a long-term combined cost function and decide on whether to cooperate or not with other agents. During the subsequent information sharing phase, agents share and process information over their sub-networks. Simulations illustrate how the clustering technique enhances the mean-square-error performance of the agents over noncooperative processing.

5.1   Introduction

We now develop a clustering technique as an alternative to reputation protocols to entice cooperation by self-interested agents. In this method, agents are allowed to decide with whom to cluster and share information. The clustering concept is widely studied in the social sciences and game theory (e.g., [111, 114, 134, 139, 140]). It enables agents to drive their cooperative behavior by selecting their partners according to whether they can help them reduce their utility costs. For adaptive networks, the challenge is to select utility functions that can drive the clustering operation. Recent results on the performance of adaptive networks [125] can be exploited to great effect for this purpose.

In the formulation studied in this chapter, we divide the operation of the network into two stages. The first stage is the cluster formation phase and the second stage is the information sharing and processing phase. During cluster formation, agents meet randomly in pairs
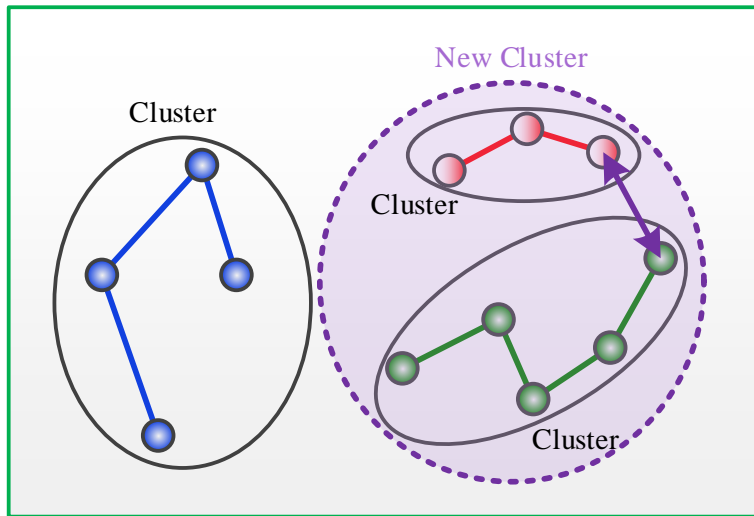
(a) Cluster formation.
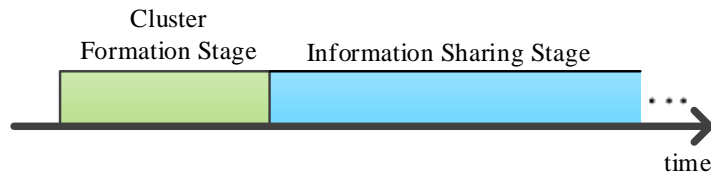
(b) Timeline.

Figure 5.1: (a) Self-interested agents establish new links to form a larger cluster. (b) Timeline illustrates two stages of cluster formation and information sharing.

following a random pairing protocol [9], similar to the random-pairing model in (4.11)–(4.13). Based on some prior reference knowledge about mutual clusters, each agent then evaluates the expected cost of its possible actions and decides on whether to propose cooperation to the other agent. If both agents agree on cooperation, then they establish a link and become part of the same larger cluster. We illustrate cluster formation and the timeline involved in Fig. 5.1. Once clusters are formed, the agents can then proceed to solve the estimation task in a distributed manner by cooperating within their sub-networks. We assume there exist harsh punishments to prevent agents from deviating from the agreement of information sharing, such as to permanently isolate the deviant agents.

5.2 Information Sharing Structure

5.2.1 Reference Knowledge and Transmission Cost

Consider a network with N self-interested agents. During the cluster formation stage, pairs of agents, say, agents k and ℓ, randomly meet and exchange some preliminary knowledge, denoted by $\mathbb{K}_k$ and $\mathbb{K}_\ell$, respectively. Based on $\mathbb{K}_k$ and $\mathbb{K}_\ell$, the agents decide on whether they want to become part of the same cluster. Membership in the same cluster implies that the agents would agree to cooperate with each other during the information sharing stage. During this second phase, agents share information denoted by $\mathbb{I}_{k,i}$ and $\mathbb{I}_{\ell,i}$ at time i. Obviously, the sharing of the information $\mathbb{I}_{k,i}$ with agent ℓ bears some transmission cost for agent k, which is denoted by $c_{k\ell} > 0$ and assumed to be known by agent k. Likewise, $c_{\ell k} > 0$ represents the cost for agent ℓ when it shares information with agent k. In the subscripts ℓk, the first letter represents the source agent and the second letter represents the destination agent. We set $c_{kk} = 0$.

5.2.2 Agreement to Cluster

When agent k first meets agent ℓ during the cluster formation stage, agent k chooses an action $\boldsymbol{a}_{k\ell} \in \{0, 1\}$ based on their shared preliminary knowledge $\mathbb{K}_k$ and $\mathbb{K}_\ell$ (as described further ahead in Sec. 5.3.) The action $\boldsymbol{a}_{k\ell} = 1$ means that agent k proposes to agent ℓ that they become part of the same cluster, and the action $\boldsymbol{a}_{k\ell} = 0$ means that agent k does not want to cluster with agent ℓ. Agent ℓ's action, $\boldsymbol{a}_{\ell k} \in \{0, 1\}$, is defined in a similar manner. The agreement to cluster must be consensual, i.e., both agents need to propose $\boldsymbol{a}_{k\ell} = 1$ and $\boldsymbol{a}_{\ell k} = 1$. This situation can be represented by the indicator value defined by:

$$\boldsymbol{I}_{k\ell} = \boldsymbol{I}_{\ell k} \triangleq \boldsymbol{a}_{k\ell} \cdot \boldsymbol{a}_{\ell k} \tag{5.1}$$

Thus, $\boldsymbol{I}_{k\ell} = 1$ means that both agents have agreed to become part of the same cluster so that agent k will share information $\mathbb{I}_{k,i}$ with agent ℓ during the information sharing stage, and vice-versa. On the other hand, $\boldsymbol{I}_{k\ell} = 0$ means that agents k and ℓ do not wish to cluster. We set $\boldsymbol{I}_{kk} = 1$.
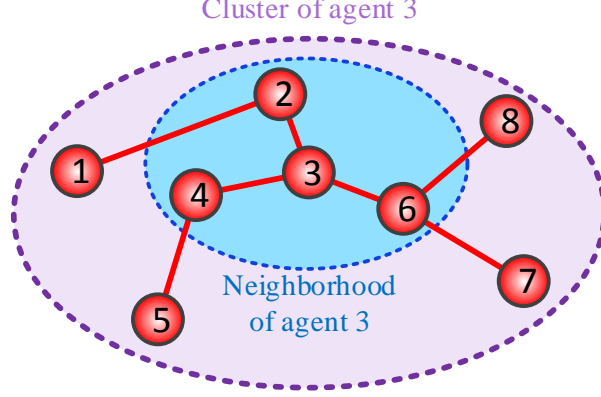
Figure 5.2: The neighborhood of agent 3 is $\mathcal{N}_3 = \{2,3,4,6\}$ and the cluster of agent 3 is $\mathcal{C}_3 = \{1,2,3,4,5,6,7,8\}$.

Table 5.1: Cost values for all four combinations of actions by the agents.

Agent k / Agent ℓ	$\boldsymbol{a}_{k\ell} = 0$	$\boldsymbol{a}_{k\ell} = 1$
$\boldsymbol{a}_{\ell k} = 0$	$\mathrm{MSD}_k(\mathcal{C}_k) + \beta_k \sum_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq}$ $\mathrm{MSD}_\ell(\mathcal{C}_\ell) + \beta_\ell \sum_{q \in \mathcal{N}_\ell} \boldsymbol{I}_{\ell q} c_{\ell q}$	$\mathrm{MSD}_k(\mathcal{C}_k) + \beta_k \sum_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq}$ $\mathrm{MSD}_\ell(\mathcal{C}_\ell) + \beta_\ell \sum_{q \in \mathcal{N}_\ell} \boldsymbol{I}_{\ell q} c_{\ell q}$
$\boldsymbol{a}_{\ell k} = 1$	$\mathrm{MSD}_k(\mathcal{C}_k) + \beta_k \sum_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq}$ $\mathrm{MSD}_\ell(\mathcal{C}_\ell) + \beta_\ell \sum_{q \in \mathcal{N}_\ell} \boldsymbol{I}_{\ell q} c_{\ell q}$	$\mathrm{MSD}_k(\mathcal{C}_k \cup \mathcal{C}_\ell) + \beta_k \sum_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq} + \beta_k c_{k\ell}$ $\mathrm{MSD}_\ell(\mathcal{C}_k \cup \mathcal{C}_\ell) + \beta_\ell \sum_{q \in \mathcal{N}_\ell} \boldsymbol{I}_{\ell q} c_{\ell q} + \beta_\ell c_{\ell k}$

5.2.3 Two-Stage Operations

Consider that each agent k seeks to solve a distributed estimation task, such as estimating and tracking some parameter vector of interest, which we denote by $w^o \in \mathbb{C}^{M \times 1}$. We assume that agents have access to the data $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$ related via the linear regression model in (4.4). We now describe a two-stage operation to facilitate the cluster formation process.

During the cluster formation stage, the cluster dynamics is evolving and, therefore, $\mathcal{C}_k$ is dependent on time during this phase. When two agents k and ℓ first meet randomly at some time i, the reference knowledge $\mathbb{K}_k$ and $\mathbb{K}_\ell$ that they share is assumed to consist of the

agents that belong to their clusters and their respective noise variances:

$$\mathbb{K}_k \triangleq \left\{ (q, \sigma_{v,q}^2) | q \in \mathcal{C}_k \right\} \tag{5.2}$$

When two agents decide to cluster, then their cluster sets are merged and all agents in these sets become part of the same larger cluster. As such, whenever two agents meet and they are not members of the same cluster, then their cluster sets are necessarily disjoint.

During the information sharing stage, agents apply the diffusion strategy (4.15)-(4.16) and share information to solve the inference task. In this chapter, although unnecessary, we assume that the combination matrix A is doubly-stochastic, i.e., the entries on each of its rows and columns add up to one, such as selecting A to be the Laplacian combination rule [17, 141] or the Metropolis combination rule [8, 141]. Then, we have

$$A^T \mathbb{1} = \mathbb{1}, \quad A \mathbb{1} = \mathbb{1} \tag{5.3}$$

In the context of algorithm (4.15)-(4.16), the information $\mathbb{I}_{k,i}$ to be shared by each agent k is the intermediate estimates $\boldsymbol{\psi}_{k,i}$. As illustrated in Fig. 5.2, it is obvious that agents in $\mathcal{N}_k$ should belong to the cluster of agent k, denoted by $\mathcal{C}_k$, i.e., $\mathcal{N}_k \subset \mathcal{C}_k$. The cluster of agent k includes two types of agents: (a) those agents which agent k has decided to cluster with and, therefore, has direct links to them, and (b) agents which agent k has a path through other intermediate agents to connect with. In other words, the set $\mathcal{C}_k$ represents a connected sub-network that includes k and its immediate neighborhood in addition to other agents. Formally, the cluster set $\mathcal{C}_k$ is constructed as follows. Representing the connection topology graphically, we connect two agents k and ℓ by an edge if $\boldsymbol{I}_{k\ell} = 1$. Then, the cluster $\mathcal{C}_k$ is the maximally connected subnetwork containing agent k. In this way, for any other agent in $\mathcal{C}_k$, there will exist at least one path connecting agent k to it either directly by an edge, or by means of a path passing through other intermediate agents.

5.3 Combined Cost for Clustering Agreement

In the cluster formation stage, when two agents k and ℓ meet randomly, they select their actions $\{\boldsymbol{a}_{k\ell}, \boldsymbol{a}_{\ell k}\}$ based on their assessment of a long-term expected return as follows. Instead

of the long-term discounted cost discussed in Sec. 4.3.1, we assume that agents construct their long-term expected returns based on their steady-state performance in the information sharing stage. Therefore, each agent k employs a combined cost function that takes into account the cost of communicating with agent ℓ *and* the contribution of agent ℓ towards the estimation task (i.e., whether it will help reduce the steady-state mean-square error). The combined cost function for agent k depends on the actions by both agents and on their existing clusters:

$$J_k(\boldsymbol{a}_{k\ell}, \boldsymbol{a}_{\ell k}|\mathcal{C}_k, \mathcal{C}_\ell) \triangleq$$
$$\begin{cases} \mathrm{MSD}_k(\mathcal{C}_k \cup \mathcal{C}_\ell) + \beta_k \left(\sum_{q \in \mathcal{N}_k \cup \{\ell\}} \boldsymbol{I}_{kq} c_{kq} \right), & \text{if } (\boldsymbol{a}_{k\ell}, \boldsymbol{a}_{\ell k}) = (1,1) \\ \mathrm{MSD}_k(\mathcal{C}_k) + \beta_k \left(\sum_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq} \right), & \text{otherwise} \end{cases} \tag{5.4}$$

where β_k is a normalization parameter, and MSD_k denotes the steady-state mean-square-deviation (MSD) measure for agent k:

$$\mathrm{MSD}_k \triangleq \lim_{i \to \infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \tag{5.5}$$

in terms of the error vector $\widetilde{\boldsymbol{w}}_{k,i} \triangleq w^o - \boldsymbol{w}_{k,i}$. Moreover, the notation $\mathrm{MSD}_k(\mathcal{C}_k)$ for cluster $\mathcal{C}_k$ is used to denote the MSD value that would be attained by agent k if its cluster is $\mathcal{C}_k$. In Table 5.1, we summarize the resulting cost values for the agents under their respective actions.

Let us now explain how the MSD values in (4.22) can be evaluated. Consider an arbitrary agent k and a cluster set $\mathcal{C}_k$ of size K. Under the assumption that the regressors $\boldsymbol{u}_{k,i}$ are spatially and temporally independent and that the step-size μ is sufficiently small, it holds that for the doubly-stochastic A, we have the following expression (refer to Equations (89) and (97) in [125] or Equation (32) in [17]):

$$\mathrm{MSD}_k(\mathcal{C}_k) \approx \frac{\mu M}{2} \cdot \frac{1}{K^2} \sum_{q \in \mathcal{C}_k} \sigma_{v,q}^2 \tag{5.6}$$

Suppose agent k meets agent ℓ with cluster $\mathcal{C}_\ell$ of size L. We note that one of two situations will occur: $\mathcal{C}_k = \mathcal{C}_\ell$ or $\mathcal{C}_k \bigcap \mathcal{C}_\ell = \emptyset$. In the trivial case that $\mathcal{C}_k = \mathcal{C}_\ell$, we have

$$\mathrm{MSD}_k(\mathcal{C}_k \cup \mathcal{C}_\ell) = \mathrm{MSD}_k(\mathcal{C}_k) \tag{5.7}$$

since agents k and ℓ have the same cluster. For $\mathcal{C}_k \bigcap \mathcal{C}_\ell = \emptyset$, if agents k and ℓ fail to reach agreement, which means $\boldsymbol{I}_{k\ell} = 0$, then we again obtain $\text{MSD}_k(\mathcal{C}_k)$ for (5.6). On the other hand, if they successfully reach agreement ($\boldsymbol{I}_{k\ell} = 1$), then

$$\text{MSD}_k(\mathcal{C}_k \cup \mathcal{C}_\ell) \approx \frac{\mu M}{2} \cdot \frac{1}{(K+L)^2} \sum_{q \in \mathcal{C}_k \cup \mathcal{C}_\ell} \sigma_{v,q}^2 \tag{5.8}$$

In this way, the combined cost values in (5.4) are given by:

$$J_k(\boldsymbol{a}_{k\ell}, \boldsymbol{a}_{\ell k} | \mathcal{C}_k, \mathcal{C}_\ell) =$$

$$\begin{cases} \frac{\mu M}{2} \cdot \frac{1}{(K+L)^2} \sum\limits_{q \in \mathcal{C}_k \cup \mathcal{C}_\ell} \sigma_{v,q}^2 + \beta_k \sum\limits_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq} + \beta_k c_{k\ell}, & \text{if } (\boldsymbol{a}_{k\ell}, \boldsymbol{a}_{\ell k}) = (1,1) \\ \frac{\mu M}{2} \cdot \frac{1}{K^2} \sum\limits_{q \in \mathcal{C}_k} \sigma_{v,q}^2 + \beta_k \sum\limits_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq}, & \text{otherwise} \end{cases} \tag{5.9}$$

Then, agents choose the actions that minimize their combined cost function (5.9). Once $\boldsymbol{I}_{kl} = 1$, agents k and ℓ start sharing estimates in the information sharing stage. To prevent agents from deviating from the agreement, we punish the deviant agents in the following manner: if any agent k violates the agreement to cooperate with agent ℓ, agent ℓ broadcasts this misbehavior to its neighbors and from there to their neighbors and agents will stop sharing estimates with agent k permanently.

We remark that the individual actions of agents could impact the combined cost values of other agents in the same cluster. However, individual actions do not worsen the marginal combined costs of other agents in a cluster. To see this, if no larger clustering (no new agreement) occurs, the combined cost of every agent in a cluster remains the same. If a new clustering agreement of agents, say, k and ℓ, is made, the MSD costs of other agents reduce but there is no addition communication cost required by them, and thus their combined costs reduce.

5.4 Cluster Formation Process

The following lemma characterizes the conditions for cluster formation.

Lemma 5.1. *Agents k and ℓ reach agreement to cluster ($\boldsymbol{I}_{k\ell} = 1$), when the following two*

conditions are met:

$$\frac{\sum\limits_{q \in \mathcal{C}_k} \sigma_{v,q}^2}{K^2} - \frac{\sum\limits_{q \in \mathcal{C}_k \cup \mathcal{C}_\ell} \sigma_{v,q}^2}{(K+L)^2} > \frac{2}{\mu M} \beta_k c_{k\ell} \tag{5.10}$$

and

$$\frac{\sum\limits_{q \in \mathcal{C}_\ell} \sigma_{v,q}^2}{L^2} - \frac{\sum\limits_{q \in \mathcal{C}_k \cup \mathcal{C}_\ell} \sigma_{v,q}^2}{(K+L)^2} > \frac{2}{\mu M} \beta_\ell c_{\ell k} \tag{5.11}$$

Proof. From Table 1, we first note that if agent ℓ selects $\boldsymbol{a}_{\ell k} = 0$, then it is indifferent to agent k selecting $\boldsymbol{a}_{k\ell} = 0$ or 1. On the other hand, in the case of $\boldsymbol{a}_{\ell k} = 1$, if we have

$$J_k(\boldsymbol{a}_{k\ell} = 0, \boldsymbol{a}_{\ell k} = 1 | \mathcal{C}_k, \mathcal{C}_\ell) > J_k(\boldsymbol{a}_{k\ell} = 1, \boldsymbol{a}_{\ell k} = 1 | \mathcal{C}_k, \mathcal{C}_\ell) \tag{5.12}$$

then agent k should choose $\boldsymbol{a}_{k\ell} = 1$ to obtain a lower combined cost. Therefore, condition (5.12) ensures the best strategy for agent k to be $\boldsymbol{a}_{k\ell} = 1$. Using (5.9) we can rewrite (5.12) as

$$\frac{\mu M}{2} \left(\frac{\sum\limits_{q \in \mathcal{C}_k} \sigma_{v,q}^2}{K^2} \right) + \beta_k \sum_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq}$$

$$> \frac{\mu M}{2} \left(\frac{\sum\limits_{q \in \mathcal{C}_k \cup \mathcal{C}_\ell} \sigma_{v,q}^2}{(K+L)^2} \right) + \beta_k \sum_{q \in \mathcal{N}_k} \boldsymbol{I}_{kq} c_{kq} + \beta_k c_{k\ell} \tag{5.13}$$

which is equivalent to (5.10). Similarly, we can obtain condition (5.11) to ensure $\boldsymbol{a}_{\ell k} = 1$ from agent ℓ's perspective. $\square$

Note that when conditions (5.10) and (5.11) hold, the dominant strategies for agents k and ℓ become $\boldsymbol{a}_{k\ell} = 1$ and $\boldsymbol{a}_{\ell k} = 1$. On the other hand, when either one of conditions (5.10) or (5.11) fails to hold, agents have no incentive to cluster. In this case, $(\boldsymbol{a}_{k\ell}, \boldsymbol{a}_{\ell k}) = (1, 1)$ will not be chosen, which results in $\boldsymbol{I}_{k\ell} = 0$. We assume agents k and ℓ select $(\boldsymbol{a}_{k\ell}, \boldsymbol{a}_{\ell k}) = (0, 0)$ if equalities occur in (5.10) and (5.11). From Lemma 1, we know that clusters $\mathcal{C}_k$ and $\mathcal{C}_\ell$ unite if both conditions (5.10) and (5.11) hold. Furthermore, we observe that low weighted transmission costs, $\beta_k c_{k\ell}$ and $\beta_\ell c_{\ell k}$, facilitate the formation of the united cluster. Now, let us consider networks with uniform $\beta_k = \beta_\ell \equiv \beta$ and $c_{k\ell} = c_{\ell k} \equiv c$. If every agent further has the same noise variance, we obtain the following result.

Lemma 5.2. *If the noise variances across the network are uniform, i.e., $\sigma_{v,q}^2 \equiv \sigma_v^2$, then the following condition guarantees the cluster formation $\mathcal{C}_k \cup \mathcal{C}_\ell$:*

$$\frac{K+L}{\sigma_v^2}\frac{2}{\mu M}\beta c < \min\left\{\frac{L}{K}, \frac{K}{L}\right\} \tag{5.14}$$

Proof. For agent k, it follows from (5.10) that we must have

$$\frac{L}{K} > \frac{K+L}{\sigma_v^2}\frac{2}{\mu M}\beta c \tag{5.15}$$

Similarly, for agent ℓ it follows from (5.11) that we must have

$$\frac{K}{L} > \frac{K+L}{\sigma_v^2}\frac{2}{\mu M}\beta c \tag{5.16}$$

Combining both results, we obtain (5.14). □

Therefore, if we want to facilitate the formation of larger clusters, Lemma 2 suggests to maximize the right-hand side of (5.14), which occurs when $K = L$ and the maximum value becomes equal to one. In other words, larger clustering is more likely to occur for clusters $\mathcal{C}_k$ and $\mathcal{C}_\ell$ of equal sizes. Now, let us examine the case in which the clusters $\mathcal{C}_k$ and $\mathcal{C}_\ell$ have the same sizes but their agents have heterogeneous noise variances.

Lemma 5.3. *If clusters $\mathcal{C}_k$ and $\mathcal{C}_\ell$ have the same sizes, i.e., $K = L$, then the following condition guarantees the cluster formation $\mathcal{C}_k \cup \mathcal{C}_\ell$:*

$$\frac{8}{\mu M}\beta c < \min\left\{\frac{1}{K}(3\bar{\sigma}_k^2 - \bar{\sigma}_\ell^2), \frac{1}{L}(3\bar{\sigma}_\ell^2 - \bar{\sigma}_k^2)\right\} \tag{5.17}$$

where

$$\bar{\sigma}_k^2 \triangleq \frac{1}{K}\sum_{q\in\mathcal{C}_k}\sigma_{v,q}^2 \quad and \quad \bar{\sigma}_\ell^2 \triangleq \frac{1}{L}\sum_{q\in\mathcal{C}_\ell}\sigma_{v,q}^2 \tag{5.18}$$

are the average noise variances of $\mathcal{C}_k$ and $\mathcal{C}_\ell$, respectively.

Proof. For agent k, we conclude from (5.10) that we must have:

$$3\frac{\sum\limits_{q\in\mathcal{C}_k}\sigma_{v,q}^2}{K} - \frac{\sum\limits_{q\in\mathcal{C}_\ell}\sigma_{v,q}^2}{L} > \frac{8K}{\mu M}\beta c \tag{5.19}$$

152

Similarly, for agent ℓ it must hold that

$$3\frac{\sum\limits_{q\in\mathcal{C}_\ell}\sigma_{v,q}^2}{L} - \frac{\sum\limits_{q\in\mathcal{C}_k}\sigma_{v,q}^2}{K} > \frac{8L}{\mu M}\beta c \tag{5.20}$$

Combining both conditions, we obtain (5.17). □

Again, the maximum of the term on the right-hand side of (5.17) occurs when

$$(3L + K)\bar{\sigma}_k^2 = (3K + L)\bar{\sigma}_L^2 \tag{5.21}$$

Therefore, for clusters of equal sizes, two clusters with the same (weighted) average noise variance will be more likely to unite.

5.5 Simulation Results

In our simulations, we consider a network with 20 agents. During the first 10 time instants, agents are uniformly and randomly paired. Then, agents proceed to cooperate within their clusters to solve the estimation problem. The length of w^o is $M = 3$ and we randomly choose its entries and normalize them to satisfy $\|w^o\| = 1$. The regressor $\{\boldsymbol{u}_{k,i}\}$ is zero-mean and $R_{u,k}$ is diagonal with entries uniformly generated between $[0,1]$. The background noise $\boldsymbol{v}_k(i)$ is temporally white and spatially independent Gaussian distributed with zero-mean and assumed to be uniform with variance $\sigma_{v,k}^2 = \sigma_v^2 = -6$ (dB). We set $\mu = 0.005$, $\beta_k = \beta = 1$, and $c_{k\ell} = c = 5 \times 10^{-5}$ for all agents.

Fig. 5.3(a) shows the topology evolution from $i = 1$ to 4. We observe that agents gradually form clusters to maximize their own utilities. The final topology with three disjoint clusters is shown in Fig. 5.3(b). Cooperating over the resulting sub-networks, agents start to share estimates and run algorithm (4.15)-(4.16). We simulate the corresponding steady-state MSD in Fig. 5.4(a) where agents are indexed and grouped according to their clusters. We observe that through clustering, every agent is able to achieve better estimation performance than if the agents were to act independently of the other agents by running their own individual LMS recursions. Fig. 5.4(b) shows the effect of transmission cost to the cluster formation and thus to the steady-state network MSD.

5.6 Concluding Remarks

In this chapter, we allow the agents to select their partners according to whether they can help them reduce their utility costs. We divided the operation of the network into two stages: a cluster formation stage and an information sharing stage. During cluster formation, agents evaluate a long-term combined cost function and decide on whether to cooperate or not with other agents. During the subsequent information sharing phase, agents share and process information over their sub-networks. Simulations illustrate that through clustering, every agent is able to achieve better estimation performance than noncooperative processing.
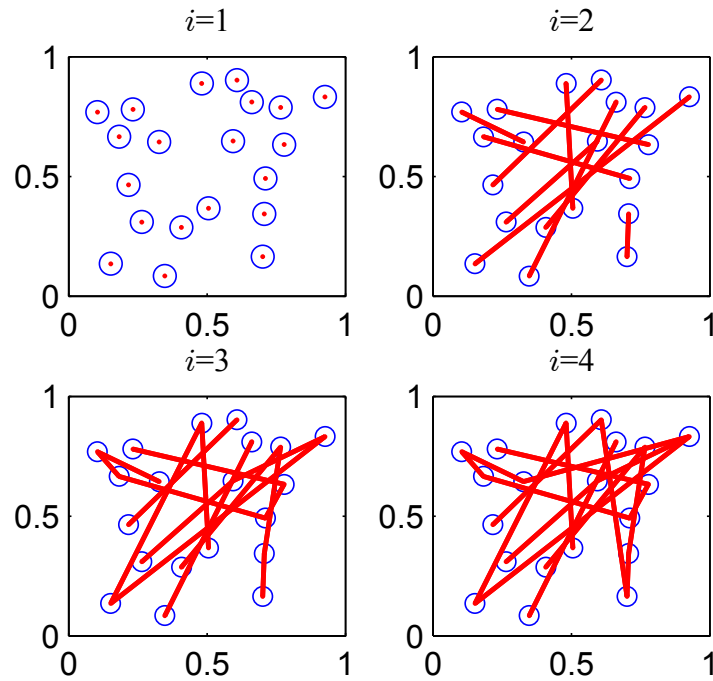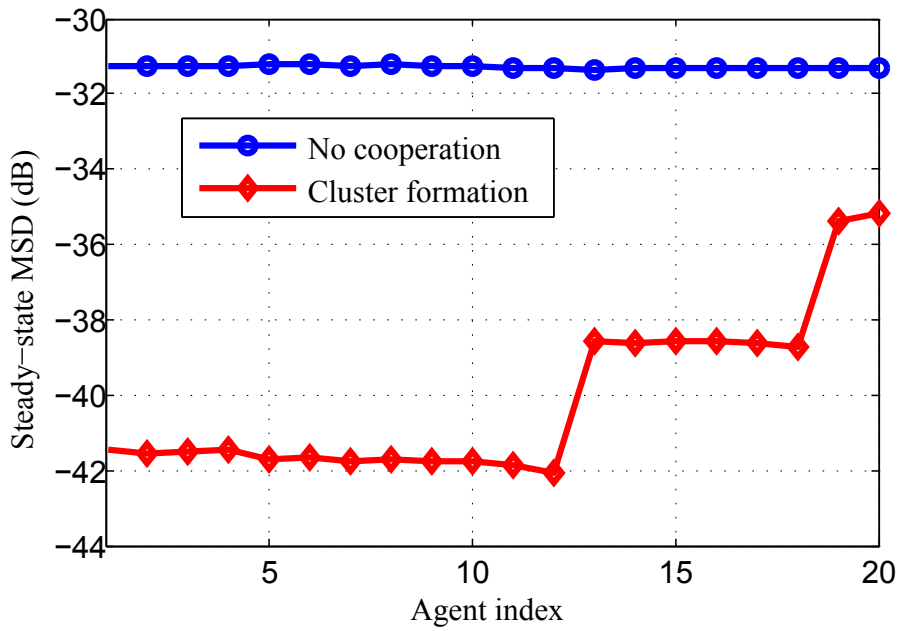
(a) Evolution of clusters.
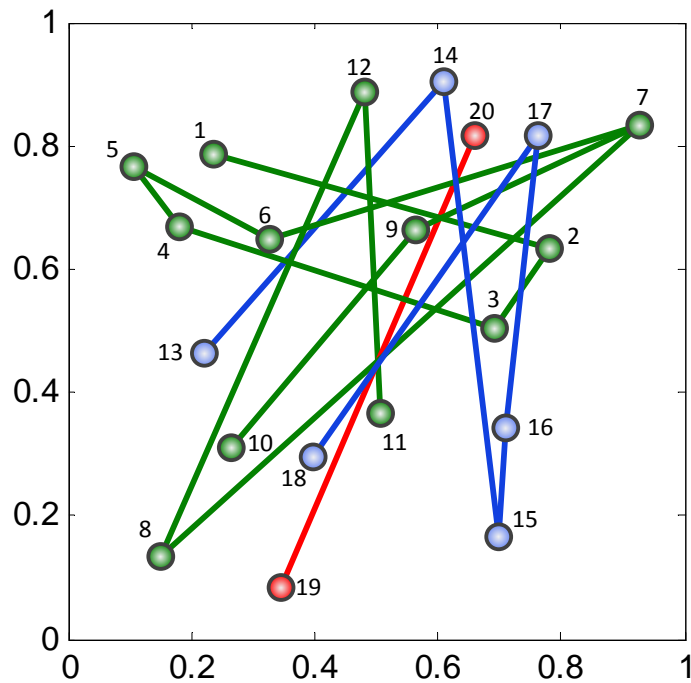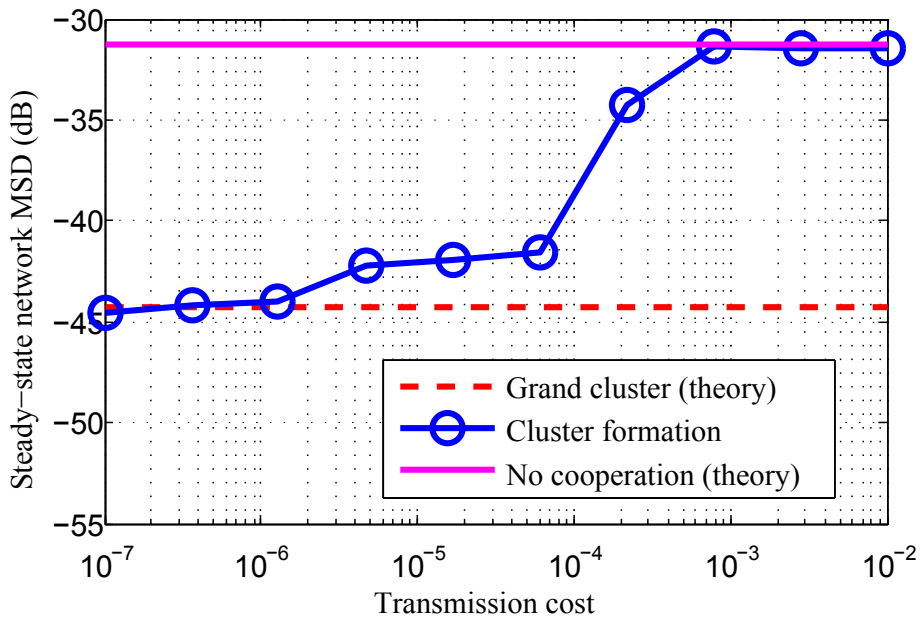

(b) Final clusters.

Figure 5.3: Cluster formation with $c = 5 \times 10^{-5}$ and $\sigma_v^2 = -6$ (dB).

155

(a) Agents in three clusters.

(b) Effect of transmission cost.

Figure 5.4: Simulations of steady-state network MSD.

CHAPTER 6

Future Work

In this dissertation, we focus on stochastic multi-objective optimization problems over network topologies. A key challenge in our formulation is the adaptive nature of the strategies, both for solving the inference task and for the agents to react to the actions of neighbors. In these scenarios, agents need to operate in response to streaming data and be able to respond to changes in the statistical properties of the data, the nature of the task, and even the behavior of neighboring agents. Both stability and efficiency, corresponding to Nash equilibria and Pareto optimal solutions, respectively, is desirable when we study these problems. We first developed distributed and online learning strategies through a penalty reformulation and showed that they approach an asymptotic Nash equilibrium. One critical extension involving cooperation with partial information can still be beneficial to agents. We also examined the case where Nash equilibria may not be Pareto optimal by associating a positive cost for information sharing. Such inefficient problems can be dealt with by the proposed adaptive reputation and cluster formation protocols. In the following, we list several topics that deserve further examination:

- One interesting future work would be to explore how the converging point (asymptotic Nash equilibrium) of the developed penalty-based algorithms relates to the variational equilibrium obtained by KKT conditions with identical Lagrange multipliers. It would be useful if we could select a target generalized Nash equilibrium in the original GNEP for our diffusion algorithms to converge to.

- Throughout this dissertation, we assume the individual cost and constraints functions are differentiable. A possible future work is to relax the differentiability assumption and explore the use of sub-gradient or proximal-based methods. This relaxation can

allow us to use non-differentiable penalty functions and thus enlarge the flexibility of our algorithms.

- Asynchronous adaptation learning is also a practical extension so that agents do not need to execute the update of actions simultaneously. It would be useful if we can study the performance degradation due to asynchronous behavior by agents.

- While the partial information learning can be beneficial for agents, the appearance of communication overhead may destroy the cooperation of agents. Therefore, an interesting problem will be developing a new mechanism for this general network where agents might have different or partially common inference targets subject to positive communication cost.

REFERENCES

[1] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, Jul. 2008.

[2] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[3] S. Sardellitti, M. Giona, and S. Barbarossa, "Fast distributed average consensus algorithms based on advection-diffusion processes," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 826–842, Feb. 2010.

[4] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[5] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, Jan. 2009.

[6] ——, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE J. Sel. Topics in Signal Process.*, vol. 5, no. 5, pp. 674–690, Aug. 2011.

[7] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.

[8] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, no. 1, pp. 65–78, Sep. 2004.

[9] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[10] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, Jul. 2008.

[11] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[12] S. Chouvardas, K. Slavakis, and S. Theodoridis, "Adaptive robust distributed learning in diffusion sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4692–4707, Oct. 2011.

[13] N. Takahashi and I. Yamada, "Link probability control for probabilistic diffusion least-mean squares over resource-constrained networks," in *Proc. IEEE ICASSP*, Dallas, TX, Mar. 2010, pp. 3518–3521.

[14] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4289–4305, Aug. 2012.

[15] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.

[16] ——, *Adaptation, Learning, and Optimization over Networks.* Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, NOW Publishers, Jul. 2014.

[17] ——, "Diffusion adaptation over networks," in *Academic Press Library in Signal Processing, R. Chellapa and S. Theodoridis, Eds.*, vol. 4, pp. 323–454, Academic Press, Elsevier, 2014.

[18] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4129–4144, Aug. 2014.

[19] ——, "Diffusion LMS over multitask networks," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2733–2748, Jun. 2015.

[20] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Multitask diffusion adaptation over asynchronous networks," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2835–2850, Jun. 2016.

[21] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.

[22] A. Ghosh, L. Cottatellucci, and E. Altman, "Nash equilibrium for femto-cell power allocation in hetnets with channel uncertainty," in *Proc. IEEE GLOBECOM*, San Diego, USA, Dec. 2015, pp. 1–7.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[24] C. Metzler, B. F. Hobbs, and J. S. Pang, "Nash-Cournot equilibria in power markets on a linearized DC network with arbitrage: Formulations and properties," *Netw. Spatial Econ.*, vol. 3, no. 2, pp. 123–150, 2003.

[25] K. Bimpikis, S. Ehsani, and R. Ilkilic, "Cournot competition in networked markets," in *Proc. of the 15th ACM Conf. on Economics and Computation*, Stanford, CA, USA, Jun. 2014, pp. 733–733.

[26] J.-S. Pang and M. Fukushima, "Quasivariational inequalities, generalized Nash equilibria, and multi-leader-follower games," *Comput. Manag. Sci.*, vol. 2, no. 1, pp. 21–56, Jan. 2005.

[27] F. Facchinei and C. Kanzow, "Generalized Nash equilibrium problems," *4OR, Quart. J. Oper. Res.*, vol. 5, no. 3, pp. 173–210, Sep. 2007.

[28] A. Fischer, M. Herrich, and K. Schönefeld, "Generalized Nash equilibrium problems - recent advances and challenges," *Pesquisa Operacional*, vol. 34, no. 3, pp. 521–558, Sep. 2014.

[29] G. Scutari, D. Palomar, F. Facchinei, and J.-S. Pang, "Convex optimization, game theory, and variational inequality theory," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 35–49, May 2010.

[30] F. Facchinei, A. Fischer, and V. Piccialli, "On generalized Nash games and variational inequalities," *Oper. Res. Lett.*, vol. 35, no. 2, pp. 159–164, 2007.

[31] A. Dreves, F. Facchinei, C. Kanzow, and S. Sagratella, "On the solution of the KKT conditions of generalized Nash equilibrium problems," *SIAM J. Optim.*, vol. 21, no. 3, pp. 1082–1108, 2011.

[32] F. Facchinei and J.-S. Pang, *Finite Dimensional Variational Inequalities and Complementarity Problems.* Springer-Verlag, NY, 2003.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge Univ. Press, 2004.

[34] L. A. Zadeh, "Optimality and non-scalar-valued performance criteria," *IEEE Trans. Autom. Control*, vol. 8, no. 1, pp. 59–60, Jan. 1963.

[35] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.

[36] A. von Heusinger and C. Kanzow, "Optimization reformulations of the generalized Nash equilibrium problem using Nikaido-Isoda-type functions," *Comput. Optim. Appl.*, vol. 43, no. 3, pp. 353–377, Jul. 2009.

[37] ——, "Relaxation methods for generalized Nash equilibrium problems with inexact line search," *J. Optim. Theory Appl.*, vol. 143, no. 1, pp. 159–183, Oct. 2009.

[38] F. Facchinei, C. Kanzow, and S. Sagratella, "Solving quasi-variational inequalities via their KKT conditions," *Math. Programming*, vol. 144, no. 1, pp. 369–412, Apr. 2014.

[39] H. e Oliveira and A. Petraglia, "Solving generalized Nash equilibrium problems through stochastic global optimization," *Applied Soft Computing*, vol. 39, pp. 21–35, Feb. 2016.

[40] G. Debreu, "A social equilibrium existence theorem," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 38, no. 10, pp. 886–893, Oct. 1952.

[41] P. T. Harker, "Generalized Nash games and quasi-variational inequalities," *European Journal of Operational Research*, vol. 54, no. 1, pp. 81–94, Sep. 1991.

[42] J. Contreras, M. Klusch, and J. B. Krawczyk, "Numerical solutions to Nash-Cournot equilibria in coupled constraint electricity markets," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 195–206, Feb. 2004.

[43] B. F. Hobbs and J. S. Pang, "Nash-Cournot equilibria in electric power markets with piecewise linear demand functions and joint constraints," *Oper. Res.*, vol. 55, no. 1, pp. 113–127, 2007.

[44] I. Abada, S. Gabriel, V. Briat, and O. Massol, "A generalized NashCournot model for the northwestern European natural gas markets with a fuel substitution demand function: The GaMMES model," *Netw. Spat. Econ.*, vol. 13, no. 1, pp. 1–42, Mar. 2013.

[45] S. T. Chung, S. J. Kim, J. Lee, and J. M. Cioffi, "A game-theoretic approach to power allocation in frequency-selective gaussian interference channels," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun. 2003, p. 316.

[46] J. B. Krawczyk, "Coupled constraint Nash equilibria in environmental games," *Resour. Energy Econ.*, vol. 27, no. 2, pp. 157–181, Jun. 2005.

[47] D. Ardagna, B. Panicucci, and M. Passacantando, "Generalized Nash equilibria for the service provisioning problem in cloud systems," vol. 6, no. 4, pp. 429–442, Apr. 2013.

[48] V. Cardellini, V. D. N. Personé, V. D. Valerio, F. Facchinei, V. Grassi, F. L. Presti, and V. Piccialli, "A game-theoretic approach to computation offloading in mobile cloud computing," *Math. Program.*, vol. 157, no. 2, pp. 421–449, Jun. 2016.

[49] F. Facchinei and C. Kanzow, "Penalty methods for the solution of generalized nash equilibrium problems," *SIAM J. Optim.*, vol. 20, no. 5, pp. 2228–2253, 2010.

[50] M. Fukushima, "Restricted generalized nash equilibria and controlled penalty algorithm," *Computat. Manage. Sci.*, vol. 8, no. 3, pp. 201–218, Aug. 2011.

[51] D. Watling, "User equilibrium traffic network assignment with stochastic travel times and late arrival penalty," *European J. Oper. Res.*, vol. 175, no. 3, pp. 1539–1556, Dec. 2006.

[52] H. Xu and D. Zhang, "Stochastic Nash equilibrium problems: sample average approximation and applications," *Comput. Optim. Appl.*, vol. 55, no. 3, pp. 597–645, Jul. 2013.

[53] J. Koshal, A. Nedic, and U. V. Shanbhag, "Regularized iterative stochastic approximation methods for stochastic variational inequality problems," *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 594–609, Mar. 2013.

[54] Z. J. Towfic and A. H. Sayed, "Adaptive penalty-based distributed stochastic convex optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3924–3938, Aug. 2014.

[55] S. Bubeck and N. Cesa-Bianchi, *Regret analysis of stochastic and nonstochastic multi-armed bandit problems.* Foundations and Trends in Machine Learning, 2012, vol. 5, no. pp. 1-122.

[56] E. Hazan, "A survey: The convex optimization approach to regret minimization," in *Optimization for Machine Learning, S. Sra, S. Nowozin, and S. J. Wright, Eds.*, pp. 287–304, MIT Press, 2012.

[57] S. Shalev-Shwartz, *Online learning and online convex optimization.* Foundations and Trends in Machine Learning, vol. 2, no. pp. 107-194, 2014.

[58] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proc. AAAI ICML*, Washington D.C., USA, Aug. 2003, pp. 928–936.

[59] O. Besbes, Y. Gur, and A. Zeevi, "Non-stationary stochastic optimization," *Oper. Res.*, vol. 63, no. 5, pp. 1227–1244, 2015.

[60] R. Sutton and A. Barto, *Reinforcement learning: An introduction*, 2nd ed. MIT Press, 2017.

[61] B. T. Polyak, *Introduction to Optimization.* New York, NY, USA: Optimization Software, 1987.

[62] M. C. Pinar and S. Zenios, "On smoothing exact penalty functions for convex constrained optimization," *SIAM J. Optim.*, vol. 4, no. 3, pp. 486–511, 1994.

[63] Y. Nesterov, "Primal-dual subgradient methods for convex problems," *Math. Program.*, vol. 120, no. 1, pp. 261–283, Aug. 2009.

[64] L. Xiao, "Dual averaging methods for regularized stochastic learning and online optimization," *J. Mach. Learn. Res.*, vol. 11, pp. 2543–2596, Oct. 2010.

[65] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Trans. Autom. Control*, vol. 57, no. 3, pp. 592–606, Mar. 2012.

[66] A. Juditsky, A. Nemirovski, and C. Tauvel, "Solving variational inequalities with stochastic mirror-prox algorithm," *Stoch. Syst.*, vol. 1, no. 1, pp. 17–58, 2011.

[67] P. Mertikopoulos, "Learning in concave games with imperfect information," Aug. 2016, *arXiv:1608.07310 [Online].* Available: https://arxiv.org/abs/1608.07310.

[68] M. Parzy and H. Bogucka, "QoS support in radio resource sharing with Cournot competition," in *2nd International Workshop on Cognitive Information Processing (CIP)*, Elba, Italy, Jun. 2010, pp. 93–98.

[69] U. Ravat and U. V. Shanbhag, "On the characterization of solution sets of smooth and nonsmooth convex stochastic Nash games," *SIAM J. Optim.*, vol. 21, no. 3, pp. 1168–1199, Sep. 2011.

[70] C. R. Johnson, "Positive definite matrices," *Amer. Math. Monthly*, vol. 77, no. 3, pp. 259–264, Mar. 1970.

[71] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms.* Wiley, NY, 1993.

[72] D. P. Bertsekas, "Necessary and sufficient conditions for a penalty method to be exact," *Math. Programming Study*, vol. 9, pp. 87–99, 1975.

[73] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course.* Kluwer Academic Publishers, 2004.

[74] A. Ben-Tal and M. Zibulevsky, "Penalty/barrier multiplier methods for convex programming problems," *SIAM J. Optim.*, vol. 7, no. 2, pp. 347–366, 1997.

[75] P. Tseng and D. P. Bertsekas, "On the convergence of the exponential multiplier method for convex programming," *Math. Programming*, vol. 60, no. 1, pp. 1–19, Jun. 1993.

[76] A. Matsui, "Best response dynamics and socially stable strategies," *Journal of Economic Theory*, vol. 57, no. 2, pp. 343–362, Aug. 1992.

[77] S. D. Flam, "Equilibrium, evolutionary stability and gradient dynamics," *Int. Game Theory Rev.*, vol. 4, no. 4, pp. 357–370, 2002.

[78] J. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, distributed convergence to Nash equilibria," *IEEE Trans. Autom. Control*, vol. 50, no. 3, pp. 312–327, Mar. 2005.

[79] N. Li and J. Marden, "Designing games for distributed optimization," *IEEE J. Sel. Topics in Signal Process.*, vol. 7, no. 2, pp. 230–242, Apr. 2013.

[80] X. Zhao and A. H. Sayed, "Asynchronous adaptation and learning over networks – Part I: Modeling and stability analysis," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 811–826, Feb. 2015.

[81] J. Chen and A. H. Sayed, "On the learning behavior of adaptive networks — Part I: Transient analysis," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3487–3517, Jun. 2015.

[82] ——, "On the learning behavior of adaptive networks — Part II: Performance analysis," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3518–3548, Jun. 2015.

[83] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.

[84] A. Kannan and U. V. Shanbhag, "Distributed computation of equilibria in monotone nash games via iterative regularization techniques," *SIAM J. Optim.*, vol. 22, no. 4, pp. 1177–1205, 2012.

[85] A. A. Kulkarni and U. V. Shanbhag, "On the variational equilibrium as a refinement of the generalized Nash equilibrium," *Automatica*, vol. 48, no. 1, pp. 45–55, Jan. 2012.

[86] ——, "Revisiting generalized Nash games and variational inequalities," *Journal of Optimization Theory and Applications*, vol. 154, no. 1, pp. 1–12, Jul. 2012.

[87] B. Ying and A. H. Sayed, "Performance limits of single-agent and multi-agent subgradient stochastic learning," in *Proc. IEEE ICASSP*, Shanghai, China, Mar. 2016, pp. 4905–4909.

[88] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications*. Academic Press, 1980.

[89] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge Univ. Press, 1990.

[90] R. Arablouei, S. Werner, Y. F. Huang, and K. Doanay, "Distributed least mean-square estimation with partial diffusion," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 472–484, Jan. 2014.

[91] C. R. J. Chen, S. K. Ting and A. H. Sayed, "Group diffusion LMS," in *Proc. IEEE ICASSP*, Shanghai, China, Mar. 2016, pp. 4925–4929.

[92] A. H. Sayed, *Adaptive Filters*. Wiley, NJ, 2008.

[93] A. J. Laub, *Matrix Analysis for Scientists and Engineers*. SIAM, PA, 2005.

[94] O. N. Gharehshiran, V. Krishnamurthy, and G. Yin, "Distributed energy-aware diffusion least mean squares: Game-theoretic learning," *IEEE J. Sel. Topics in Signal Process.*, vol. 7, no. 5, pp. 821–836, Oct. 2013.

[95] C.-K. Yu, M. van der Schaar, and A. H. Sayed, "Distributed spectrum sensing in the presence of selfish users," in *Proc. IEEE CAMSAP*, Saint Martin, Dec. 2013, pp. 392–395.

[96] A. Filali, A. Hafid, and M. Gendreau, "Bandwidth and computing resources provisioning for grid applications and services," in *Proc. IEEE ICC*, Dresden, Germany, Jun. 2009, pp. 1–6.

[97] R. Bekkerman, M. Bilenko, and J. Langford, *Scaling up Machine Learning, Parallel and Distributed Approaches*. Cambridge University Press, 2011.

[98] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dyn. Games Appl.*, vol. 1, no. 1, pp. 3–49, Mar. 2011.

[99] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Nonbayesian social learning," *Game. Econ. Behav.*, vol. 76, no. 1, pp. 210–225, Sep. 2012.

[100] X. Zhao and A. H. Sayed, "Learning over social networks via diffusion adaptation," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Nov. 2012, pp. 709–713.

[101] R. S. G Gigerenzer, *Bounded Rationality: The Adaptative Toolbox*. MIT Press, 2002.

[102] D. Foster and R. Vohra, "Regret in the on-line decision problem," *Games and Economic Behavior*, vol. 29, no. 1-2, pp. 7–35, Oct. 1999.

[103] O. N. Gharehshiran, V. Krishnamurthy, and G. Yin, "Distributed tracking of correlated equilibria in regime switching noncooperative games," *IEEE Trans. Autom. Control*, vol. 58, no. 10, pp. 2435–2450, Oct. 2013.

[104] G. J. Mailath and L. Samuelson, *Repeated Games and Reputations: Long-Run Relationships.* Oxford University Press, 2006.

[105] J. Xu and M. van der Schaar, "Social norm design for information exchange systems with limited observations," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 11, pp. 2126–2135, Dec. 2012.

[106] Y. Zhang and M. van der Schaar, "Reputation-based incentive protocols in crowdsourcing applications," in *Proc. IEEE INFOCOM*, Orlando, Florida, USA, Mar. 2012, pp. 2140–2148.

[107] J. Carter, E. Bitting, and A. A. Ghorbani, "Reputation formalization for an information-sharing multi-agent system," in *Computational Intelligence*, vol. 18, no. 4, 2002, pp. 515–534.

[108] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for wimax and lte-advanced mobile systems," *IEEE Commun. Mag.*, vol. 47, no. 10, pp. 100–105, Oct. 2009.

[109] C. Aliprantis, G. Camera, and D. Puzzello, "A random matching theory," *Games Econ. Behav.*, vol. 59, pp. 1–16, 2007.

[110] G. Ellison, "Cooperation in the prisoner's dilemma with anonymous random matching," *Rev. Econ. Stud.*, vol. 61, no. 3, pp. 567–588, 1994.

[111] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game Theoretic and Logical Foundations.* Cambridge University Press, 2008.

[112] J. Xu, Y. Song, and M. van der Schaar, "Sharing in networks of strategic agents," *IEEE J. Sel. Topics in Signal Process.*, vol. 8, no. 4, pp. 717–731, Aug. 2014.

[113] H. A. Simon, "A behavioral model of rational choice," *Quart. J. Econ.*, vol. 69, no. 1, pp. 99–118, Feb. 1955.

[114] H. Park and M. van der Schaar, "On the impact of bounded rationality in peer-to-peer networks," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 675–678, Aug. 2009.

[115] M. J. Osborne and A. Rubinstein, *A Course in Game Theory.* Cambridge, MA, 1994.

[116] A. Rapaport and A. Chammah, *The Prisoners Dilemma.* Univ. of Michigan Press, 1965.

[117] R. Jurca and B. Faltings, "Eliciting truthful feedback for binary reputation mechanisms," in *Proc. IEEE/WIC/ACM International Conference on Web Intelligence*, Beijing, China, Sep. 2004, pp. 214–220.

[118] M. Fan, Y. Tan, and A. B. Whinston, "Evaluation and design of online cooperative feedback mechanisms for reputation management," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 2, pp. 244–254, Feb. 2005.

[119] D. Houser and J. Wooders, "Reputation in auctions: Theory, and evidence from ebay," *J. Econom. Manage. Strat.*, vol. 15, no. 2, pp. 353–369, Jun. 2006.

[120] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decis. Support Syst.*, vol. 43, no. 2, pp. 618–644, Mar. 2007.

[121] A. Nayyar, A. Gupta, C. Langbort, and T. Basar, "Common information based Markov perfect equilibria for stochastic games with asymmetric information: Finite games," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 555–570, Mar. 2014.

[122] G. Wyatt, "Risk-taking and risk-avoiding behavior – The impact of some dispositional and situational variables," *Journal of Psychology*, vol. 124, pp. 437–447, 1990.

[123] Ø. L. Rørtveit, J. H. Husøy, and A. H. Sayed, "Diffusion LMS with communication constraints," in *Proc. 44th Asilomar Conference on Signals*, Nov. 2010, pp. 1645–1649.

[124] J. Chen and A. H. Sayed, "Distributed Pareto optimization via diffusion strategies," *IEEE J. Sel. Topics in Signal Process.*, vol. 7, no. 2, pp. 205–220, Apr. 2013.

[125] X. Zhao and A. H. Sayed, "Performance limits for distributed estimation over LMS adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5107–5124, Oct. 2012.

[126] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 2002.

[127] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[128] M. Mishali and Y. Eldar, "Wideband spectrum sensing at sub-Nyquist rates," *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 102–135, Jul. 2011.

[129] J. Bazerque and G. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847–1862, Mar. 2010.

[130] Z. Tian, "Compressed wideband sensing in cooperative cognitive radio networks," in *Proc. IEEE GLOBECOM*, New Orleans, LO, Nov. 2008, pp. 1–5.

[131] Z. Li, F. Yu, and M. Huang, "A distributed consensus-based cooperative spectrum-sensing scheme in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 383–393, Jan. 2010.

[132] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, May 2011.

[133] G. Scutari, D. Palomar, and S. Barbarossa, "Cognitive MIMO radio," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 46–59, Nov. 2008.

[134] H. Park and M. van der Schaar, "Coalition-based resource reciprocation strategies for P2P multimedia broadcasting," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 557–567, Sep. 2008.

[135] W. Saad, Z. Han, M. Debbah, A. Hjorungnes, and T. Basar, "Coalitional games for distributed collaborative spectrum sensing in cognitive radio networks," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, Nov. 2009, pp. 2114–2122.

[136] Y. Zhang and M. van der Schaar, "Robust reputation protocol design for online communities: A stochastic stability analysis," *IEEE J. Selected Topics in Signal Process.*, vol. 7, no. 5, pp. 907–920, Oct. 2013.

[137] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in ofdm systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 223–229, Sep. 2002.

[138] Z. Quan, S. Cui, V. H. Poor, and A. H. Sayed, "Collaborative wideband sensing for cognitive radios," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 60–73, Nov. 2008.

[139] J. P. Kahan and A. Rapoport, *Theories of Coalition Formation.* Hillsdale, NJ: L. Erlbaum Associates, 1984.

[140] W. A. Gamson, "A theory of coalition formation," *Amer. Sociolog. Rev., 26 (3)*, pp. 373–382, 1961.

[141] D. S. Scherber and H. C. Papadopoulos, "Locally constructed algorithms for distributed computations in ad-hoc networks," in *Proc. Information Processing Sensor Networks (IPSN)*, Apr. 2004, pp. 11–19.