ABSTRACT

Title of dissertation: GENERALIZED DISTRIBUTED CONSENSUS-
-BASED ALGORITHMS FOR UNCERTAIN
SYSTEMS AND NETWORKS

Ion Matei,
Doctor of Philosophy, 2010

Dissertation directed by: Professor John S. Baras
Professor Nuno C. Martins
Department of Electrical and
Computer Engineering

We address four problems related to multi-agent optimization, filtering and agreement. First, we investigate collaborative optimization of an objective function expressed as a sum of local convex functions, when the agents make decisions in a distributed manner using local information, while the communication topology used to exchange messages and information is modeled by a graph-valued random process, assumed independent and identically distributed. Specifically, we study the performance of the consensus-based multi-agent distributed subgradient method and show how it depends on the probability distribution of the random graph. For the case of a constant stepsize, we first give an upper bound on the difference between the objective function, evaluated at the agents' estimates of the optimal decision vector, and the optimal value. In addition, for a particular class of convex functions, we give an upper bound on the distances between the agents' estimates of the optimal decision vector and the minimizer and we provide the rate of convergence to zero of the time varying component of the aforementioned upper

bound. The addressed metrics are evaluated via their expected values. As an application, we show how the distributed optimization algorithm can be used to perform collaborative system identification and provide numerical experiments under the randomized and broadcast gossip protocols.

Second, we generalize the asymptotic consensus problem to convex metric spaces. Under minimal connectivity assumptions, we show that if at each iteration an agent updates its state by choosing a point from a particular subset of the generalized convex hull generated by the agents current state and the states of its neighbors, then agreement is achieved asymptotically. In addition, we give bounds on the distance between the consensus point(s) and the initial values of the agents. As an application example, we introduce a probabilistic algorithm for reaching consensus of opinion and show that it in fact fits our general framework.

Third, we discuss the linear asymptotic consensus problem for a network of dynamic agents whose communication network is modeled by a randomly switching graph. The switching is determined by a finite state, Markov process, each topology corresponding to a state of the process. We address both the cases where the dynamics of the agents are expressed in continuous and discrete time. We show that, if the consensus matrices are doubly stochastic, average consensus is achieved in the mean square and almost sure senses if and only if the graph resulting from the union of graphs corresponding to the states of the Markov process is strongly connected.

Fourth, we address the consensus-based distributed linear filtering problem, where a discrete time, linear stochastic process is observed by a network of sensors. We assume that the consensus weights are known and we first provide sufficient conditions under

which the stochastic process is detectable, i.e. for a specific choice of consensus weights there exists a set of filtering gains such that the dynamics of the estimation errors (without noise) are asymptotically stable. Next, we develop a distributed, sub-optimal filtering scheme based on minimizing an upper bound on a quadratic filtering cost. In the stationary case, we provide sufficient conditions under which this scheme converges; conditions expressed in terms of the convergence properties of a set of coupled Riccati equations. We continue by presenting a connection between the consensus-based distributed linear filter and the optimal linear filter of a Markovian jump linear system, appropriately defined. More specifically, we show that if the Markovian jump linear system is (mean square) detectable, then the stochastic process is detectable under the consensus-based distributed linear filtering scheme. We also show that the optimal gains of a linear filter for estimating the state of a Markovian jump linear system, appropriately defined, can be used to approximate the optimal gains of the consensus-based linear filter.

GENERALIZED DISTRIBUTED CONSENSUS-BASED
ALGORITHMS FOR UNCERTAIN SYSTEMS AND NETWORKS

by

Ion Matei

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Professor John S. Baras, Chair/Advisor
Professor Nuno C. Martins, Co-Chair/Co-Advisor
Professor P.S. Krishnaprasad
Professor William S. Levine
Professor Satyandra K. Gupta, Dean's Representative

Dedication

To my family.

Acknowledgments

I owe my gratitude to all the people who have made this thesis possible and thanks to whom my graduate experience has been one that I will always remember. First and foremost, I am grateful to my advisor and co-advisor, Prof. John S. Baras and Prof. Nuno C. Martins, for their continuous support and encouragement during the entire period of my graduate studies at the University of Maryland. Their energy, enthusiasm, persistence, deep mathematical insight and expertise in various topics has always been a constant source of motivation for me. The fact that they foster a research environment free of constraints, where a student is encouraged to explore different research areas and problems, allowed me to work on various problems which introduced me to numerous mathematical tools and methodologies. This involvement helped me build a strong mathematical background and increased my research maturity. Besides the academic dimension of Profs. Baras and Martins, I also appreciate immensely their human dimension. The constant attention, support and encouragement shown to their students made me appreciate even more the place I spent my years as a graduate student. I am also grateful to other committee members, Profs. P.S. Krishnaprasad, William S. Levine and Satyandra K. Gupta for agreeing to serve on my committee.

During my graduate studies at the University of Maryland I had the privilege to take classes with some exceptional teachers. I would like to take this opportunity to thank Profs. P. S. Krishnaprasad, Steve Marcus and Andre Tits. Their dedication to teaching and the extra effort they always made to organize and present the course material in an interesting manner made the task of learning more enjoyable. Their deep knowledge,

experience, intuition, relaxed style and graciousness towards their students had a profound impact.

The life of a graduate student involves dealing with bureaucratic issues from time to time. I would like to thank Kim Edwards for her efficiency in keeping all these administrative details to an absolute minimum. I would also like to thank the staff of both ECE and ISR for always trying to do their best helping students with official matters.

A special thanks goes to my friend and colleague George Papageorgiou for his help with the review of the manuscript and his very useful suggestions for improving it and to Prof. Andre Tits for his recommendation on how to improve Chapter 3. I would also like to thank a number of colleagues that made the long hours in the office seem not so bad after all: George Papageorgiou, Pedram Hovareshti, Punyaslok Purkayastha, Tao Jiang, Senni Perumal, Kiran Somasundaram, George Theodorakopoulos, Kaustubh Jain, Vahid Tabatabaee, Ayan Roy-Chowdhury and Vladimir Ivanov.

Table of Contents

List of Figures

Chapter 1

Introduction

This chapter serves as an introduction to the rest of the thesis, by providing the motivation for the current work. Moreover, it introduces the problems that are addressed and our contributions.

1.1   Motivation

In the following chapters we address problems related to multi-agent optimization and filtering. We design and analyze distributed algorithms which are based on the consensus/agreement asymptotic algorithm for performing localized (i.e. using only information from neighbors) computations. A consensus problem consists of a group of dynamic agents who seek to agree upon certain quantities of interest by exchanging information among them according to a set of rules. This problem can model many phenomena involving information exchange between agents such as cooperative control of vehicles, formation control, flocking, synchronization, parallel computing, etc. Distributed computation over networks has a long history in control theory starting with the work of Borkar and Varaiya [5], Tsitsikils, Bertsekas and Athans [51, 52] on asynchronous agreement problems and parallel computing. A theoretical framework for solving consensus problems was introduced by Olfati-Saber and Murray in [42, 43], while Jadbabaie et al. studied alignment problems [18] for reaching an agreement. Relevant extensions of the consen-

sus problem were done by Ren and Beard [39], by Moreau in [29] or, more recently, by Nedic and Ozdaglar in [32, 33] or by Olshevsky and Tsitsiklis in [36].

Typically agents are connected via a network that changes with time due to link failures, packet drops, node failure, etc. Such variations in topology can happen randomly which motivates the investigation of consensus problems under a stochastic framework. Hatano and Mesbahi consider in [17] an agreement problem over random information networks, where the existence of an information channel between a pair of elements at each time instance is probabilistic and independent of other channels. In [38], Porfiri and Stilwell provide sufficient conditions for reaching consensus almost surely in the case of a discrete linear system, where the communication flow is given by a directed graph derived from a random graph process, independent of other time instances. Under a similar model of the communication topology, Tahbaz-Salehi and Jadbabaie give necessary and sufficient conditions for almost sure convergence to consensus in [44], while in [45], the authors extend the applicability of their necessary and sufficient conditions to strictly stationary ergodic random graphs.

The consensus algorithm proves to be a useful tool for solving distributively optimization and estimation problems. Multi-agent distributed optimization problems appear naturally in many distributed processing problems (such as network resource allocation, collaborative control and estimation, etc.), where the optimization cost is a convex function which is not necessarily separable. A distributed subgradient method for multi-agent optimization of a sum of convex functions was proposed in [33], where each agent has only local knowledge of the optimization cost, i.e. knows only one term of the sum. The agents exchange information according to a communication topology, modeled as an

undirected, time varying graph, which defines the communication neighborhoods of the agents. The agents maintain *estimates* of the optimal decision vector, which are updated in two stages. The first stage consists of a consensus step among the estimates of an agent and its neighbors. In the second stage, the result of the consensus step is updated in the direction of a subgradient of the local knowledge of the optimization cost. Another multi-agent subgradient method was proposed in [20], where the communication topology is assumed time invariant and where the order of the two stages mentioned above is inverted.

A fundamental problem in sensor networks is developing distributed algorithms for the state estimation of a process of interest. Generically, a process is observed by a group of (mobile) sensors organized in a network. The goal of each sensor is to compute accurate state estimates. The distributed filtering (estimation) problem has received a lot of attention during the past thirty years. An important contribution was made by Borkar and Varaiya [5], who addressed the distributed estimation problem of a random variable by a group of sensors. The particularity of their formulation is that both estimates and measurements are shared among neighboring sensors. The authors show that if the sensors form a communication ring, through which information is exchanged infinitely often, then the estimates converge asymptotically to the same value, i.e. they asymptotically agree. An extension of the results in reference [5] is given in [50]. The recent technological advances in mobile sensor networks have re-ignited the interest in the distributed estimation problem. Most papers focusing on distributed estimation propose different mechanisms for combining the Kalman filter with a consensus filter in order to ensure that the estimates asymptotically converge to the same value, schemes which will be henceforth

called consensus based distributed filtering (estimation) algorithms. In [41] and [40], several algorithms based on the idea mentioned above are introduced. In [8], the authors study the interaction between the consensus matrix, the number of messages exchanged per sampling time, and the Kalman gain for scalar systems. It is shown that optimizing the consensus matrix for fastest convergence and using the centralized optimal gain is not necessarily the optimal strategy if the number of exchanged messages per sampling time is small. In [48], the weights are adaptively updated to minimize the variance of the estimation error. Both the estimation and the parameter optimization are performed in a distributed manner. The authors derive an upper bound on the error variance at each node which decreases with the number of neighboring nodes.

1.2   Contributions of the thesis

Our contributions are as follows. In Chapter 2 we study the performance metrics (rate of convergence and guaranteed region of convergence) of the consensus-based, multi-agent subgradient method proposed in [33], for the case of a constant stepsize. The communication among agents is modeled by a random graph, independent of other time instances, and the performance metrics are viewed in the expectation sense. Random graphs are suitable models for networks that change with time due to link failures, packet drops, node failure, etc. Our focus is on providing upper bounds on the performance metrics, which explicitly depend on the probability distribution of the random graph. The explicit dependence on the probability distribution allows us to determine the optimal probability distributions in the sense that they would ensure the best guaranteed upper

bounds on the performance metric. As an example of possible applications of our results, we address a scenario where the goal is to tune the communication protocol parameters of a wireless network so that the performance of the multi-agent subgradient method is improved, in the context of a distributed parametric system identification application.

In Chapter 2 we emphasize the effect and importance of the agreement step in solving an optimization problem distributively. It is often the case that we need to solve optimization problems that go beyond the $\mathbb{R}^n$ setup. In [47], the authors formulate optimization problems for the *trusted routing problem* routing under a semiring framework. In [28, 27], the popular *particle swarm optimization algorithm* is extended to combinatorial spaces, such as Euclidean, Manhattan, and Hamming spaces. Related to the distributed optimization algorithm introduced in Chapter 2, a first step to extend the applicability of the algorithm is to formulate and analyze the agreement problem in more general spaces. Consequently, in Chapter 3 we generalize the asymptotic consensus problem to the more general case of convex metric spaces and emphasize the fundamental role of the generalized notion of convexity and in particular of the generalized convex hull of a finite set of points. Tsitsiklis showed in [51] that, under some minimal connectivity assumptions on the communication network, if an agent updates its value by choosing a point from the (interior) of the convex hull of its current value and the current values of its neighbors, then asymptotic convergence to consensus is achieved. We will show that this idea extends naturally to the case of convex metric spaces. As an application we present a probabilistic consensus of opinion algorithm and show that it fits our general framework for a particular convex metric space.

In Chapter 2 we assume that the communication topology, which dictates how the

consensus step is performed, is modeled by a random graph, independent of other time instances. In Chapter 4, we generalize the communication model and study the linear consensus problem where the communication flow between agents is modeled by a (possibly directed) switching random graph. The switching is determined by a homogeneous, finite-state Markov chain, each communication pattern corresponding to a state of the Markov process. We address both the continuous and discrete time cases and, under certain assumptions on the matrices involved in the linear scheme, we give necessary and sufficient conditions such that average consensus is achieved in the mean square sense and in the almost sure sense. The Markovian switching model goes beyond the common i.i.d. assumption on the random communication topology and appears in the cases where Rayleigh fading channels are considered. Our aim is to show how mathematical techniques used in the stability analysis of Markovian jump linear systems, together with results inspired by matrix and graph theory, can be used to prove (intuitively clear) convergence results for the (linear) stochastic consensus problem.

In Chapter 5 we address the consensus-based distributed linear filtering problem. We assume that each agent updates its (local) estimate in two steps. In the first step, an update is produced using a Luenberger observer type of filter. In the second step, called the *consensus step*, every sensor computes a convex combination between its local update and the updates received from the neighboring sensors. For given consensus weights, we will first give sufficient conditions for the existence of filter gains such that the dynamics of the estimation errors (without noise) are asymptotically stable. Next, we present a distributed, sub-optimal filtering algorithm, valid for time varying topologies as well, resulting from minimizing an upper bound on a quadratic cost expressed in terms of the

covariances matrices of the estimation errors. We will also present a connection between the consensus-based linear filter and the linear filtering of a Markovian jump linear system appropriately defined, a connection which was inspired by our previous work on state estimation for switching systems (see for instance [24], [25]).

Chapter 2

Distributed Optimization under Random Communication Topologies

2.1   Introduction

We investigate the collaborative optimization problem in a multi-agent setting, when the agents make decisions in a distributed manner using local information, while the communication topology used to exchange messages and information is modeled by a graph-valued random process, assumed independent and identically distributed (i.i.d.). Specifically, we study the performance of the consensus-based multi-agent distributed subgradient method proposed in [33], for the case of a constant stepsize.

Random graphs are suitable models for networks that change with time due to link failures, packet drops, node failures, etc. An analysis of the multi-agent subgradient method under random communication topologies is addressed in [22]. The authors assume that the consensus weights are lower bounded by some positive scalar and give upper bounds on the performance metrics as functions of this scalar and other parameters of the problem. More precisely, the authors give upper bounds on the distance between the cost function and the optimal solution (in expectation), where the cost is evaluated at the (weighted) time average of the optimal decision vector's estimate. *Our main goal is to provide upper bounds on the performance metrics, which explicitly depend on the probability distribution of the random graph.* We first derive an upper bound on the difference between the cost function, evaluated at the estimate, and the optimal value.

Next, for a particular class of convex functions, we focus on the distance between the estimate of the optimal decision and the minimizer. The upper bound we provide has a constant component and a time varying component. For the latter, we provide the rate of convergence to zero. The performance metrics are evaluated via their expected values. The explicit dependence on the graph's probability distribution may be useful to design probability distributions that would ensure the best guaranteed upper bounds on the performance metrics. This idea has relevance especially in the wireless networks, where the communication topology has a random nature with a probability distribution (partially) determined by the communication protocol parameters (the reader can consult [21, 35], where the authors introduce probabilistic models for successful transmissions as functions of the transmission powers). As an example of possible application, we show how the distributed optimization algorithm can be used to perform collaborative system identification and we present numerical experiments results under the randomized [7] and broadcast [1] gossip protocols. Similar performance metrics as our are studied in [2], where the authors generalizes the randomized incremental subgradient method and where the stochastic component in the algorithm is described by a Markov chain, which can be constructed in a distributed fashion using local information only. Newer results on the distributed optimization problem can be found in [13], where the authors analyze distributed algorithms based on dual averaging of subgradients, and provide sharp bounds on their convergence rates as a function of the network size and topology.

Notations: Let X be a subset of $\mathbb{R}^n$ and let y be a point in $\mathbb{R}^n$. By slight abuse of notation, let $\|y - X\|$ denote the distance from the point y to the set X, i.e. $\|y - X\| \triangleq \min_{x \in X} \|y - x\|$, where $\|\cdot\|$ is the standard Euclidean norm. For a twice differentiable func-

9

tion $f(x)$, we denote by $\nabla f(x)$ and $\nabla^2 f(x)$ the gradient and Hessian of f at x, respectively. Given a symmetric matrix A, by $(A \geq 0)$ $A > 0$ we understand A is positive (semi) definite. The symbol $\otimes$ represents the Kronecker product.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. We denote by $\partial f(x)$ the subdifferential of f at x, i.e. the set of all subgradients of f at x:

$$\partial f(x) = \{d \in \mathbb{R}^n | f(y) \geq f(x) + d'(y - x), \ \forall y \in \mathbb{R}^n\}. \tag{2.1}$$

Let $\epsilon \geq 0$ be a nonnegative real number. We denote by $\partial_\epsilon f(x)$ the ϵ-subdifferential of f at x, i.e. the set of all ϵ-subgradients of f at x:

$$\partial_\epsilon f(x) = \{d \in \mathbb{R}^n | f(y) \geq f(x) + d'(y - x) - \epsilon, \ \forall y \in \mathbb{R}^n\}. \tag{2.2}$$

The gradient of the differentiable function $f(x)$ on $\mathbb{R}^n$ satisfies a *Lipschitz condition with constant L* if

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \ \forall x, y \in \mathbb{R}^n.$$

The differentiable, convex function $f(x)$ on $\mathbb{R}^n$ is *strongly convex with constant l* if

$$f(y) \geq f(x) + \nabla f(x)'(y - x) + \frac{l}{2}\|y - x\|^2, \ \forall x, y \in \mathbb{R}^n.$$

We will denote by LEM and SLEM the largest and second largest eigenvalue in modulus of a matrix, respectively. We will use CBMASM as the abbreviation for **C**onsensus-**B**ased **M**ulti-**A**gent **S**ubgradient **M**ethod and pmf for probability mass function.

Chapter structure: Section 2.2 contains the problem formulation. In Section 2.3 we introduce a set of preliminary results, which mainly consist of providing upper bounds for a number a quantities of interest. Using these preliminary results, in Section 2.4 we give

upper bounds for the expected value of two performance metrics: the distance between the cost function evaluated at the estimate and the optimal solution and the (squared) distance between the estimate and the minimizer. Section 2.5 shows how the distributed optimization algorithm can be used for collaborative system identification.

2.2 Problem formulation

2.2.1 Communication model

Consider a network of N agents, indexed by $i = 1, \ldots, N$. The communication topology is time varying and is modeled by a random graph $G(k) = (V, \mathcal{E}(k))$, where V is the set of N vertices (nodes) and $\mathcal{E}(k) = (e_{ij}(k))$ is the set of edges, and where we used k to denote the time index. The edges in the set $\mathcal{E}(k)$ correspond to the communication links among agents. Given a positive integer M, the graph $G(k)$ takes values in a finite set $\mathcal{G} = \{G_1, G_2, \ldots, G_M\}$ at each k, where the graphs $G_i = (V, \mathcal{E}_i)$ are assumed *undirected* and *without self loops*. In other words, we will consider only bidirectional communication topologies. The underlying random process of $G(k)$ is assumed i.i.d. with probability distribution $Pr(G(k) = G_i) = p_i$, $\forall k \geq 0$, where $\sum_{i=1}^{M} p_i = 1$ and $p_i > 0$.

Assumption 2.2.1. *(Connectivity assumption) The graph $\bar{G} = (V, \bar{E})$ resulting from the union of all graphs in the $\mathcal{G}$ is* connected, *where*

$$\bar{G} = \bigcup_{i=1}^{M} G_i = \left(V, \bigcup_{i=1}^{M} \mathcal{E}_i \right).$$

Let G be an undirected graph with N nodes and no self loops and let $A \in \mathbb{R}^{N \times N}$ be a row stochastic matrix, with positive diagonal entries. We say that the matrix A

corresponds to the graph G or the graph G is *induced* by A if any non-zero entry (i,j) of A, with $i \neq j$ implies a link from j to i in G and vice-versa.

2.2.2 Optimization model

The task of the N agents consists of minimizing a convex function $f : \mathbb{R}^n \to \mathbb{R}$. The function f is expressed as a sum of N functions, i.e.

$$f(x) = \sum_{i=1}^{N} f_i(x), \tag{2.3}$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ are convex. Formally expressed, the agents want to cooperatively solve the following optimization problem

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{N} f_i(x). \tag{2.4}$$

The fundamental assumption is that each agent i, has access only to the function f_i.

Let f^* denote the optimal value of f and let X^* denote the set of optimizers of f, i.e. $X^* = \{x \in \mathbb{R}^n | f(x) = f^*\}$. Let $x_i(k) \in \mathbb{R}^n$ designate the *estimate of the optimal decision vector* of (2.4), maintained by agent i, at time k. The agents exchange estimates among themselves subject to the communication topology described by the random graph $G(k)$.

As proposed in [33], the agents update their estimates using a modified incremental subgradient method. Compared to the standard subgradient method, the local estimate $x_i(k)$ is replaced by a convex combination of $x_i(k)$ with the estimates received from the neighbors:

$$x_i(k+1) = \sum_{j=1}^{N} a_{ij}(k) x_j(k) - \alpha(k) d_i(k), \tag{2.5}$$

where $a_{ij}(k)$ is the $(i,j)^{th}$ entry of a stochastic random matrix $A(k)$ which corresponds to the communication graph $G(k)$. The matrices $A(k)$ form an i.i.d. random process

taking values in a finite set of *symmetric* stochastic matrices with *positive diagonal entries* $\mathcal{A} = \{A_i\}_{i=1}^M$, where A_i is a stochastic matrix corresponding to the graph $G_i \in \mathcal{G}$, for $i = 1, \ldots, M$. The probability distribution of $A(k)$ is inherited from $G(k)$, i.e. $Pr(A(k) = A_i) = Pr(G(k) = G_i) = p_i$. The real valued scalar $\alpha(k)$ is the stepsize, while the vector $d_i(k) \in \mathbb{R}^n$ is a subgradient of f_i at $x_i(k)$, i.e. $d_i(k) \in \partial f_i(x_i(k))$. Obviously, when $f_i(x)$ are assumed differentiable, $d_i(k)$ becomes the gradient of f_i at $x_i(k)$, i.e. $d_i(k) = \nabla f_i(x_i(k))$.

Note that the first part of equation (2.5) is a consensus step, a problem that has received a lot of attention in recent years, both in a deterministic ([6, 14, 18, 29, 39, 51, 52]) and a stochastic ([17, 23, 44, 45]) framework.

The consensus problem under different gossip algorithms was studied in [1, 7, 12]. We note that there is direct connection between our communication model and the communication models used in the *randomized gossip protocol* [7] and *broadcast communication protocol* [1]. Indeed, in the case of the randomized communication protocol, the set $\mathcal{G}$ is formed by the graphs G_{ij} with only one link (i, j), where $Pr(G(k) = G_{ij}) = \frac{1}{N}P_{ij}$ for some $P_{ij} > 0$ with $\sum_{i=1}^N P_{ij} = 1$, while the set $\mathcal{A}$ is formed by stochastic matrices A_{ij} of the form $A_{ij} = I - \frac{1}{2}(e_i - e_j)(e_i - ej)'$, where vectors the e_i represent the standard basis. In the case of the broadcast communication protocol, the set $\mathcal{G}$ is formed by the graphs G_i, where G_i contains links between the node i and the nodes in its neighborhood, denoted by N_i. The probability distribution of $G(k)$ is given by $Pr(G(k) = G_i) = \frac{1}{N}$ and the set $\mathcal{A}$ is formed by matrices of the form $A_i = I - \delta_i \sum_{j \in N_i}(e_i - e_j)(e_i - ej)'$, for some $0 < \delta_i \le \frac{1}{|N_i|}$.

The following assumptions, which will not necessarily be used simultaneously, introduce properties of the function $f(x)$.

Assumption 2.2.2. *(Non-differentiable functions)*

(a) The subgradients of the functions $f_i(x)$ are uniformly bounded, i.e. there exists a positive scalar φ such that

$$\|d\| \leq \varphi, \forall d \in \partial f_i(x), \ \forall x \in \mathbb{R}^n, \ i = 1, \ldots, N,$$

(b) The stepsize is constant, i.e.

$$\alpha(k) = \alpha, \ \forall k \geq 0,$$

(c) The optimal solution set X^* is nonempty.

Assumption 2.2.3. *(Differentiable functions)*

(a) The functions $f_i(x)$ are twice differentiable on $\mathbb{R}^n$,

(b) There exists positive scalars l_i, L_i such that

$$l_i I \leq \nabla^2 f_i(x) \leq L_i I, \ \forall x \in \mathbb{R}^n \ and \ \forall i,$$

(c) The stepsize is constant, i.e. $\alpha(k) = \alpha$ for all k and satisfies the inequality

$$0 < \alpha < \min\left\{\frac{\underline{\lambda} + 1}{L}, \frac{1}{l}\right\},$$

where $\underline{\lambda}$ is the smallest among all eigenvalues of matrices A_i, $l = \min_i l_i$ and $L = \max_i L_i$.

Assumption 2.2.3 -(b) is satisfied if the gradient of $f_i(x)$ satisfies a Lipschitz condition with constant L_i and if $f_i(x)$ is strongly convex with constant l_i. Also, under Assumptions 2.2.3, X^* has one element which is the unique minimizer of $f(x)$, denote henceforth by x^*.

2.3 Preliminary Results

In this section we lay the foundation for our main results in Section 2.4. The preliminary results introduced here revolve around the idea of providing upper-bounds on a number of quantities of interest. The first quantity is represented by the distance between the estimate of the optimal decision vector and the average of all estimates. The second quantity is described by the distance between the average of all estimates and the minimizer.

We introduce the *average* vector of estimates of the optimal decision vector, denoted by $\bar{x}(k)$ and defined by

$$\bar{x}(k) \triangleq \frac{1}{N} \sum_{i=1}^{N} x_i(k). \tag{2.6}$$

The dynamic equation for the average vector can be derived from (2.5) and takes the form

$$\bar{x}(k+1) = \bar{x}(k) - \frac{\alpha(k)}{N} h(k), \tag{2.7}$$

where $h(k) = \sum_{i=1}^{N} d_i(k)$.

We introduce also the *deviation* of the local estimates $x_i(k)$ from the average estimate $\bar{x}(k)$, which is denoted by $z_i(k)$ and defined by

$$z_i(k) \triangleq x_i(k) - \bar{x}(k), \ i = 1 \dots N. \tag{2.8}$$

and let β be a positive scalar such that

$$\|z_i(0)\| \leq \beta, \ i = 1 \dots N.$$

Let us define the *aggregate* vectors of estimates, average estimates, deviations and (sub)gradients, respectively:

$$\mathbf{x}(k)' \triangleq [x_1(k)', x_2(k)', \dots, x_N(k)'] \in \mathbb{R}^{Nn},$$

15

$$\bar{\mathbf{x}}(k)' \triangleq [\bar{x}(k)', \bar{x}(k)', \dots, \bar{x}(k)'] \in \mathbb{R}^{Nn},$$

$$\mathbf{z}(k)' \triangleq [z_1(k)', z_2(k)', \dots, z_N(k)'] \in \mathbb{R}^{Nn}$$

and

$$\mathbf{d}(k)' \triangleq [d_1(k)', d_2(k)', \dots, d_N(k)'] \in \mathbb{R}^{Nn}.$$

From (2.6) we note that the aggregate vector of average estimates can be expressed as

$$\bar{\mathbf{x}}(k) = \mathbf{J}\mathbf{x}(k),$$

where $\mathbf{J} = \frac{1}{N}\mathbb{1}\mathbb{1}' \otimes I$, with I the identity matrix in $\mathbb{R}^{n \times n}$ and $\mathbb{1}$ the vector of all ones in $\mathbb{R}^N$.

Consequently, the aggregate vector of deviations can be written as

$$\mathbf{z}(k) = (\mathbf{I} - \mathbf{J})\mathbf{x}(k), \tag{2.9}$$

where $\mathbf{I}$ is the identity matrix in $\mathbb{R}^{nN \times nN}$. The next Proposition characterizes the dynamics

of the vector $\mathbf{z}(k)$.

Proposition 2.3.1. *The dynamic evolution of the aggregate vector of deviations is given*

by

$$z(k+1) = W(k)z(k) - \alpha(k)(I - J)d(k), \ z(0) = z_0, \tag{2.10}$$

where $W(k) = A(k) - J$ *and* $A(k) = A(k) \otimes I$, *with solution*

$$z(k) = \Phi(k,0)z(0) - \sum_{s=0}^{k-1} \alpha(s)\Phi(k,s+1)d(s), \tag{2.11}$$

where $\Phi(k,s)$ *is the transition matrix of (2.10) defined by* $\Phi(k,s) \triangleq W(k-1)W(k-2)\cdots W(s)$,

with $\Phi(k,k) = I$.

Proof. From (2.5) the dynamics of the aggregate vector of estimates is given by

$$\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k) - \alpha(k)\mathbf{d}(k). \tag{2.12}$$

From (2.9) together with (2.12), we can further write

$$\mathbf{z}(k+1) = (\mathbf{I} - \mathbf{J})\mathbf{x}(k+1) = (\mathbf{A}(k) - \mathbf{J})\mathbf{x}(k) - \alpha(k)(\mathbf{I} - \mathbf{J})\mathbf{d}(k).$$

By noting that

$$(\mathbf{A}(k) - \mathbf{J})\mathbf{z}(k) = (\mathbf{A}(k) - \mathbf{J})(\mathbf{I} - \mathbf{J})\mathbf{x}(k) = (\mathbf{A}(k) - \mathbf{J})\mathbf{x}(k),$$

we obtain (2.10). The solution (2.11) follows from (2.10) together with the observation that $\Phi(k,s)(\mathbf{I} - \mathbf{J}) = \Phi(k,s)$. $\qquad\square$

Remark 2.3.1. *The transition matrix $\Phi(k,s)$ of the stochastic linear equation (2.10) can also be represented as*

$$\Phi(k,s) = \left(\prod_{i=1}^{s} A(k-i)\right) - \mathbf{J}, \tag{2.13}$$

where $\mathbf{J} = \left(\frac{1}{N}\mathbb{1}\mathbb{1}'\right) \otimes I$. This follows from the fact that for any $i \in \{1, 2, \ldots, s-1\}$ we have

$$(A(k-i) - \mathbf{J})(A(k-i-1) - \mathbf{J}) = A(k-i)A(k-i-1) - \mathbf{J}.$$

Remark 2.3.2 (On the first and second moments of the transition matrix $\Phi(k,s)$). *Let m be a positive integer and consider the transition matrix $\Phi(k+m,k) = W(k+m-1)\ldots W(k)$, generated by a sequence of length m of random graphs, i.e. $G(k)\ldots G(k+m-1)$, for some $k \geq 0$. The random matrix $\Phi(k+m,k)$ takes values of the form $W_{i_1} W_{i_2} \cdots W_{i_m}$, with $i_j \in \{1, 2, \ldots, M\}$ and $j = 1, \ldots, m$. The norm of a particular realization of $\Phi(k+m,k)$ is given by the LEM of the matrix product $W_{i_1} W_{i_2} \cdots W_{i_m}$ or the*

17

SLEM of $A_{i_1} A_{i_2} \cdots A_{i_m}$, denoted henceforth by $\lambda_{i_1 \ldots i_m}$. Let $q_{i_1 \ldots i_m} = \prod_{j=1}^m p_{i_j}$ be the probability of the sequence of graphs $G_{i_1} \ldots G_{i_m}$ that appear during the time interval $[k, k+m]$. Let I_m be the set of sequences of indices of length m for which the union of graphs with the respective indices produces a connected graph, i.e. $I_m = \{ i_1 i_2 \ldots i_m | \bigcup_{j=1}^m G_{i_j} = \text{connected} \}$. Using the previous notations, the first and second moments of the norm of $\Phi(k+m, k)$ can be expressed as

$$E[\|\Phi(k+m, k)\|] = \eta_m, \tag{2.14}$$

$$E[\|\Phi(k+m, k)\|^2] = \rho_m, \tag{2.15}$$

where $\eta_m = \sum_{j \in I_m} q_j \lambda_j + 1 - \sum_{j \in I_m} q_j$ and $\rho_m = \sum_{j \in I_m} q_j \lambda_j^2 + 1 - \sum_{j \in I_m} q_j$. The integer j was used as an index for the elements of set I_m, i.e. for an element of the form $i_1 \ldots i_m$.

The above formulas follow from results introduced in [18], Lemma 1, or in [39], Lemma 3.9, which state that for any sequence of indices $i_1 \ldots i_m \in I_m$, the matrix product $A_{i_1} \cdots A_{i_m}$ is ergodic, and therefore $\lambda_j < 1$, for any $j \in I_m$. Conversely, if $j \notin I_m$ then $\lambda_j = 1$. We also note that $\sum_{j \in I_m} q_j$ is the probability of having a connected graph over a time interval of length m. Due to Assumption 2.2.1, for sufficiently large values of m, the set I_m is nonempty. In fact for $m \geq M$, I_m is always non-empty. Therefore, for any m such that I_m is not empty, we have that $0 < \rho_m < \eta_m < 1$. *In general for large values of m, it may be difficult to compute all eigenvalues λ_j, $j \in I_m$. We can omit the necessity of computing the eigenvalues λ_j, and this way decrease the computational burden, by using the following upper bounds on η_m and ρ_m*

$$\eta_m \leq \lambda_m \boldsymbol{p}_m + 1 - \boldsymbol{p}_m, \tag{2.16}$$

$$\rho_m \leq \lambda_m^2 \boldsymbol{p}_m + 1 - \boldsymbol{p}_m, \tag{2.17}$$

18

where $\lambda_m = \max_{j \in I_m} \lambda_j$ and $\boldsymbol{p}_m = \sum_{j \in I_m} q_j$ is the probability to have a connected graph over a time interval of length m. For notational simplicity, in what follows we will omit the index m when referring to the scalars η_m and ρ_m.

Throughout this chapter we will use the symbols m, η and ρ in the sense defined within the Remark 2.3.2. Moreover, the value of m is chosen such that I_m is nonempty. The existence of such a value is guaranteed by Assumption 2.2.1.

The next proposition gives upper bounds on the expected values of the norm and the squared norm of the transition matrix $\Phi(k, s)$.

Proposition 2.3.2. *Let Assumption 2.2.1 hold, and let $r \leq s \leq k$ be three nonnegative integer values and m a positive integer, such that the set I_m is non-empty. Then, the following inequalities involving the transition matrix $\Phi(k, s)$ of (2.10), hold*

$$E[\|\Phi(k, s)\|] \leq \eta^{\left\lfloor \frac{k-s}{m} \right\rfloor}, \tag{2.18}$$

$$E[\|\Phi(k, s)\|^2] \leq \rho^{\left\lfloor \frac{k-s}{m} \right\rfloor}, \tag{2.19}$$

$$E[\|\Phi(k, r)\Phi(k, s)'\|] \leq \rho^{\left\lfloor \frac{k-s}{m} \right\rfloor} \eta^{\left\lfloor \frac{s-r}{m} \right\rfloor}, \tag{2.20}$$

where η and ρ are defined in Remark 2.3.2.

Proof. We fix an m such that the probability of having a connected graph over a time interval of length m is positive, i.e. I_m is non-empty. Note that, by Assumption 2.2.1, such a value always exists (pick $m \geq M$). Let t be the number of intervals of length m between s and k, i.e.

$$t = \left\lfloor \frac{k - s}{m} \right\rfloor,$$

and let $s_0, s_1, \ldots, s_t$ be a sequence of nonnegative integers such that $s = s_0 < s_1 < \ldots < s_t \le k$

where $s_{i+1} - s_i = m$ and $i = 0, \ldots, m-1$. By the semigroup property of transition matrices,

it follows that

$$\Phi(k, s) = \Phi(k, s_t)\Phi(s_t, s_{t-1}) \cdots \Phi(s_1, s),$$

or

$$\|\Phi(k, s)\| \le \|\Phi(s_t, s_{t-1})\| \cdots \|\Phi(s_1, s)\|,$$

where we use the fact that $\|\Phi(k, s_t)\| \le 1$. Using the i.i.d. assumption on the random

process $\mathbf{A}(k)$, we can further write

$$E[\|\Phi(k, s)\|] \le E[\|\Phi(s_t, s_{t-1})\|] \cdots E[\|\Phi(s_1, s)\|],$$

which together with (2.14) leads to inequality (2.18).

Similarly, inequality (2.19) follows from (2.15) and from the i.i.d. assumption on

the random graph process.

We now turn to inequality (2.20). By the semigroup property we get

$$E[\|\Phi(k, r)\Phi(k, s)'\|] \le E[\|\Phi(k, s)\|^2 \|\Phi(s, r)\|] \le E[\|\Phi(k, s)\|^2]E[\|\Phi(s, r)\|],$$

where the second inequality followed by the independence of $\mathbf{A}(k)$. Inequality (2.20)

follows from (2.18) and (2.19). $\qquad\square$

In the next lemma we show that, under Assumption 2.2.3, for small enough α the

gradients $\nabla f_i(x_i(k))$ remain bounded with probability one for all k.

Lemma 2.3.1. *Let Assumption 2.2.3 hold and let $\mathcal{F} : \mathbb{R}^{Nn} \to \mathbb{R}$ be a function given by*

$\mathcal{F}(\boldsymbol{x}) = \sum_{i=1}^{N} f_i(x_i)$ *where* $\boldsymbol{x}' = (x_1', \ldots, x_N')$. *There exists a positive scalar φ such that*

$$\|\nabla f_i(x_i(k))\| \le \varphi, \forall \ i, k \ w.p. \ 1,$$

$$\|\nabla f_i(\bar{x}(k))\| \leq \varphi, \forall\, i, k \text{ w.p. } 1,$$

where $\varphi = L\|\boldsymbol{x}(0) - \tilde{\boldsymbol{x}}\| + L\left(\frac{2}{1-q} + 1\right)\|\tilde{\boldsymbol{x}}\|$, $q = \max\{|\underline{\lambda} - \alpha L|, |1 - \alpha l|\}$, $\tilde{\boldsymbol{x}}$ is the unique minimizer of $\mathcal{F}(\boldsymbol{x})$, and $x_i(k)$ and $\bar{x}(k)$ satisfy (2.5) and (2.7), respectively.

Proof. We first note that since the matrices A_i have positive diagonal entries, they are aperiodic and therefore $\underline{\lambda} \in (-1\ 1]$. From Assumption 2.2.3 it follows immediately that $\mathcal{F}(\mathbf{x})$ is a convex, twice differentiable function satisfying

$$l\mathbf{I} \leq \nabla^2 \mathcal{F}(\mathbf{x}) \leq L\mathbf{I}, \tag{2.21}$$

where $l = \min_i l_i$, $L = \max_i L_i$ and $\mathbf{I}$ is the identity matrix in $\mathbb{R}^{nN \times nN}$. In addition, $\mathcal{F}(\mathbf{x})$ has a unique minimizer denoted by $\tilde{\mathbf{x}}$. The dynamics described by (2.5) can be compactly written as

$$\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k) - \alpha\nabla\mathcal{F}(\mathbf{x}(k)),\ \mathbf{x}(0) = \mathbf{x}_0, \tag{2.22}$$

with $\mathbf{x}(k)' = (x_1(k)', \ldots, x_N(k)')$.

We observe that equation (2.22) is a modified version of the gradient method with constant step, where instead of the identity matrix, we have that $\mathbf{A}(k)$ multiplies $\mathbf{x}(k)$. In what follows we show that the stochastic dynamics (2.22) is stable with probability one.

Using a similar idea as in Theorem 3, page 25 of [37], we have that

$$\nabla\mathcal{F}(\mathbf{x}(k)) = \nabla\mathcal{F}(\tilde{\mathbf{x}}) + \int_0^1 \nabla^2\mathcal{F}(\tilde{\mathbf{x}} + \tau(\mathbf{x}(k) - \tilde{\mathbf{x}}))(\mathbf{x}(k) - \tilde{\mathbf{x}})d\tau = \mathcal{H}(k)(\mathbf{x}(k) - \tilde{\mathbf{x}}),$$

where $l\mathbf{I} \leq \mathcal{H}(k) \leq L\mathbf{I}$ by virtue of (2.21). Hence, with probability one

$$\|\mathbf{x}(k+1) - \tilde{\mathbf{x}}\| = \|\mathbf{A}(k)\mathbf{x}(k) - \tilde{\mathbf{x}} - \alpha\nabla\mathcal{F}(\mathbf{x}(k)) + \mathbf{A}(k)\tilde{\mathbf{x}} - \mathbf{A}(k)\tilde{\mathbf{x}}\| \leq$$

$$\leq \|\mathbf{A}(k) - \alpha\mathcal{H}(k)\|\,\|\mathbf{x}(k) - \tilde{\mathbf{x}}\| + \|\mathbf{A}(k) - \mathbf{I}\|\,\|\tilde{\mathbf{x}}\|.$$

But since

$$(\underline{\lambda} - \alpha L)\mathbf{I} \leq \mathbf{A}(k) - \mathcal{H}(k) \leq (1 - \alpha l)\mathbf{I},$$

it follows that

$$\|\mathbf{x}(k+1) - \tilde{\mathbf{x}}\| \leq q\|\mathbf{x}(k) - \tilde{\mathbf{x}}\| + |\underline{\lambda} - 1|\|\tilde{\mathbf{x}}\|,$$

where $q = \max\{|\underline{\lambda} - \alpha L|, |1 - \alpha l|\}$. Since by Assumption 2.2.3-(c) $\alpha < \min\left\{\frac{\lambda+1}{L}, \frac{1}{l}\right\}$ we get that $q < 1$ and therefore the dynamics (2.22) is stable with probability one and

$$\|\mathbf{x}(k) - \tilde{\mathbf{x}}\| \leq q^k\|\mathbf{x}(0) - \tilde{\mathbf{x}}\| + \frac{2}{1-q}\|\tilde{\mathbf{x}}\| \leq \|\mathbf{x}(0) - \tilde{\mathbf{x}}\| + \frac{2}{1-q}\|\tilde{\mathbf{x}}\|, \ \forall k.$$

From Assumption 2.2.3 we have that

$$\|\nabla f_i(x_i(k))\| \leq \|\nabla \mathcal{F}(\mathbf{x}(k))\| \leq L\|\mathbf{x}(k) - \tilde{\mathbf{x}}\| \leq L\|\mathbf{x}(0) - \tilde{\mathbf{x}}\| + \frac{2L}{1-q}\|\tilde{\mathbf{x}}\|. \tag{2.23}$$

We also have that

$$\|\bar{\mathbf{x}}(k) - \tilde{\mathbf{x}}\| = \|\mathbf{J}\mathbf{x}(k) - \mathbf{J}\tilde{\mathbf{x}} + \mathbf{J}\tilde{\mathbf{x}} - \tilde{\mathbf{x}}\| \leq \|\mathbf{x}(k) - \tilde{\mathbf{x}}\| + \|\tilde{\mathbf{x}}\|,$$

from where it follows that

$$\|\nabla f_i(\bar{x}(k))\| \leq \|\nabla \mathcal{F}(\bar{\mathbf{x}}(k))\| \leq L\|\bar{\mathbf{x}}(k) - \tilde{\mathbf{x}}\| \leq L\|\mathbf{x}(0) - \tilde{\mathbf{x}}\| + L\left(\frac{2}{1-q} + 1\right)\|\tilde{\mathbf{x}}\|. \tag{2.24}$$

Taking the maximum among the right hand side terms of the inequalities (2.23) and (2.24), the result follows. □

Remark 2.3.3. *If the stochastic matrices A_i are generated using a Laplacian based scheme, e.g.*

$$A_i = I - \varepsilon \mathcal{L}_i, \forall i,$$

22

where $\mathcal{L}_i$ is the Laplacian of the graph G_i and $\varepsilon \leq \frac{1}{N}$, then it turns out that $\underline{\lambda} \geq 0$. Hence, the inequality in Assumption 2.2.3-(c) is satisfied if

$$0 < \alpha < \frac{1}{L},$$

which is a sufficient condition for the stability of (2.5). In the case of the randomized and broadcast gossip protocols it can be checked that $\underline{\lambda} = 0$.

Remark 2.3.4. *Throughout the rest of the chapter φ should be interpreted in the context of the assumptions used, i.e. under Assumption 2.2.2, φ is the uniform bound on the subgradients of $f_i(x)$, while under Assumption 2.2.3, φ is the bound on the gradients $\nabla f_i(x_i(k))$ and $\nabla f_i(\bar{x}(k))$ given by Lemma 2.3.1.*

The following lemma gives upper bounds on the first and the second moments of the distance between the estimate $x_i(k)$ and the average of the estimates, $\bar{x}(k)$.

Lemma 2.3.2. *Under Assumptions 2.2.1 and 2.2.2 or 2.2.1 and 2.2.3, for the sequences $\{x_i(k)\}_{k \geq 0}$, $i = 1, \ldots, N$ generated by (2.5) with a constant stepsize α, the following inequalities hold*

$$E[\|x_i(k) - \bar{x}(k)\|] \leq \beta \sqrt{N} \eta^{\lfloor \frac{k}{m} \rfloor} + \alpha \varphi \sqrt{N} \frac{m}{1 - \eta} \tag{2.25}$$

$$E[\|x_i(k) - \bar{x}(k)\|^2] \leq N\beta^2 \rho^{\lfloor \frac{k}{m} \rfloor} + N\alpha^2 \varphi^2 \left(1 + 2\frac{m}{1-\eta}\right) \frac{m}{1-\rho} + 2N\alpha\beta\varphi m \frac{\rho^{\lfloor \frac{k-1}{m} \rfloor + 1} - \eta^{\lfloor \frac{k-1}{m} \rfloor + 1}}{\rho - \eta},$$

$$\tag{2.26}$$

where η, ρ and m are defined in Remark 2.3.2.

Proof. Note that the norm of the deviation $z_i(k) = x_i(k) - \bar{x}(k)$ is upper bounded by the norm of the aggregate vector of deviations $\mathbf{z}(k)$ (with probability one), i.e. $\|z_i(k)\| \leq \|\mathbf{z}(k)\|$.

Hence, by Proposition 2.3.1, we have

$$\|z_i(k)\| \le \|\mathbf{z}(k)\| = \|\Phi(k,0)\mathbf{z}(0) - \alpha \sum_{s=0}^{k-1} \Phi(k,s+1)\mathbf{d}(s)\|,$$

or

$$E[\|z_i(k)\|] \le \beta \sqrt{N} E[\|\Phi(k,0)\|] + \alpha\varphi \sqrt{N} \sum_{s=0}^{k-1} E[\|\Phi(k,s+1)\|],$$

where we used the fact that $\|z_i(0)\| \le \beta$ and $\|d_i(k)\| \le \varphi$, $\forall k \ge 0$.

By inequality (2.18) of Proposition 2.3.2, we get

$$E[\|z_i(k)\|] \le \beta \sqrt{N} \eta^{\lfloor \frac{k}{m} \rfloor} + \alpha\varphi \sqrt{N} \sum_{s=0}^{k-1} \eta^{\lfloor \frac{k-s-1}{m} \rfloor}.$$

Noting that the sum $\sum_{s=0}^{k-1} \eta^{\lfloor \frac{k-s-1}{m} \rfloor}$ can be upper bounded by

$$\sum_{s=0}^{k-1} \eta^{\lfloor \frac{k-s-1}{m} \rfloor} \le m \sum_{s=0}^{\lfloor \frac{k-1}{m} \rfloor} \eta^s = m \frac{1 - \eta^{\lfloor \frac{k-1}{m} \rfloor + 1}}{1 - \eta} \le m \frac{1}{1 - \eta},$$

inequality (2.25) follows.

We now turn to obtaining an upper bound on the second moment of $\|\mathbf{z}(k)\|$.

Let $\mathbf{Z}(k) \in \mathbb{R}^{Nn \times Nn}$ be the symmetric, semi-positive definite matrix defined by

$$\mathbf{Z}(k) \triangleq \mathbf{z}(k)\mathbf{z}(k)'.$$

Using Proposition 2.3.1, it follows that $\mathbf{Z}(k)$ satisfies the following dynamic equation

$$\mathbf{Z}(k+1) = \mathbf{W}(k)\mathbf{Z}(k)\mathbf{W}(k)' + \mathbf{F}(k), \tag{2.27}$$

where $\mathbf{F}(k)$ is given by

$$\mathbf{F}(k) = \alpha^2(\mathbf{I} - \mathbf{J})\mathbf{d}(k)\mathbf{d}(k)'(\mathbf{I} - \mathbf{J})' - \alpha\mathbf{W}(k)\mathbf{z}(k)\mathbf{d}(k)'(\mathbf{I} - \mathbf{J})' - \alpha(\mathbf{I} - \mathbf{J})\mathbf{d}(k)\mathbf{z}(k)'\mathbf{W}(k)'.$$

$$\tag{2.28}$$

24

The solution of (2.27) is given by

$$\mathbf{Z}(k) = \Phi(k,0)\mathbf{Z}(0)\Phi(k,0)' + \sum_{s=0}^{k-1} \Phi(k,s+1)\mathbf{F}(s)\Phi(k,s+1)'. \tag{2.29}$$

For simplicity, in what follows, we will omit the matrix $\mathbf{I} - \mathbf{J}$ from $\mathbf{F}(k)$ since it disappears by multiplication with the transition matrix (see Proposition 2.3.1). We can further write

$$\|\mathbf{Z}(k)\| \leq \|\Phi(k,0)\|^2 \|\mathbf{Z}(0)\| + \sum_{s=0}^{k-1} \|\Phi(k,s+1)\mathbf{F}(s)\Phi(k,s+1)'\|,$$

and by noting that $\|\mathbf{Z}(k)\| = \|\mathbf{z}(k)\|^2$, we obtain

$$E[\|\mathbf{z}(k)\|^2] \leq E[\|\Phi(k,0)\|^2]\|\mathbf{z}(0)\|^2 + \sum_{s=0}^{k-1} E[\|\Phi(k,s+1)\mathbf{F}(s)\Phi(k,s+1)'\|]. \tag{2.30}$$

From (2.19) of Proposition 2.3.2 we obtain

$$E[\|\Phi(k,0)\|^2] \leq \rho^{\left\lfloor \frac{k}{m} \right\rfloor}.$$

We now focus on the terms of the sum in the right hand-side of (2.30). We have

$$\Phi(k,s+1)\mathbf{F}(s)\Phi(k,s+1)' = \alpha^2 \Phi(k,s+1)\mathbf{d}(s)\mathbf{d}(s)'\Phi(k,s+1)' -$$

$$-\alpha\Phi(k,s+1)\mathbf{W}(s)\mathbf{z}(s)\mathbf{d}(s)'\Phi(k,s+1)' - \alpha\Phi(k,s+1)\mathbf{d}(s)\mathbf{z}(s)'\mathbf{W}(s)'\Phi(k,s+1)'.$$

Using the solution of $\mathbf{z}(k)$ given in (2.11), we get

$$\Phi(k,s+1)\mathbf{W}(s)\mathbf{z}(s)\mathbf{d}(s)'\Phi(k,s+1)' =$$

$$= \Phi(k,s+1)\mathbf{W}(s)\left(\Phi(s,0)\mathbf{z}(0) - \alpha\sum_{r=0}^{s-1}\Phi(s,r+1)\mathbf{d}(r)\right)\mathbf{d}(s)'\Phi(k,s+1)'$$

$$= \Phi(k,0)\mathbf{z}(0)\mathbf{d}(s)'\Phi(k,s+1)' - \alpha\sum_{r=0}^{s-1}\Phi(k,r+1)\mathbf{d}(r)\mathbf{d}(s)'\Phi(k,s+1)'. \tag{2.31}$$

Similarly,

$$\Phi(k,s+1)\mathbf{d}(s)\mathbf{z}(s)'\mathbf{W}(s)'\Phi(k,s+1)' =$$

25

$$\Phi(k,s+1)\mathbf{d}(s)\mathbf{z}(0)'\Phi(k,0)' - \alpha\sum_{r=0}^{s-1}\Phi(k,s+1)\mathbf{d}(s)\mathbf{d}(r)'\Phi(k,r+1)'. \tag{2.32}$$

We now give a more explicit formula for $\Phi(k,s+1)\mathbf{F}(s)\Phi(k,s+1)'$:

$$\Phi(k,s+1)\mathbf{F}(s)\Phi(k,s+1)' = \alpha^2\Phi(k,s+1)\mathbf{d}(s)\mathbf{d}(s)'\Phi(k,s+1)'-$$

$$-\alpha\Phi(k,0)\mathbf{z}(0)\mathbf{d}(s)'\Phi(k,s+1)+\alpha^2\sum_{r=0}^{s-1}\Phi(k,r+1)\mathbf{d}(r)\mathbf{d}(s)'\Phi(k,s+1)'-$$

$$-\alpha\Phi(k,s+1)\mathbf{d}(s)\mathbf{z}(0)'\Phi(k,0)'+\alpha^2\sum_{r=0}^{s-1}\Phi(k,s+1)\mathbf{d}(s)\mathbf{d}(r)'\Phi(k,r+1)'.$$

By applying the norm operator, we get

$$\|\Phi(k,s+1)\mathbf{F}(s)\Phi(k,s+1)'\| \le N\alpha^2\varphi^2\|\Phi(k,s+1)\|^2+$$

$$+N\alpha^2\varphi^2\sum_{r=0}^{s-1}\|\Phi(k,r+1)\Phi(k,s+1)'\|+N\alpha^2\varphi^2\sum_{r=0}^{s-1}\|\Phi(k,s+1)\Phi(k,r+1)'\|+$$

$$+N\alpha\beta\varphi\|\Phi(k,s+1)\Phi(k,0)'\|+N\alpha\beta\varphi\|\Phi(k,0)\Phi(k,s+1)'\|,$$

or

$$\|\Phi(k,s+1)\mathbf{F}(s)\Phi(k,s+1)'\| \le N\alpha^2\varphi^2\|\Phi(k,s+1)\|^2+$$

$$+2N\alpha^2\varphi^2\sum_{r=0}^{s-1}\|\Phi(k,r+1)\Phi(k,s+1)'\|+2N\alpha\beta\varphi\|\Phi(k,s+1)\Phi(k,0)'\|. \tag{2.33}$$

Next we derive bounds for the expected values of each of the terms in (2.33). Based on the results of Proposition 2.3.2 we can write

$$E[\|\Phi(k,s+1)\|^2] \le \rho^{\lfloor\frac{k-s-1}{m}\rfloor},$$

$$\sum_{r=0}^{s-1}E[\|\Phi(k,r+1)\Phi(k,s+1)'\|] \le \sum_{r=0}^{s-1}\rho^{\lfloor\frac{k-s-1}{m}\rfloor}\eta^{\lfloor\frac{s-r}{m}\rfloor} \le m\rho^{\lfloor\frac{k-s-1}{m}\rfloor}\sum_{r=0}^{\lfloor\frac{s}{m}\rfloor}\eta^r \le$$

$$\le m\rho^{\lfloor\frac{k-s-1}{m}\rfloor}\frac{1-\eta^{\lfloor\frac{s}{m}\rfloor+1}}{1-\eta} \le m\rho^{\lfloor\frac{k-s-1}{m}\rfloor}\frac{1}{1-\eta}$$

26

and

$$E[\|\Phi(k, s+1)\Phi(k, 0)'\|] \leq \rho^{\left\lfloor \frac{k-s-1}{m} \right\rfloor} \eta^{\left\lfloor \frac{s+1}{m} \right\rfloor}.$$

Therefore we obtain

$$E[\|\Phi(k, s+1)\mathbf{F}(s)\Phi(k, s+1)'\|] \leq N\alpha^2\varphi^2\left(1 + \frac{2m}{1-\eta}\right)\rho^{\left\lfloor \frac{k-s-1}{m} \right\rfloor} + 2N\alpha\beta\varphi\rho^{\left\lfloor \frac{k-s-1}{m} \right\rfloor}\eta^{\left\lfloor \frac{s+1}{m} \right\rfloor}.$$

We know compute an upper bound for $\sum_{s=0}^{k-1} E[\|\Phi(k, s+1)\mathbf{F}(s)\Phi(k, s+1)'\|]$. Using the

fact that

$$\sum_{s=0}^{k-1} \rho^{\left\lfloor \frac{k-s-1}{m} \right\rfloor} \leq m \sum_{s=0}^{\left\lfloor \frac{k-1}{m} \right\rfloor} \rho^s \leq m\frac{1 - \rho^{\left\lfloor \frac{k-1}{m} \right\rfloor + 1}}{1 - \rho} \leq m\frac{1}{1 - \rho}$$

and

$$\sum_{s=0}^{k-1} \rho^{\left\lfloor \frac{k-s-1}{m} \right\rfloor}\eta^{\left\lfloor \frac{s+1}{m} \right\rfloor} \leq \sum_{s=0}^{k-1} \rho^{\left\lfloor \frac{k-s-1}{m} \right\rfloor}\eta^{\left\lfloor \frac{s}{m} \right\rfloor} \leq$$

$$\leq m \sum_{s=0}^{\left\lfloor \frac{k-1}{m} \right\rfloor} \rho^{\left\lfloor \frac{k-1}{m} \right\rfloor - s}\eta^s = m\frac{\rho^{\left\lfloor \frac{k-1}{m} \right\rfloor + 1} - \eta^{\left\lfloor \frac{k-1}{m} \right\rfloor + 1}}{\rho - \eta},$$

we obtain

$$\sum_{s=0}^{k-1} E[\|\Phi(k, s+1)\mathbf{F}(s)\Phi(k, s+1)'\|] \leq N\alpha^2\varphi^2\left(1 + \frac{2m}{1-\eta}\right)\frac{m}{1-\rho} +$$

$$+ 2N\alpha\beta\varphi m\frac{\rho^{\left\lfloor \frac{k-1}{m} \right\rfloor + 1} - \eta^{\left\lfloor \frac{k-1}{m} \right\rfloor + 1}}{\rho - \eta}.$$

Finally we obtain an upper bound for the second moment of $\|\mathbf{z}(k)\|$:

$$E[\|\mathbf{z}(k)\|^2] \leq N\beta^2\rho^{\left\lfloor \frac{k}{m} \right\rfloor} + N\alpha^2\varphi^2\left(1 + \frac{2m}{1-\eta}\right)\frac{m}{1-\rho} + 2N\alpha\beta\varphi m\frac{\rho^{\left\lfloor \frac{k-1}{m} \right\rfloor + 1} - \eta^{\left\lfloor \frac{k-1}{m} \right\rfloor + 1}}{\rho - \eta}.$$

□

The following lemma allows us to interpret $d_i(k)$ as an ϵ-subgradient of f_i at $\bar{x}(k)$

(with ϵ being a random variable).

Lemma 2.3.3. *Let Assumptions 2.2.2 or 2.2.3 hold. Then the vector $d_i(k)$ is an $\epsilon(k)$-subdifferential of f_i at $\bar{x}(k)$, i.e. $d_i(k) \in \partial_{\epsilon(k)} f_i(\bar{x}(k))$ and $h(k) = \sum_{i=1}^{N} d_i(k)$ is an $N\epsilon(k)$-subdifferential of f at $\bar{x}(k)$, i.e. $h(k) \in \partial_{N\epsilon(k)} f(\bar{x}(k))$, for any $k \geq 0$, where*

$$\epsilon(k) = 2\varphi\beta \sqrt{N}\|\Phi(k,0)\| + 2\alpha\varphi^2 \sqrt{N} \sum_{s=0}^{k-1} \|\Phi(k,s+1)\|. \tag{2.34}$$

Proof. The proof is somewhat similar to the proof of Lemma 3.4.5 of [19]. Let $\bar{d}_i(k)$ be a subgradient of f_i at $\bar{x}(k)$. By the subgradient definition we have that

$$f_i(x_i(k)) \geq f_i(\bar{x}(k)) + \bar{d}_i(k)'(x_i(k) - \bar{x}(k)) \geq f_i(\bar{x}(k)) - \|\bar{d}_i(k)\|\|(x_i(k) - \bar{x}(k))\|,$$

or

$$f_i(x_i(k)) \geq f_i(\bar{x}(k)) - \varphi\|z_i(k)\|.$$

Furthermore, for any $y \in \mathbb{R}^n$ we have that

$$f_i(y) \geq f_i(x_i(k)) + d_i(k)'(y - x_i(k)) = f_i(x_i(k)) + d_i(k)'(y - \bar{x}(k)) + d_i(k)'(\bar{x}(k) - x_i(k)) \geq$$

$$\geq f_i(\bar{x}(k)) + d_i(k)'(y - \bar{x}(k)) - 2\varphi\|z_i(k)\| \geq f_i(\bar{x}(k)) + d_i(k)'(y - \bar{x}(k)) - 2\varphi\|\mathbf{z}(k)\|,$$

or

$$f_i(y) \geq f_i(\bar{x}(k)) + d_i(k)'(y - \bar{x}(k)) - \epsilon(k),$$

where $\epsilon(k) = 2\varphi\|\mathbf{z}(k)\|$. Using the definition of the ϵ-subgradient, it follows that $d_i(k) \in \partial_{\epsilon(k)} f_i(\bar{x}(k))$. Summing over all i we get that $\sum_{i=1}^{N} d_i(k) \in \partial_{N\epsilon(k)} f(\bar{x}(k))$. Note, that $\epsilon(k)$ has a random characteristic due to the assumptions on $\mathbf{A}(k)$. $\square$

For twice differentiable cost functions with lower and upper bounded hessians, the next result gives an upper bound on the second moment of the distance between the average vector $\bar{x}(k)$ and the minimizer of f.

28

Lemma 2.3.4. *Let Assumptions 2.2.1 and 2.2.3 hold and let $\{\bar{x}(k)\}_{k \geq 0}$ be a sequence of vectors defined by iteration (2.7). Then, the following inequality holds*

$$E[\|\bar{x}(k) - x^*\|^2] \leq \|\bar{x}(0) - x^*\|^2 \gamma^k + \frac{4\alpha\varphi\beta\sqrt{N}}{\gamma\eta} \frac{\gamma^k - \eta^{\frac{k}{m}}}{\gamma - \eta^{\frac{1}{m}}} + \frac{\alpha^2\varphi^2}{1-\gamma}\left(4\sqrt{N}\frac{m}{1-\eta} + 1\right), \quad (2.35)$$

where $\gamma = 1 - \alpha l$, with $l = \min_i l_i$ and η is defined in Remark 2.3.2.

Proof. Under Assumption 2.2.3, $f(x)$ is a strongly convex function with constant Nl, where $l = \min_i l_i$ and therefore it follows that

$$f(x) - f^* \geq \frac{Nl}{2}\|x - x^*\|^2. \quad (2.36)$$

We use the same idea as in the proof of Proposition 2.4 in [30], formulated under a deterministic setup. By (2.7), where we use a constant stepsize α, we obtain

$$\|\bar{x}(k+1) - x^*\|^2 = \|\bar{x}(k) - x^* - \frac{\alpha}{N}h(k)\|^2 = \|\bar{x}(k) - x^*\|^2 - 2\frac{\alpha}{N}h(k)'(\bar{x}(k) - x^*) + \alpha^2\varphi^2.$$

Using the fact that, by Lemma 2.3.3, $h(k)$ is a $N\epsilon(k)$-subdifferential of f at $\bar{x}(k)$, we have

$$f(x^*) \geq f(\bar{x}(k)) + h(k)'(x^* - \bar{x}(k)) - N\epsilon(k),$$

or, from inequality (2.36),

$$-h(k)'(\bar{x}(k) - x^*) \leq -\frac{Nl}{2}\|\bar{x}(k) - x^*\|^2 + N\epsilon(k).$$

Further, we can write

$$\|\bar{x}(k+1) - x^*\|^2 \leq (1 - \alpha l)\|\bar{x}(k) - x^*\|^2 + 2\alpha\epsilon(k) + \alpha^2\varphi^2$$

or

$$E[\|\bar{x}(k) - x^*\|^2] \leq (1 - \alpha l)^k \|\bar{x}(0) - x^*\|^2 + \sum_{s=0}^{k-1}(1 - \alpha l)^{k-s-1}(2\alpha E[\epsilon(s)] + \alpha^2\varphi^2).$$

29

Note that from Assumption 2.2.3-(c), $0 < \alpha < \frac{1}{l}$ and therefore the quantity $\gamma^k = (1 - \alpha l)^k$ does not grow unbounded. It follows that

$$E[\|\bar{x}(k) - x^*\|^2] \le \gamma^k \|\bar{x}(0) - x^*\|^2 + \sum_{s=0}^{k-1} \gamma^{k-s-1}(2\alpha E[\epsilon(s)] + \alpha^2 \varphi^2). \qquad (2.37)$$

From the expression of $\epsilon(k)$ in Lemma 2.3.3, we immediately obtain the following inequality

$$E[\epsilon(s)] \le 2\varphi\beta \sqrt{N} \eta^{\lfloor \frac{s}{m} \rfloor} + \frac{2\alpha\varphi^2 \sqrt{N}m}{1 - \eta}. \qquad (2.38)$$

The inequality

$$\sum_{s=0}^{k-1} \gamma^{k-1-s} \eta^{\lfloor \frac{s}{m} \rfloor} \le \gamma^{k-1} \eta^{-1} \sum_{s=0}^{k-1} \left(\frac{\eta^{\frac{1}{m}}}{\gamma} \right)^s = (\gamma\eta)^{-1} \frac{\gamma^k - \eta^{\frac{k}{m}}}{\gamma - \eta^{\frac{1}{m}}}$$

yields

$$\sum_{s=0}^{k-1} \gamma^{k-s-1} E[\epsilon(s)] \le \frac{2\varphi\beta \sqrt{N}}{\gamma\eta} \frac{\gamma^k - \eta^{\frac{k}{m}}}{\gamma - \eta^{\frac{1}{m}}} + \frac{2\alpha\varphi^2 \sqrt{N}m}{1 - \eta} \frac{1}{1 - \gamma}, \qquad (2.39)$$

which combined with (2.37), generates the inequality (2.35). $\qquad \square$

2.4 Main Results - Error bounds

In the following we provide upper bounds for two performance metrics of the CB-MASM. First, we give a bound on the difference between the best recorded value of the cost function f, evaluated at the estimate $x_i(k)$, and the optimal value f^*. Second, we focus on the second moment of the distance between the estimate $x_i(k)$ and the minimizer of f^*. For a particular class of twice differentiable functions, we give an upper bound on this metric and show how fast the time varying part of this bound converge to zero. The bounds we give in these section emphasize the effect of the random topology on the performance metrics.

The following result shows how close the cost function f evaluated at the estimate $x_i(k)$ gets to the optimal value f^*. A similar result for the standard sub-gradient method can be found in [31], for example.

Corollary 2.4.1. *Let Assumptions 2.2.1 and 2.2.2 or 2.2.1 and 2.2.3 hold and let $\{x_i(k)\}_{k \geq 0}$ be a sequence generated by the iteration (2.5), $i = 1, \ldots N$. Let $\bar{f}_i^{best}(k) = \min_{s=0\ldots k} E[f(x_i(s))]$ be the smallest cost value (in average) achieved by agent i at iteration k. Then*

$$\lim_{k \to \infty} \bar{f}_i^{best}(k) \leq f^* + 3\alpha\varphi^2 N \sqrt{N} \frac{m}{1-\eta} + \frac{N\alpha\varphi^2}{2}. \tag{2.40}$$

Proof. Using the subgradient definition of f_i at $x_i(k)$ we have that

$$f_i(x_i(k)) \leq f_i(\bar{x}(k)) + \varphi\|z_i(k)\|, \text{ for all } i = 1, \ldots, N.$$

Summing over all i, we get

$$f(x_i(k)) \leq f(\bar{x}(k)) + N\varphi\|\mathbf{z}(k)\|,$$

which holds with probability one. Subtracting f^* from both sides of the above inequality, and applying the expectation operator, we further get

$$E[f(x_i(k))] - f^* \leq E[f(\bar{x}(k))] - f^* + N\varphi E[\|\mathbf{z}(k)\|],$$

or

$$\bar{f}_i^{best}(k) - f^* \leq \min_{s=0\ldots k} \{E[f(\bar{x}(s))] - f^* + N\varphi E[\|\mathbf{z}(s)\|]\}. \tag{2.41}$$

Let $x^* \in X^*$ be an optimal point of f. By (2.7), where we use a constant stepsize α, we obtain

$$\|\bar{x}(k+1) - x^*\|^2 = \|\bar{x}(k) - x^* - \frac{\alpha}{N}h(k)\|^2 \leq \|\bar{x}(k) - x^*\|^2 - 2\frac{\alpha}{N}h(k)'(\bar{x}(k) - x^*) + \alpha^2\varphi^2$$

31

and since, by Lemma 2.3.3, $h(\bar{x}(k))$ is a $N\epsilon(k)$-subdifferential of f at $\bar{x}(k)$, we have

$$\|\bar{x}(k+1) - x^*\|^2 \le \|\bar{x}(k) - x^*\|^2 - \frac{2\alpha}{N}(f(\bar{x}(k)) - f^*) + 2\alpha\epsilon(k) + \alpha^2\varphi^2,$$

or

$$\|\bar{x}(k) - x^*\|^2 \le \|\bar{x}(0) - x^*\|^2 - \frac{2\alpha}{N}\sum_{s=0}^{k-1}(f(\bar{x}(s)) - f^*) + 2\alpha\sum_{s=0}^{k-1}\epsilon(s) + k\alpha^2\varphi^2.$$

Since $\|\bar{x}(k) - x^*\|^2 \ge 0$

$$\frac{2\alpha}{N}\sum_{s=0}^{k-1}(f(\bar{x}(s)) - f^*) \le \|\bar{x}(0) - x^*\|^2 + 2\alpha\sum_{s=0}^{k-1}\epsilon(s) + k\alpha^2\varphi^2,$$

or

$$\sum_{s=0}^{k-1}(E[f(\bar{x}(s))] - f^*) \le \frac{N}{2\alpha}\|\bar{x}(0) - x^*\|^2 + N\sum_{s=0}^{k-1}E[\epsilon(s)] + \frac{kN\alpha\varphi^2}{2}.$$

Adding and subtracting $N\varphi E[\|\mathbf{z}(s)\|]$ inside the sum of the left-hand side of the above inequality and recalling from Lemma 2.3.3 that $\epsilon(k) = 2\varphi\|\mathbf{z}(k)\|$, we obtain

$$\sum_{s=0}^{k-1}(E[f(\bar{x}(s))] - f^* + N\varphi E[\|\mathbf{z}(s)\|]) \le \frac{1}{2\alpha}\|\bar{x}(0) - x^*\|^2 + \frac{3N}{2}\sum_{s=0}^{k-1}E[\epsilon(s)] + \frac{kN\alpha\varphi^2}{2}.$$

Using the fact that

$$\sum_{s=0}^{k-1}(E[f(\bar{x}(s))] - f^* + N\varphi E[\|\mathbf{z}(s)\|]) \ge k\min_{s=0,\dots,k-1}\{E[f(\bar{x}(s))] - f^* + N\varphi E[\|\mathbf{z}(s)\|]\},$$

we get

$$\min_{s=0,\dots,k-1}\{E[f(\bar{x}(s))] - f^* + N\varphi E[\|\mathbf{z}(s)\|]\} \le \frac{1}{2\alpha k}\|\bar{x}(0) - x^*\|^2 + \frac{3N}{2k}\sum_{s=0}^{k-1}E[\epsilon(s)] + \frac{N\alpha\varphi^2}{2}.$$

Using inequality (2.38) from Lemma 2.3.3 we obtain

$$\sum_{s=0}^{k-1}E[\epsilon(s)] \le 2\varphi\beta\sqrt{N}\frac{m}{1-\eta} + k2\alpha\varphi^2\sqrt{N}\frac{m}{1-\eta}.$$

It follows that

$$\min_{s=0,\dots,k-1} \{E[f(\bar{x}(s))] - f^* + N\varphi E[\|\mathbf{z}(s)\|]\} \le \frac{1}{2\alpha k}\|\bar{x}(0) - x^*\|^2 +$$

$$+ \frac{3N}{2k}\left(2\varphi\beta\sqrt{N}\frac{m}{1-\eta} + k2\alpha\varphi^2\sqrt{N}\frac{m}{1-\eta}\right) + \frac{N\alpha\varphi^2}{2}. \tag{2.42}$$

Combining inequalities (2.41) and (2.42) and taking the limit, we obtain

$$\lim_{k\to\infty} \bar{f}_i^{best}(k) \le f^* + 3\alpha\varphi^2 N\sqrt{N}\frac{m}{1-\eta} + \frac{N\alpha\varphi^2}{2}.$$

$\square$

In the case of twice differentiable functions, the next result introduces an error bound which essentially says that the estimates "converge in the mean square sense to within some guaranteed distance" from the optimal point, distance which can be made arbitrarily small by an appropriate choice of the stepsize. In addition, the time varying component of the error bound converges to zero at least linearly.

Corollary 2.4.2. *Let Assumptions 2.2.1 and 2.2.3 hold. Then, for the sequence $\{x_i(k)\}_{k\ge0}$ generated by iteration (2.5) we have*

(a)

$$\limsup_{k\to\infty} E[\|x_i(k) - x^*\|^2] \le C_1 + C_2 + 2\sqrt{C_1 C_2}, \tag{2.43}$$

where

$$C_1 = \frac{\alpha^2\varphi^2}{1-\gamma}\left(\frac{4m\sqrt{N}}{1-\eta} + 1\right), \quad C_2 = N\alpha^2\varphi^2\left(1 + \frac{2m}{1-\eta}\right)\frac{m}{1-\rho}. \tag{2.44}$$

(b)

$$E[\|x_i(k) - x^*\|^2] \le \psi(k) + C, \tag{2.45}$$

33

where $\psi(k) = c\delta^k$ with c a positive constant depending on the initial conditions, $\delta = $

$\max\{\gamma, \eta^{\frac{1}{m}}\}$, $\gamma = 1 - \alpha l$, and where $C = 4\max\{C_1, C_2\}$.

Proof. By the triangle inequality we have

$$\|x_i(k) - x^*\|^2 \le \|x_i(k) - \bar{x}(k)\|^2 + 2\|x_i(k) - \bar{x}(k)\|\|\bar{x}(k) - x^*\| + \|\bar{x}(k) - x^*\|^2.$$

or

$$E[\|x_i(k) - x^*\|^2] \le E[\|x_i(k) - \bar{x}(k)\|^2] + 2E[\|x_i(k) - \bar{x}(k)\|\|\bar{x}(k) - x^*\|] + E[\|\bar{x}(k) - x^*\|^2].$$

By the Cauchy-Schwarz inequality for the expectation operator, we get

$$E[\|x_i(k) - x^*\|^2] \le E[\|x_i(k) - \bar{x}(k)\|^2] + 2E[\|x_i(k) - \bar{x}(k)\|^2]^{\frac{1}{2}} E[\|\bar{x}(k) - x^*\|^2]^{\frac{1}{2}} + E[\|\bar{x}(k) - x^*\|^2].$$

$$(2.46)$$

Inequality (2.35) can be further upper bounded by

$$E[\|\bar{x}(k) - x^*\|^2] \le \psi_1(k) + C_1,$$

where

$$\psi_1(k) = \underbrace{\left[\|\bar{x}(0) - x^*\|^2 + \frac{8\alpha\varphi\beta\sqrt{N}}{\gamma\eta} \frac{1}{\gamma - \eta^{\frac{1}{m}}}\right]}_{c_1} \delta^k = c_1 \delta^k,$$

with $\delta = \max\{\gamma, \eta^{\frac{1}{m}}\}$ and C_1 being given in (2.44). Using the inequalities

$$\rho^{\left\lfloor\frac{k-1}{m}\right\rfloor + 1} \le \rho^{-\frac{1}{m}}\rho^{\frac{k}{m}} \text{ and } \eta^{\left\lfloor\frac{k-1}{m}\right\rfloor + 1} \le \eta^{-\frac{1}{m}}\eta^{\frac{k}{m}},$$

from (2.26), a new bound for $E[\|x_i(k) - \bar{x}(k)\|^2]$ is given by

$$E[\|x_i(k) - \bar{x}(k)\|^2] \le \psi_2(k) + C_2,$$

34

where C_2 is given in (2.44) and

$$\psi_2(k) = \underbrace{\left[N\beta^2\rho^{-1} + \frac{2N\alpha\beta\varphi m}{\eta - \rho}\left(\eta^{-\frac{1}{m}} + \rho^{-\frac{1}{m}}\right)\right]}_{c_2}\delta^k = c_2\delta^k.$$

Taking the limit of (2.46) and recalling that under Assumptions 2.2.1 and 2.2.3, $\gamma < 1$ and $\eta^{\frac{1}{m}} < 1$ for any $m \in I_m$, we obtain (2.43).

Inequality (2.46) can be further upper bounded by

$$E[\|x_i(k) - x^*\|^2] \leq 2\max\{c_1, c_2\}\delta^k + 2\left(\max\{c_1, c_2\}\delta^k + \max\{C_1, C_2\}\right) = \psi(k) + C,$$

where $\psi(k) = c\delta^k$, with $c = 4\max\{c_1, c_2\}$ and $C = 4\max\{C_1, C_2\}$. Hence, we obtained that the time varying component of the error bound converges linearly to zero with a factor $\delta = \max\{\gamma, \eta^{\frac{1}{m}}\}$.

$\square$

2.4.1 Discussion of the results

We obtained upper bounds on two performance metrics relevant to the CBMASM. First we studied the difference between the cost function evaluated at the estimate and the optimal solution (Corollary 2.4.1) - for non-differentiable and differentiable functions with bounded (sub)gradients. Second, for a particular class of convex functions (see Assumptions 2.2.3), we gave an upper bound for the second moment of the distance between the estimates of the agents and the minimizer. We also showed that the time varying component of this upper bound converges linearly to zero with a factor reflecting the contribution of the random topology. We introduced Assumption 2.2.3 to cover part of the class of convex functions for which uniform boundness of the (sub)gradients can

not be guaranteed.

From our results we can notice that the stepsize has a similar influence as in the case of the standard subgradient method, i.e. a small value of α implies good precision but slow rate of convergence, while a larger value of α increases the rate of convergence but at a cost in accuracy. More importantly, we can emphasize the influence of the consensus step on the performance of the distributed algorithm. When possible, by appropriately designing the probability distribution of the random graph (together with an appropriate choice of the integer m) we can improve the guaranteed precision of the algorithm (intuitively, this means making the quantities $m/(1 - \eta)$ and $m/(1 - \rho)$ as small as possible). In addition, the rate of convergence of the time varying component of the error bound (2.45) can be improved by making the quantity $\eta^{\frac{1}{m}}$ as small as possible. Note however that there are limits with respect to the positive effect of the consensus step on the the rate of convergence of $\psi(k)$, since the latter is determined by the maximum between γ and $\eta^{\frac{1}{m}}$. Indeed, if the stepsize is small enough, i.e.

$$\alpha < \frac{1}{l}(1 - \eta^{\frac{1}{m}}), \tag{2.47}$$

then the rate of convergence of $\psi(k)$ is given by γ. This suggests that having a fast consensus step will not necessarily be helpful in the case of a small stepsize, which is in accordance with the intuition on the role of a small value of α. In the case inequality (2.47) is not satisfied, the rate of convergence of $\psi(k)$ is determined by $\eta^{\frac{1}{m}}$. However, this does not necessarily means that the estimates will not "converge faster to within some distance of the minimizer", since we are providing only an error bound.

Assume that we are using the centralized subgradient method to minimize the con-

vex function $f(x) = \sum_{i=1}^{N} f_i(x)$ satisfying Assumption 2.2.2 (the subgradients of $f_i(x)$ are uniformly bounded by φ), where the stepsize used is N times smaller than the stepsize of the distributed algorithm, i.e.

$$x(k+1) = x(k) - \frac{\alpha}{N} d(k),$$

where $d(k)$ is a subgradient of f at $x(k)$, with $\|d(k)\| \leq N\varphi$. Then, from the optimization literature we get

$$\lim_{k \to \infty} f^{best}(k) \leq f^* + \frac{N\alpha\varphi^2}{2},$$

where $f^{best}(k) = \min_{s=0,\dots,k} f(x(s))$. From above we note that, compared with the centralized subgradient method with a step size N times smaller than the agents' stepsize, the distributed optimization algorithm introduced an additional term in the error bound given by $3\alpha\varphi^2 N \sqrt{N} \frac{m}{1-\eta}$, which reflects the influence of the dimension of the network and of the random topology on the guaranteed accuracy of the algorithm.

Let us now assume that we are minimizing the function $f(x)$, satisfying Assumptions 2.2.3-(a)(b), using a centralized gradient algorithm:

$$x(k+1) = x(k) - \frac{\alpha}{N} \nabla f(x(k)),$$

where we have that α is small enough ($0 < \alpha < \frac{2}{L}$) so that the algorithm is stable and there exit φ_c so that $\|\nabla f_i(x(k))\| \leq \varphi_c$. It follows that we can get the following upper bound on the distance between the estimate of the optimal decision vector and the minimizer

$$\|x(k) - x^*\|^2 \leq \|x(0) - x^*\|^2 \gamma_c^k + \frac{\alpha\varphi_c^2}{l},$$

with $\gamma_c = 1 - \alpha l$. Therefore, we can see that $\gamma = \gamma_c$ which shows that the rates of convergence, at which the time-varying components of the error bounds converge to zero in the

centralized and distributed cases, are the same. However, please note that we assumed the stepsise in the centralized case to be N times smaller than the stepsize used by the agents.

The error bounds (2.40) and (2.45) are functions of three quantities induced by the consensus step: $\frac{m}{1-\eta}$, $\frac{m}{1-\rho}$ and $\eta^{\frac{1}{m}}$. These quantities show the dependence of the performance metrics on the pmf of $G(k)$ and on the corresponding random matrix $A(k)$. The scalars η and ρ represent the first and second moments of the SLEM of the random matrix $A(k+1)\ldots A(k+m)$, corresponding to a random graph formed over a time interval of length m, respectively. We notice from our results that the performance of the CBMASM is improved by making $\frac{m}{1-\eta}$, $\frac{m}{1-\rho}$ and $\eta^{\frac{1}{m}}$ as small as possible, i.e. by optimizing these quantities having as decision variables m and the pmf of $G(k)$. For instance if we are interested in obtaining a tight bound on $E[\|x_i(k) - x^*\|^2]$ and having a fast decrease to zero of $\psi(k)$, we can formulate the following multi-criteria optimization problem:

$$\min_{m,p_i} \quad \{\eta^{\frac{1}{m}}, C_1 + C_2 + 2\sqrt{C_1 C_2}\}$$

$$\text{subject to:} \quad m \geq 1,$$

$$\eta^{\frac{1}{m}} \geq \gamma, \tag{2.48}$$

$$\sum_i^M p_i = 1, \; p_i \geq 0.$$

where C_1 and C_2 were defined in (2.44). The second inequality constraint was added to emphasize the fact that making $\eta^{\frac{1}{m}}$ too small is pointless, since that rate of convergence of $\psi(k)$ is limited by γ. If we are simultaneously interested in tightening the upper bounds of both metrics, we can introduce the quantity $\frac{m}{1-\eta}$ in the optimization problem since $\frac{m}{1-\eta}$ and $\frac{m}{1-\rho}$ are not necessarily minimized by the same probability distribution. The solution to the above problem is a set of Pareto points, i.e. solution points for which improvement in one objective can only occur with the worsening of at least one other objective.

We note that for each fixed value of m, the three quantities are minimized if the scalars η and ρ are minimized as functions of the pmf of the random graph. An approximate solution of (2.48) can be obtained by focusing only on minimizing $\frac{m}{1-\eta}$, since both $\eta^{\frac{1}{m}}$ and $\frac{m}{1-\rho}$ are upper bounded by this quantity. Therefore, an approximate solution can be obtained by minimizing η (i.e. computing the optimal pmf) for each value of m, and then picking the best value m with the corresponding η that minimizes $\frac{m}{1-\eta}$. Depending on the communication model used, the pmf of the random graph can be a quantity dependent on a set of parameters of the communication protocol (transmission power, probability of collisions, etc) and therefore we can potentially tune these parameters so that the performance of the CBMASM is improved.

In what follows we provide a simple example where we show how η, the optimal probability distribution, $\frac{m}{1-\eta}$ and $\eta^{\frac{1}{m}}$ evolve as functions of m.

Example 2.4.1. *Let $G(k)$ be a random graph process taking values in the set $\mathcal{G} = \{G_1, G_2\}$, with probability p and $1-p$, respectively. The graphs G_1 and G_2 are shown in Figure 2.1. Also, let $A(k)$ be a (stochastic) random matrix, corresponding to $G(k)$, taking value in the set $\mathcal{A} = \{A_1, A_2\}$, with*

$$
A_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, A_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}
$$

Figure 2.2(a) shows the optimal probability p^* that minimizes η for different values of m. Figure 2.2(b) shows the optimized η (computed at p^*) as a function of m. Figures 2.2(c) and 2.2(d) show the evolution of the optimized $\frac{m}{1-\eta}$ and $\eta^{\frac{1}{m}}$ as functions of m, from

39
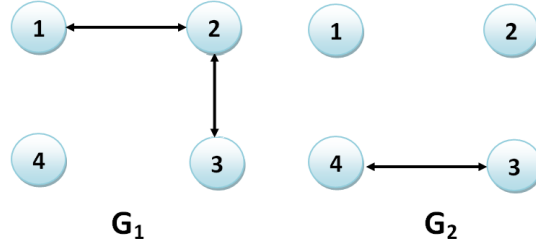
Figure 2.1: The sample space of the random graph $\mathbf{G}(k)$

where we notice that a Pareto solution is obtained for $m = 5$ and $p^* = 0.582$.

In order to obtain the solution of problem (2.48), we need to compute the probability of all possible sequences of length m produced by $G(k)$, together with the SLEM of their corresponding stochastic matrices. This task, for large values of m and M may prove to be numerically expensive. We can somewhat simplify the computational burden by using the bounds on η and ρ introduced in (2.16) and (2.17), respectively. Note that every result concerning the performance metrics still holds. In this case, for each value of m, the upper bound on η is minimized, when $\mathbf{p}_m$ is maximized, which can be interpreted as having to choose a pmf that maximizes the probability of connectivity of the union of random graph obtained over a time interval of length m.

Even in the case where we use the bound on η, it may be very difficult to compute the expression for $\mathbf{p}_m$, for large values of m (the set $\mathcal{G}$ may allow for a large number of possible unions of graphs that produce connected graphs). Another way to simplify our problem even more, is to (intelligently) fix a value for m and try to maximize $\mathbf{p}_m$ having as decision variable the pmf. We note that m should be chosen such that, within a time interval of length m, a connected graph can be obtained. Also, a very large value

40
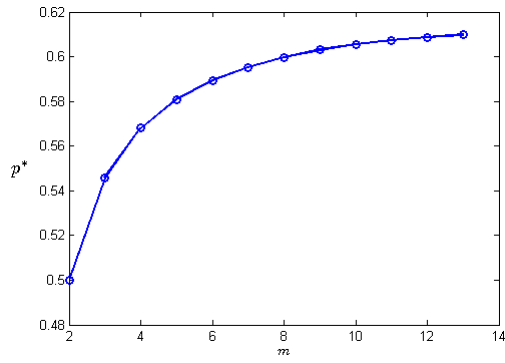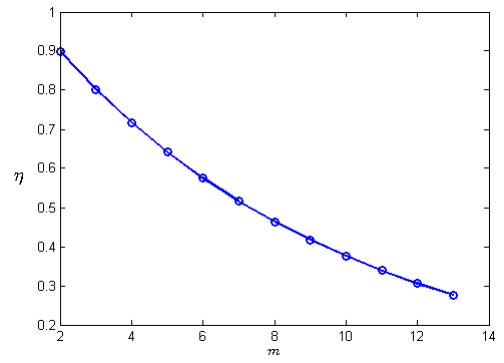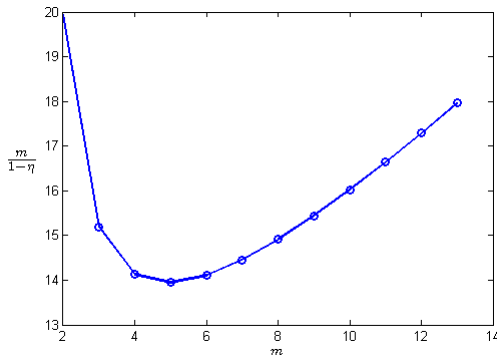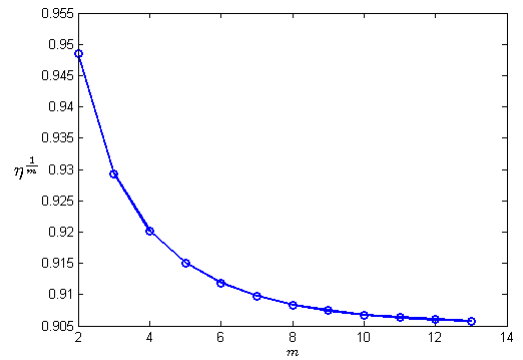
Figure 2.2: (a) Optimal p as a function of m; (b) Optimized η as a function of m; (c) Optimized $\frac{m}{1-\eta}$ as a function of m; (d) Optimized $\eta^{\frac{1}{m}}$ as a function of m.

41

for m should be avoided, since $\frac{m}{1-\eta}$ is lower bounded by m. Although in general the uniform distribution does not necessarily minimize η, it becomes the optimizer under some particular assumptions, stated in what follows. Let $\mathcal{G}$ be such that a connected graph can be obtained only over a time interval of length M (i.e. in order to form a connected graph, all graphs in $\mathcal{G}$ must appear within a sequence of length M). Choose M as the value for m. It follows that $\mathbf{p}_m$ can be expressed as:

$$\mathbf{p}_m = m! \prod_{i=1}^{M} p_i.$$

We can immediately observe that $\mathbf{p}_m$ is maximized for the uniform distribution, i.e. $p_i = \frac{1}{m}$, for $i = 1, \ldots, M$.

2.5 Application - Distributed System Identification

In this section we show how the distributed optimization algorithm analyzed in the previous section can be used to perform collaborative system identification. We assume the following scenario: a group of sensors track an object by taking measurements of its position. These sensors have memory and computation capabilities and are organized in a communication network modeled by a random graph process $G(k)$ satisfying the assumptions introduced in Section II. The task of the sensors/agents is to determine a parametric model of the object's trajectory. The measurements are affected by noise, whose effect may differ from sensor to sensor (i.e. some sensors take more accurate measurements than others). This can happen for instance when some sensors are closer to the object than other (allowing a better reading of the position), or sensors with different precision classes are used. Determining a model for the time evolution of the object's

position can be useful in motion prediction when the motion dynamics of the object in unknown to the sensors. The notations used in the following are independent from the ones used in the previous sections.

2.5.1 System identification model

Let $p(t)' = [x(t), y(t), z(t)]$ be the position vector of the tracked object. We model the time evolution of each of the axis of the position vector as a time dependent polynomial of degree n_a, i.e.

$$x(t) = a_0^x + a_1^x t + \ldots + a_{n_a}^x t^{n_a},$$
$$y(t) = a_0^y + a_1^y t + \ldots + a_{n_a}^y t^{n_a}, \tag{2.49}$$
$$z(t) = a_0^z + a_1^z t + \ldots + a_{n_a}^z t^{n_a}.$$

The measurements of each sensor i are given by

$$x_i(t) = x(t) + e_{i,x}(t),$$
$$y_i(t) = y(t) + e_{i,y}(t), \tag{2.50}$$
$$z_i(t) = z(t) + e_{i,z}(t),$$

where $e_{i,x}(t)$, $e_{i,y}(t)$ and $e_{i,y}(t)$ are assumed white noises of (unknown) variances $\sigma_{i,x}^2$, $\sigma_{i,y}^2$ and $\sigma_{i,z}^2$ respectively. Equivalently, we have

$$x_i(t) = \varphi(t)' \theta_x + e_{i,x}(t),$$
$$y_i(t) = \varphi(t)' \theta_y + e_{i,y}(t), \tag{2.51}$$
$$z_i(t) = \varphi(t)' \theta_z + e_{i,z}(t),$$

where $\varphi(t)' = [1, t, \ldots, t^{n_a}]$ and $\theta_x = [a_{0,x}, \ldots, a_{n_a,x}]'$, $\theta_y = [a_{0,y}, \ldots, a_{n_a,y}]'$ and $\theta_z = [a_{0,z}, \ldots, a_{n_a,z}]'$.

In the following we focus only on one coordinate of the position vector, say $x(t)$. The analysis, however can be mimicked in a similar way for the other two coordinates. Let

T be the total number of measurements taken by the sensors and consider the following quadratic cost functions

$$\mathcal{J}_i(\theta_x) = \sum_{t=1}^{T} (x_i(t) - \varphi(t)'\theta_x)^2, \ \forall i.$$

Using its own measurements, sensor i can determine a parametric model for the time evolution of the coordinate $x(t)$ by solving the optimization problem:

$$\min_{\theta_x} \mathcal{J}_i(\theta_x). \tag{2.52}$$

Let $\mathbf{X}_i' = [x_i(1), \ldots, x_i(T)]$ be the vector of measurements of sensor i and let $\Phi' = [\varphi(1), \ldots, \varphi(T)]$ be the matrix formed by the regression vectors. It is well known that the optimal solution of (2.52) is given by

$$\hat{\theta}_{i,x} = (\Phi'\Phi)^{-1}\Phi'\mathbf{X}_i. \tag{2.53}$$

Remark 2.5.1. *It can be shown that $\Phi'\Phi$ is invertible for any T, but it becomes ill conditioned for large values of T. That is why, for our numerical simulations, we will in fact use an orthogonal basis to model the time evolution of the coordinates $x(t)$, $y(t)$, and $z(t)$.*

Performing a localized system identification does not take into account the measurements of the other sensors, which can potentially enhance the identified model. If all the measurements are centralized, a model for the time evolution of $x(t)$ can be computed by solving

$$\min_{\theta_x} \mathcal{J}(\theta_x),$$

where

$$\mathcal{J}(\theta_x) = \sum_{i=1}^{N} \mathcal{J}_i(\theta_x). \tag{2.54}$$

44

Note that (2.54) fits the framework of the distributed optimization problem formulated in the previous sections, and therefore can be solved distributively, eliminating the need for sharing all measurements with all other sensors.

Remark 2.5.2. *If each sensor has a priori information about its accuracy, then the cost function (2.54) can be replaced with*

$$\mathcal{J}(\theta_x) = \sum_{i=1}^{N} \delta_{i,x} \mathcal{J}_i(\theta_x), \tag{2.55}$$

where $\delta_{i,x}$ is a positive scalar such that the more accurate sensor i is, the larger δ_i is. The scalar $\delta_{i,x}$ can be interpreted as trust *in the measurements taken by sensor i. The sensors can use local identification to compute $\delta_{i,x}$. For instance, $\delta_{i,x}$ can be chosen as $\delta_{i,x} = \frac{1}{\hat{\sigma}_{i,x}^2}$, where $\hat{\sigma}_{i,x}^2$ is given by*

$$\hat{\sigma}_{i,x}^2 = \frac{1}{T} \sum_{t=1}^{T} (x_i(t) - \varphi(t)' \hat{\theta}_{i,x})^2,$$

where $\hat{\theta}_{i,x}$ is the local estimate of the model for the time evolution of $x(t)$.

The distributed optimization algorithm (2.5) can be written as

$$\theta_{i,x}(k+1) = \sum_{j=1}^{N} a_{ij}(k) \theta_{j,x}(k) - \alpha \nabla \mathcal{J}_i(k), \tag{2.56}$$

where $\nabla \mathcal{J}_i(k) = -2\Phi'(\mathbf{X}_i - \Phi \theta_{i,x}(k))$.

2.5.2 Numerical simulations

In this section we simulate the distributed system identification algorithm under two gossip communication protocols: the randomized gossip protocol [7] and the broadcast gossip protocol [1]. We perform the simulations on a circular graph, where we assume

that the cardinality of the neighborhoods of the nodes is two. This graph is a particular

example of small world graphs [53] (for an analysis of the consensus problem under small

world like communication topologies, the reader can consult [3] for example).
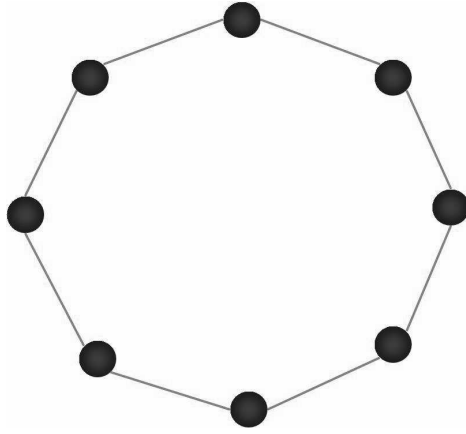
Figure 2.3: Circular graph with $N = 8$

In the case of the randomized gossip protocol, the set of consensus matrices is given

by

$$\mathcal{A}^r = \{A_{ij}, i = 1 \ldots N, j \in \{i-1, i+1\}\},$$

where $A_{ij} = I - \frac{1}{2}(e_i - e_j)(e_i - e_j)'$ and where by convention we assume that if $i = N$ then

$i + 1 = 1$ and if $i = 1$ then $i - 1 = N$. We assume that if node i wakes up, it chooses with

uniform distribution between its two neighbors. Hence the probability distribution of the

random matrix $A(k)$ is given by

$$Pr(A(k) = A_{ij}) = \frac{1}{2N}.$$

We note that the minimum value of m such that $\eta_m < 1$ is $N - 1$. Recall that m is the length

46

of a time interval such that $Pr\left(\bigcup_{l=0}^{m-1} G(k+l)\right) > 0$ for any k. It turns out that for $m = N-1$

$$p_c^r = Pr\left(\bigcup_{l=0}^{N-2} G(k+l)\right) = N!\left(\frac{1}{2N}\right)^{N-1}$$

Interestingly, the matrix products of length $N-1$ of the form $\prod_{i=1}^{N-1} A_{i+i_0, i+1+i_0}$ with $i_0 \in$

$\{0, \ldots, N-1\}$, and the matrix products that may be obtained by the permutations of the

matrices in the aforementioned matrix products, have the same SLEM (where the sum-

mations in the indices are seen as modulo N). In fact it is exactly this property that allows

us to give the following explicit expression for η_{N-1}

$$\eta_{N-1}^r = p_c^r \lambda^r + 1 - p_c^r, \tag{2.57}$$

where λ^r is the SLEM of the matrix product $A_{1,2} A_{2,3} \cdots A_{N-1,N}$.

In the case of the broadcast gossip protocol, the set $\mathcal{A}$ is given by

$$\mathcal{A}^b = \{A_i, i = 1 \ldots N\},$$

where $A_i = I - \frac{1}{3}\left[(e_i - e_{i+1})(e_i - e_{i+1})' + (e_i - e_{i-1})(e_i - e_{i-1})'\right]$ and $Pr(A(k) = A_i) = \frac{1}{N}$. For

odd values of N (and $N \geq 3$), the minimum value of m such that $\eta_m < 1$ is given by

$m = \frac{N-1}{2}$. In addition, we have that

$$p_c^b = Pr\left(\bigcup_{l=0}^{\frac{N-3}{2}} G(k+l)\right) = N\left(\frac{N-1}{2}\right)!\left(\frac{1}{N}\right)^{\frac{N-1}{2}}.$$

Observing a similar phenomenon as in the case of the randomized gossip protocol, namely

that the matrix products $A_{1+i_0} A_{3+i_0} \ldots A_{N-2+i_0}$ for $i_0 \in \{0, \ldots N-1\}$ and their permutations

have the same SLEM (where as before the summations of indices are seen as modulo N),

we obtain the formula

$$\eta_{\frac{N-1}{2}}^b = p_c^b \lambda^b + 1 - p_c^b,$$

47

where λ^b is the SLEM of the matrix product $A_1 A_3 \cdots A_{N-2}$.

The values for η^r_{N-1} and $\eta^b_{\frac{N-1}{2}}$ computed above, in the case of the two gossip protocols, do not necessarily provide tight error bounds, since we considered minimal time interval lengths so that $\eta_m < 1$. Even for this relatively simple type of graph, analytical formulas for η_m, for large values of m, are more difficult to obtain due to an increase in combinatorial complexity and because different matrix products that appear in the expression of η do not necessarily have the same SLEM. However, we did compute numerical estimates for different values of m. Figures 4 and 5 show estimates of the three quantities of interest, η, $\frac{m}{1-\eta}$ and $\eta^{\frac{1}{m}}$, as functions of m, for $N = 11$ (the estimates were computed by taking averages over 2000 realizations and are shown together with the 95% confidence intervals). We can see that $\frac{m}{1-\eta}$ is minimized for $m \approx 55$ in the case of the randomized gossip protocol and for $m \approx 30$ in the case of the broadcast gossip protocol, while the best achievable $\eta^{\frac{1}{m}}$ are approximately equal for the two protocols, (i.e. 0.985. for the randomized gossip protocols and 0.982 for the broadcast gossip protocols).

Next we present numerical simulations of the distributed system identification algorithm presented in the previous subsection, under the randomized and broadcast gossip protocols. We would like to point out that, in order to maintain numerical stability, in our numerical simulation we used an orthogonalized version of Φ, given by $\tilde{\Phi} = \Phi H$, where $\tilde{\Phi}$'s columns form an orthogonal basis of the range of Φ, and the new vector of the parameters is given $\tilde{\theta} = H\theta$, where H is a linear transformation matrix, whose entries depend on the orthogonalization process used (Gram-Schmidt, Householder transformations, etc.).

(a)

(b)

(c)

49
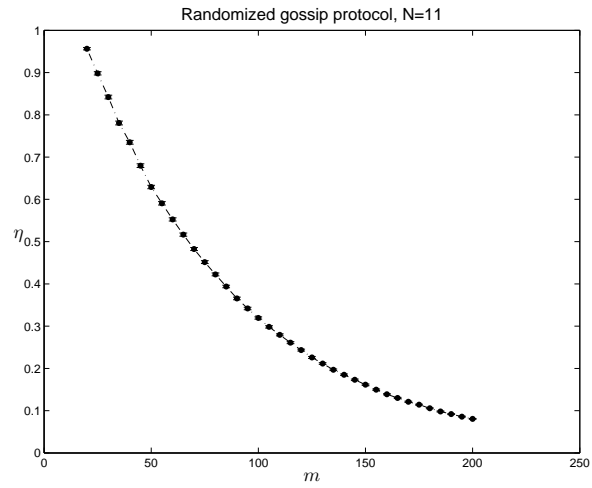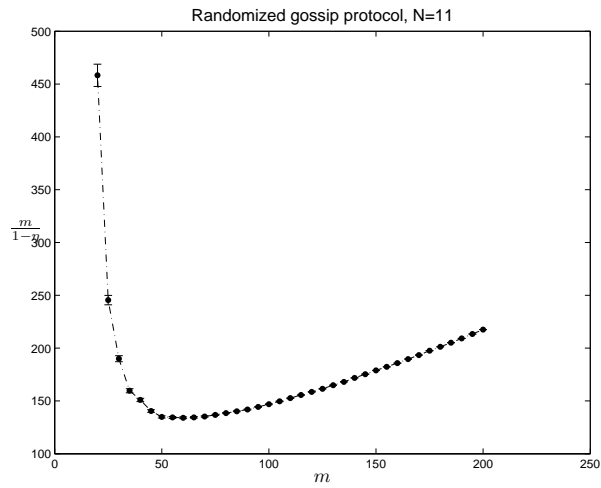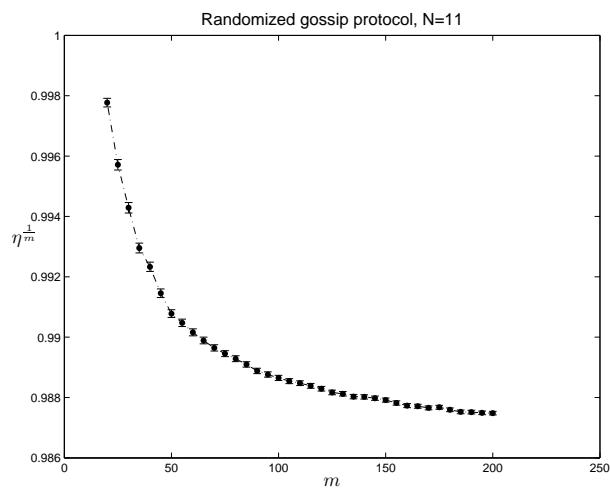
Figure 2.4: Estimates of η, $\frac{m}{1-\eta}$ and $\eta^{\frac{1}{m}}$ for the randomized gossip protocol and for $N = 11$

(a)

(b)

(c)

50
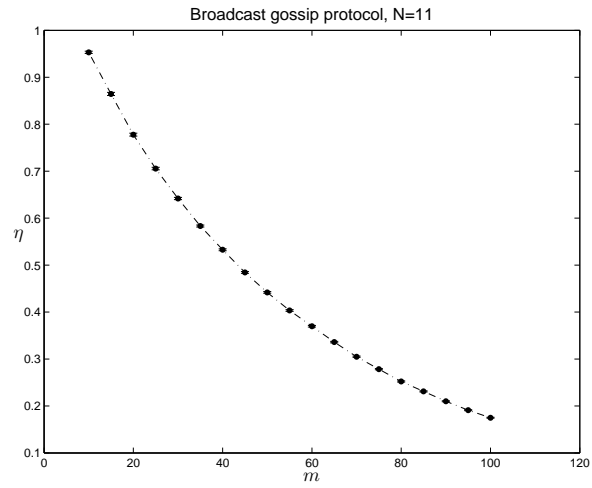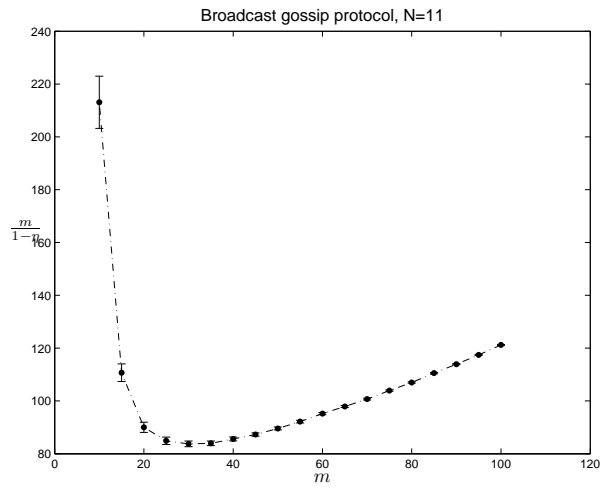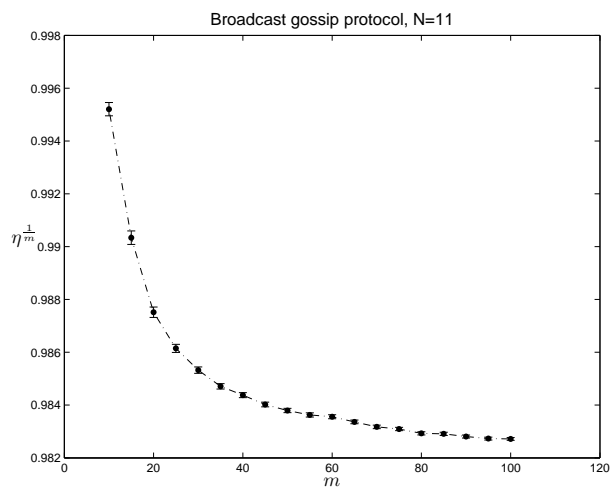
Figure 2.5: Estimates of η, $\frac{m}{1-\eta}$ and $\eta^{\frac{1}{m}}$ for the broadcast gossip protocol and for $N = 11$

Therefore, the cost function we are minimizing can be rewritten as

$$\mathcal{J}(\tilde{\theta}_x) = \sum_{i=1}^{n} \mathcal{J}_i(\tilde{\theta}_x),$$

where $\mathcal{J}_i(\tilde{\theta}_x) = \left\| \mathbf{X}_i - \tilde{\Phi}\tilde{\theta}_x \right\|^2$.

It is easy to check that in the case of the two protocols, $\underline{\lambda}$ (the smallest of all eigenvalues of matrices belonging to the set $\mathcal{A}$) is zero. In addition, Assumption 2.2.3-(a)(b) are satisfied for $l_i = L_i = 2$, and for $\alpha < \frac{1}{2}$ the distributed optimization algorithm is guaranteed to be stable with probability one (recall Lemma 2.3.1). From above we see that $\eta^{\frac{1}{m}}$ can not attain less than 0.98 for both protocols, for any m. Therefore, although we can choose $\alpha > 0.01$ which in turn implies $\gamma < 0.98$, our analysis cannot guarantee a rate of convergence for $\psi(k)$ smaller than 0.98, since the rate of convergence is upper bounded by the maximum between γ and $\eta^{\frac{1}{m}}$. However, this does not mean that faster rates of convergence can not be achieved, which in fact is shown in our numerical simulations.

Figures 6 and 7 present numerical simulations of the distributed system identification algorithm for the two protocols and for a circular graph with $N = 11$. In our numerical experiments we considered a number $T = 786$ of measurements of the x-coordinate of the trajectory depicted in Figure 2.6. We assumed that the x-coordinate measurements are affected by white, Gaussian noise with a signal-to-noise ration given by $SNR_i = 5 \times i \, dB$, for $i = 1 \ldots 11$. The time polynomials modeling the trajectory evolution are chosen of degree ten, i.e. $n_a = 10$. We plot estimates of two metrics: $\max_i E[\|\tilde{\theta}_{i,x}(k) - \tilde{\theta}^*\|]$ and $\max_i E[f(\tilde{\theta}_{i,x}(k))] - f^*$ for different values of α (the estimates were computed by taking averages over 500 realizations). We note that for larger values of α, under the two protocols, the algorithm has roughly the same rate of convergence, but the broadcast protocol

51
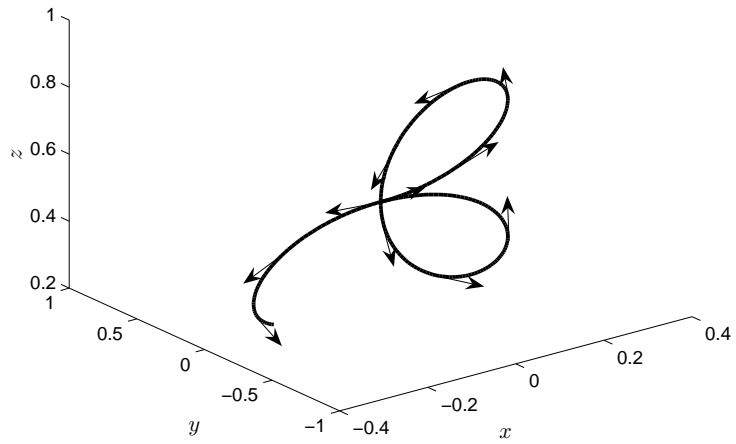
Figure 2.6: Trajectory of the object

is more accurate. This is in accordance with our analysis, since as Figures 4 and 5 show,

$\frac{m}{1-\eta^b} \leq \frac{m}{1-\eta^r}$ for any m, quantities which control the guaranteed accuracy. For smaller

values of α, under both protocols the algorithm becomes more accurate and the rate of

convergence decreases since the parameters γ becomes larger and therefore dominant.

(a)

(b)

(c)
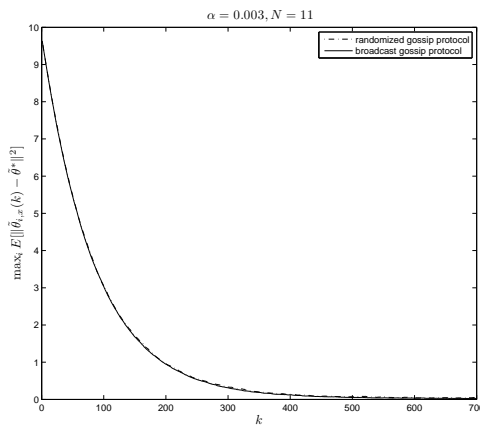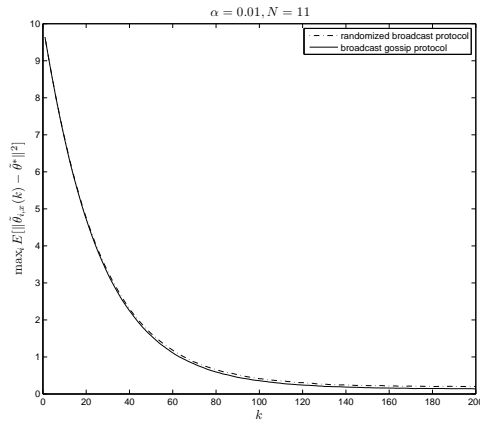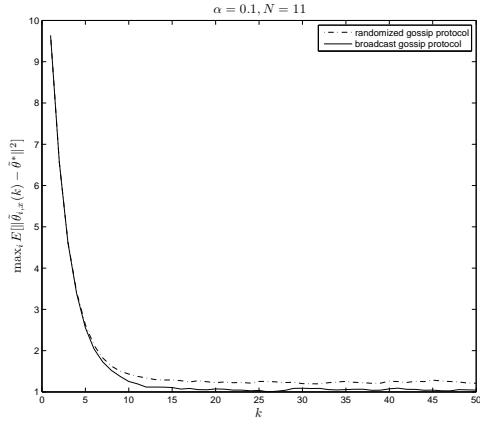
Figure 2.7: Estimate of $\max_i E[\|\tilde{\theta}_{i,x}(k) - \tilde{\theta}^*\|]$ for the randomized and broadcast protocol gossip protocols

53
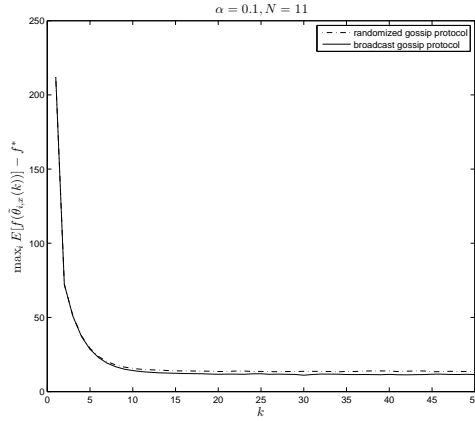
(a)
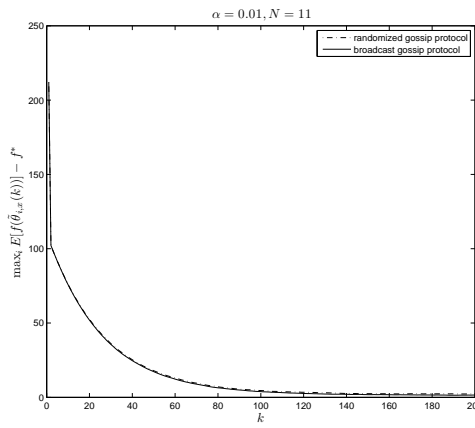
(b)
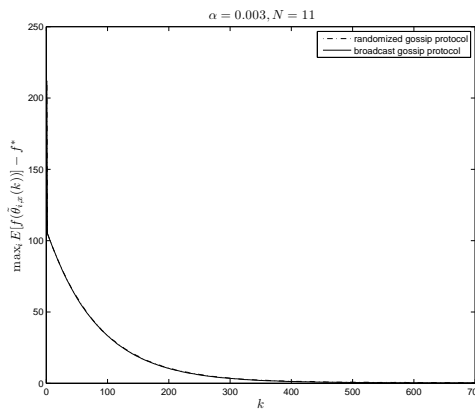
(c)

Figure 2.8: Estimate of $\max_i E[f(\tilde{\theta}_{i,x}(k))] - f^*$ for the randomized and broadcast protocol

gossip protocols   54

Chapter 3

Distributed Asymptotic Agreement Problem on Convex Metric Spaces

3.1 Introduction

A convex metric space is a metric space endowed with a convex structure. In this chapter we generalize the asymptotic consensus problem to the more general case of convex metric spaces and emphasize the fundamental role of convexity and in particular of the convex hull of a finite set of points. Tsitsiklis showed in [51] that, under some minimal connectivity assumptions on the communication network, if an agent updates its value by choosing a point (in $\mathbb{R}^n$) from the (interior) of the convex hull of its current value and the current values of its neighbors, then asymptotic convergence to consensus is achieved. We will show that this idea extends naturally to the more general case of convex metric spaces.

Our main contributions are as follows. *First*, after citing relevant results concerning convex metric spaces, we study the properties of the distance between two points belonging to two, possibly overlapping convex hulls of two finite sets of points. These properties will prove to be crucial in proving the convergence of the agreement algorithm. *Second*, we provide a dynamic equation for an upper bound of the vector of distances between the current values of the agents. We show that the agents asymptotically reach agreement, by showing that this upper bound asymptotically converges to zero. *Third*, we characterize the agreement point(s) compared to the initial values of the agents, be giving upper

bounds on the distance between the agreement point(s) and the initial values in terms of the distances between the initial values of the agents. *Forth*, we emphasize the relevance of our framework, by providing an application under the form of a consensus of opinion algorithm. For this example we define a particular convex metric space and we study in more depth the properties of the convex hull of a finite set of points.

The chapter is organized as follows. Section 3.2 introduces the main concepts related to the convex metric spaces and focuses in particular on the convex hull of a finite set. Section 3.3 formulates the problem and states our main theorem. Section 3.4 gives the proof of our main theorem together with some auxiliary results. In Section 3.6 we present an application of our main result by providing an iterative algorithm for reaching consensus of opinion.

Some basic notations: Given $W \in \mathbb{R}^{n \times n}$ by $[W]_{ij}$ we refer to the (i, j) element of the matrix. The *underlying graph* of W is a graph of order n for which every edge corresponds to a non-zero, non-diagonal entry of W. We will denote by $\mathbb{1}_{\{A\}}$ the indicator function of event A. Given some space $\mathcal{X}$ we denote by $\mathcal{P}(\mathcal{X})$ the set of all subsets of $\mathcal{X}$.

3.2 Convex Metric Spaces

The first part of this section deals with a set of definitions and basic results about convex metric spaces. The second part focuses on the convex hull of a finite set in convex metric spaces.

3.2.1 Definitions and Results on Convex Metric Spaces

For more details about the following definitions and results the reader is invited to consult [46],[49].

Definition 3.2.1. *Let (X, d) be a metric space. A mapping $\psi : X \times X \times [0, 1] \to X$ is said to be a* convex structure *on X if*

$$d(u, \psi(x, y, \lambda)) \le \lambda d(u, x) + (1 - \lambda) d(u, y), \ \forall x, y, u \in X \text{ and } \forall \lambda \in [0, 1]. \tag{3.1}$$

Definition 3.2.2. *The metric space (X, d) together with the convex structure ψ is called a convex metric space.*

A Banach space and each of its subsets are convex metric spaces. There are examples of convex metric spaces not embedded in any Banach space. The following two examples are taken from [49].

Example 3.2.1. *Let I be the unit interval $[0, 1]$ and X be the family of closed intervals $[a_i, b_i]$ such that $0 \le a_i \le b_i \le 1$. For $I_i = [a_i, b_i]$, $I_j = [a_j, b_j]$ and $\lambda \in I$, we define a mapping ψ by $\psi(I_i, I_j, \lambda) = [\lambda a_i + (1 - \lambda) a_j, \lambda b_i + (1 - \lambda) b_j]$ and define a metric d in X by the Hausdorff distance, i.e.*

$$d(I_i, I_j) = \max\{|a_i - a_j|, |b_i - b_j|\}.$$

Example 3.2.2. *We consider a linear space L which is also a metric space with the following properties:*

(a) For $x, y \in L$, $d(x, y) = d(x - y, 0)$;

(b) For $x, y \in L$ and $\lambda \in [0,1]$,

$$d(\lambda x + (1-\lambda)y, 0) \leq \lambda d(x, 0) + (1-\lambda)d(y, 0).$$

Hence L, together with the convex structure $\psi(x, y, \lambda) = \lambda x + (1-\lambda)y$, is a convex metric space.

Definition 3.2.3. *Let X be a convex metric space. A nonempty subset $K \subset X$ is said to be convex if $\psi(x, y, \lambda) \in K$, $\forall x, y \in K$ and $\forall \lambda \in [0, 1]$.*

We define the set valued mapping $\tilde{\psi} : \mathcal{P}(X) \to \mathcal{P}(X)$ as

$$\tilde{\psi}(A) \triangleq \{\psi(x, y, \lambda) \mid \forall x, y \in A, \forall \lambda \in [0, 1]\}, \tag{3.2}$$

where A is an arbitrary set in X.

In [49] it is shown that, in a convex metric space, an arbitrary intersection of convex sets is also convex and therefore the next definition makes sense.

Definition 3.2.4. *The* convex hull *of the set $A \subset X$ is the intersection of all convex sets in X containing A and is denoted by $co(A)$.*

Another characterization of the convex hull of a set in X is given in what follows. By defining $A_m \triangleq \tilde{\psi}(A_{m-1})$ with $A_0 = A$ for some $A \subset X$, it is discussed in [46] that the set sequence $\{A_m\}_{m \geq 0}$ is increasing and $\limsup A_m$ exists, and $\limsup A_m = \liminf A_m = \lim A_m = \bigcup_{m=0}^{\infty} A_m$.

Proposition 3.2.1 ([46]). *Let X be a convex metric space. The convex hull of a set $A \subset X$ is given by*

$$co(A) = \lim A_m = \bigcup_{m=0}^{\infty} A_m. \tag{3.3}$$

It follows immediately from above that if $A_{m+1} = A_m$ for some m, then $co(A) = A_m$.

58

3.2.2 On the convex hull of a finite set

For a positive integer n, let $A = \{x_1, \ldots, x_n\}$ be a finite set in $\mathcal{X}$ with convex hull $co(A)$ and let z belong to $co(A)$. By Proposition 3.2.1 it follows that there exists a positive integer m such that $z \in A_m$. But since $A_m = \tilde{\psi}(A_{m-1})$ it follows that there exits $z_1, z_2 \in A_{m-1}$ and $\lambda_{(1,2)} \in [0,1]$ such that $z = \psi(z_1, z_2, \lambda_{(1,2)})$. Similarly, there exits $z_3, z_4, z_5, z_6 \in A_{m-2}$ and $\lambda_{(3,4)}, \lambda_{5,6} \in [0,1]$ such that $z_1 = \psi(z_3, z_4, \lambda_{(3,4)})$ and $z_2 = \psi(z_5, z_6, \lambda_{(5,6)})$. By further decomposing z_3, z_4, z_5 and z_6 and their followers until they are expressed as functions of elements of A and using a graph theory terminology, we note that z can be viewed as the root of a weighted binary tree with leaves belonging to the set A. Each node α (except the leaves) has two children α_1 and α_2, and are related through the operator ψ in the sense $\alpha = \psi(\alpha_1, \alpha_2, \lambda)$ for some $\lambda \in [0,1]$. The weights of the edges connecting α with α_1 and α_2 are given by λ and $1 - \lambda$ respectively.

From the above discussion we note that for any point $z \in co(A)$ there exits a non-negative integer m such that z is the root of a binary tree of height m, and has as leaves elements of A. The binary tree rooted at z may or may not be a *perfect binary tree*, i.e. a full binary tree in which all leaves are at the same depth. That is because on some branches of the tree the points in A are reached faster then on others. Let n_i denote the number of times x_i appears as a leaf node, with $\sum_{i=1}^{n} n_i \leq 2^m$ and let m_{i_l} be the length of the i_l^{th} path from the root z to the node x_i, for $l = 1 \ldots n_i$. We formally describe the paths from the root z to x_i as the set

$$P_{z,x_i} \triangleq \left\{ \left(\{y_{i_l,j}\}_{j=0}^{m_{i_l}}, \{\lambda_{i_l,j}\}_{j=1}^{m_{i_l}} \right) \mid l = 1 \ldots n_i \right\}, \tag{3.4}$$

where $\{y_{i_l,j}\}_{j=0}^{m_{i_l}}$ is the set of points forming the i_l^{th} path, with $y_{i_l,0} = z$ and $y_{i_l,m_{i_l}} = x_i$ and

59

where $\{\lambda_{i_l,j}\}_{j=1}^{m_{i_l}}$ is the set of weights corresponding to the edges along the paths, in partic-

ular $\lambda_{i_l,j}$ being the weight of the edge $(y_{i_l,j-1}, y_{i_l,j})$. We define the aggregate weight of the

paths from root z to node x_i as

$$\mathcal{W}(P_{z,x_i}) \triangleq \sum_{l=1}^{n_i} \prod_{j=1}^{m_{i_l}} \lambda_{i_l,j}. \tag{3.5}$$

It is not difficult to note that all the aggregate weights of the paths from the root z

to the leaves $\{x_1, \ldots, x_n\}$ sum up to one, i.e.

$$\sum_{i=1}^{n} \mathcal{W}(P_{z,x_i}) = 1.$$
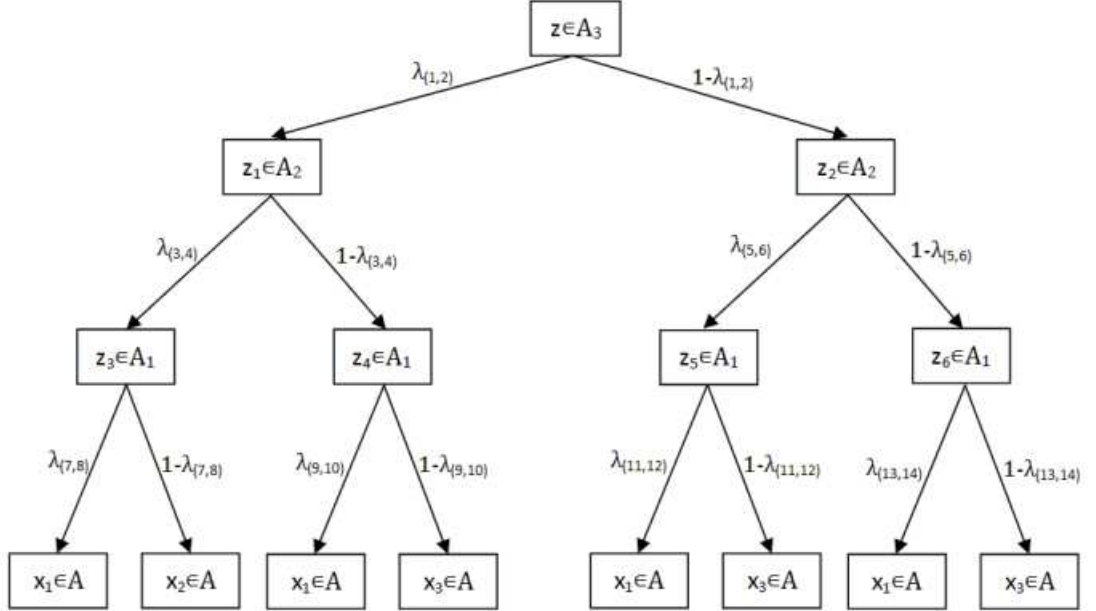
Figure 3.1: The decomposition of a point $z \in A_3$ with $A = \{x_1, x_2, x_3\}$

Example 3.2.3. *Figure 3.1 shows a binary tree corresponding to a point $z \in A_3$, where*
$A = \{x_1, x_2, x_3\}$. For this particular example, the paths from to root z to the leaves x_i are
given by

$$P_{z,x_1} = \{(\{z, z_1, z_3, x_1\}, \{\lambda_{(1,2)}, \lambda_{(3,4)}, \lambda_{(7,8)}\}), (\{z, z_1, z_4, x_1\}, \{\lambda_{(1,2)}, (1 - \lambda_{(3,4)}), \lambda_{(9,10)}\}),$$

60

$$(\{z, z_2, z_5, x_1\}, \{(1-\lambda_{(1,2)}), \lambda_{(5,6)}, \lambda_{(11,12)}\}), (\{z, z_2, z_6, x_1\}, \{(1-\lambda_{(1,2)}), (1-\lambda_{(5,6)}), \lambda_{(13,14)}\})\},$$

$$P_{z,x_2} = \{(\{z, z_1, z_3, x_2\}, \{\lambda_{(1,2)}, \lambda_{(3,4)}, (1-\lambda_{(7,8)})\})\}$$

$$P_{z,x_3} = \{(\{z, z_1, z_4, x_3\}, \{\lambda_{(1,2)}, (1-\lambda_{(3,4)}), (1-\lambda_{(9,10)})\}), (\{z, z_2, z_5, x_3\}, \{(1-\lambda_{(1,2)}), \lambda_{(5,6)}, (1-\lambda_{(11,12)})\}),$$

$$(\{z, z_2, z_6, x_3\}, \{(1-\lambda_{(1,2)}), (1-\lambda_{(5,6)}), (1-\lambda_{(13,14)})\})\}$$

 and the path weights are

$$\mathcal{W}(P_{z,x_1}) = \lambda_{(1,2)}\lambda_{(3,4)}\lambda_{(7,8)} + \lambda_{(1,2)}(1-\lambda_{(3,4)})\lambda_{(9,10)} + (1-\lambda_{(1,2)}), \lambda_{(5,6)}, \lambda_{(11,12)},$$

$$\mathcal{W}(P_{z,x_2}) = \lambda_{(1,2)}\lambda_{(3,4)}(1-\lambda_{(7,8)}),$$

$$\mathcal{W}(P_{z,x_3}) = \lambda_{(1,2)}(1-\lambda_{(3,4)})(1-\lambda_{(9,10)}) + (1-\lambda_{(1,2)})\lambda_{(5,6)}(1-\lambda_{(11,12)}) + (1-\lambda_{(1,2)})(1-\lambda_{(5,6)})(1-\lambda_{(13,14)}).$$

Definition 3.2.5. *Given a small enough positive scalar $\varepsilon < 1$ we define the following subset of $co(A)$ consisting of all points in $co(A)$ whose aggregate weights are lower bounded by ε, i.e.*

$$co_\varepsilon(A) \triangleq \{z \mid z \in co(A), \mathcal{W}(P_{z,x_i}) \geq \varepsilon, \ \forall x_i \in A\}. \tag{3.6}$$

Remark 3.2.1. *By a* small enough *value of ε we understand a value such that the inequality $\mathcal{W}(P_{z,x_i}) \geq \varepsilon$ is satisfied for all i. Obviously, for n agents ε needs to satisfy*

$$\varepsilon \leq \frac{1}{n},$$

but usually we would want to choose a value much smaller then $1/n$ since this implies a richer set $co_\varepsilon(A)$.

Remark 3.2.2. *We can iteratively generate points for which we can make sure that they belong to the interior of the convex hull of a finite set $A = \{x_1, \ldots, x_n\}$. Given a set of positive scalars $\{\lambda_1, \ldots, \lambda_{n-1}\} \in (0,1)$, consider the iteration*

$$y_{i+1} = \psi(y_i, x_{i+1}, \lambda_i) \ for \ i = 1 \ldots n-1 \ with \ y_1 = x_1. \tag{3.7}$$

It is not difficult to note that y_n is guaranteed to belong to the interior of co(A). In addition,

if we impose the condition

$$\varepsilon^{\frac{1}{n-1}} \le \lambda_i \le \frac{1-(n-1)\varepsilon}{1-(n-2)\varepsilon}, \ i = 1\ldots n-1, \tag{3.8}$$

and ε respects the inequality

$$\varepsilon^{\frac{1}{n-1}} \le \frac{1-(n-1)\varepsilon}{1-(n-2)\varepsilon}, \tag{3.9}$$

then $y_n \in co_\varepsilon(A)$. We should note that for any $n \ge 2$ we can find a small enough value of ε

such that inequality (3.9) is satisfied.

The next result characterizes the distance between two points $x, y \in X$ belonging to

the convex hulls of two (possibly overlapping) finite sets X and Y.

Proposition 3.2.2. *Let $X = \{x_1, \ldots, x_{n_x}\}$ and $Y = \{y_1, \ldots, y_{n_y}\}$ be two finite sets on X and*

let $\varepsilon < 1$ be a positive scalar small enough.

(a) If $x \in co(X)$ and $y \in X$ then

$$d(x,y) \le \sum_{i=1}^{n_x} \lambda_i d(x_i, y), \tag{3.10}$$

for some $\lambda_i \ge 0$ with $\sum_{i=1}^{n_x} \lambda_i = 1$.

(b) If $x \in co(X)$ and $y \in co(Y)$ then

$$d(x,y) \le \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \lambda_{ij} d(x_i, y_j), \tag{3.11}$$

for some $\lambda_{ij} \ge 0$ with $\sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \lambda_{ij} = 1$.

(c) If $x \in co_\varepsilon(X)$, $y \in co_\varepsilon(Y)$, then

$$\lambda_i \ge \varepsilon \text{ and } \lambda_{ij} \ge \varepsilon^2, \ \forall \ i, j, \tag{3.12}$$

where λ_i and λ_{ij} where introduced in part (a) and part (b), respectively.

(d) If $x \in co_\varepsilon(X)$, $y \in co_\varepsilon(Y)$ and $X \cap Y \neq \emptyset$, then

$$\sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \lambda_{ij} \mathbb{1}_{\{d(x_i, y_j) \neq 0\}} \leq 1 - \varepsilon^2, \tag{3.13}$$

where λ_{ij} were introduced in part (b).

Proof. (a) Mimicking the idea introduced at the beginning of this section, since $x \in co(X)$ it follows that there exists a positive integer m such that $z \in X_m$, where $X_{m+1} = \tilde{\psi}(X_m)$ with $X_0 = X$. Further, there exist $z_1, z_2 \in X_{m-1}$ and $\lambda_{12} \in [0,1]$ such that $z = \psi(z_1, z_2, \lambda_{12})$. Using the definition of the convex structure, it follows that the distance between z and y can be upper bounded by

$$d(x,y) \leq \lambda_{12} d(z_1, y) + (1 - \lambda_{12}) d(z_2, y).$$

Inductively decomposing z_1, z_2 and their *children*, it can be easily argued that

$$d(x,y) \leq \sum_{i=1}^{n_x} \lambda_i d(x_i, y),$$

for some positive weights $\lambda_i \geq 0$ summing up to one.

(b) To obtain (3.11) we proceed as in part (a) and obtain upper bounds on $d(x_i, y)$. More precisely we get that

$$d(x_i, y) \leq \sum_{j=1}^{n_y} \mu_j d(x_i, y_j), \quad \forall i,$$

with $\mu_j > 0$ and $\sum_{j=1}^{n_y} \mu_j = 1$, and it follows that

$$d(x,y) \leq \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \lambda_{ij} d(x_i, y_j),$$

where $\lambda_{ij} = \lambda_i \mu_j \geq 0$ and $\sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \lambda_{ij} = 1$.

(c) We note that $\lambda_i = \mathcal{W}(P_{x,x_i})$ and $\mu_j = \mathcal{W}(P_{y,y_j})$, $\forall i, j$. But since $x \in co_\varepsilon(X)$ and $y \in co_\varepsilon(Y)$ it immediately follows that $\lambda_i \geq \varepsilon$ and $\mu_j \geq \varepsilon$, and therefore $\lambda_{ij} = \varepsilon^2$.

(d) If $X \cap Y \neq \emptyset$ then there exists at least one pair (i, j) such that $d(x_i, y_j) = 0$. But since $\lambda_{ij} \geq \varepsilon^2$ the inequality (3.13) follows. $\qquad \square$

3.3 Problem formulation and statement of the main result

We consider a convex metric space (X, d) and a set of n agents indexed by i which take values on X. Denoting by k the time index, the agents exchange information based on a communication network modeled by a time varying graph $G(k) = (V, E(k))$, where V is the finite set of vertices (the agents) and $E(k)$ is the set of edges. An edge (communication link) $e_{ij}(k) \in E(k)$ exists if node i receives information from node j. Each agent has an initial value in X. At each subsequent time-slot is adjusting its value based on the observations about the values of its neighbors. The goal of the agents is to asymptotically agree on the same value. In what follows we denote by $x_i(k) \in X$ the value or *state* of agent i at time k.

Definition 3.3.1. *We say that the agents asymptotically reach* consensus *(or agreement) if*

$$\lim_{k \to \infty} d(x_i(k), x_j(k)) = 0, \ \forall i, j, \ i \neq j. \tag{3.14}$$

Similar to the communication models used in [52], [4], [34], we impose minimal assumptions on the connectivity of the communication graph $G(k)$. Basically these assumption consists of having the communication graph connected *infinitely often* and having *bounded intercommunication interval* between neighboring nodes.

Assumption 3.3.1 (Connectivity). *The graph* (V, E_∞) *is connected, where* E_∞ *is the set of*

edges (i, j) representing agent pairs communicating directly infinitely many times, i.e.,

$$E_\infty = \{(i, j) \mid (i, j) \in E(k) \text{ for infinitely many indices } k\}$$

Assumption 3.3.2 (Bounded intercommunication interval)**.** *There exists an integer $B \geq 1$ such that for every $(i, j) \in E_\infty$ agent j sends its information to the neighboring agent i at least once every B consecutive time slots, i.e. at time k or at time $k+1$ or ... or (at latest) at time $k + B - 1$ for any $k \geq 0$.*

Assumption 3.3.2 is equivalent to the existence of an integer $B \geq 1$ such that

$$(i, j) \in E(k) \cup E(k+1) \cup \ldots \cup E(k + B - 1), \ \forall (i, j) \in E_\infty \text{ and } \forall k.$$

Let $\mathcal{N}_i(k)$ denote the communication neighborhood of agent i, which contains all nodes sending information to i at time k, i.e. $\mathcal{N}_i(k) = \{j \mid e_{ij}(k) \in E(k)\} \cup \{i\}$, which by convention contains the node i itself. We denote by $A_i(k) \triangleq \{x_j(k), \forall j \in \mathcal{N}_i(k)\}$ the set of the states of agent i's neighbors (its own included), and by $A(k) \triangleq \{x_i(k), i = 1 \ldots n\}$ the set of all states of the agents.

The following theorem states our main result regarding the asymptotic agreement problem on metric convex space.

Theorem 3.3.1. *Let Assumptions 3.3.1 and 3.3.2 hold for $G(k)$ and let $\varepsilon < 1$ be a positive scalar sufficiently small. If agents update their state according to the scheme*

$$x_i(k+1) \in co_\varepsilon(A_i(k)), \ \forall i, \tag{3.15}$$

then they asymptotically reach consensus, i.e.

$$\lim_{k \to \infty} d(x_i(k), x_j(k)) = 0, \ \forall i, j, \ i \neq j. \tag{3.16}$$

Remark 3.3.1. *We would like to point out that the result refers strictly to the convergence of the distances between states and not to the convergence of the states themselves. It may be the case that the sequences $\{x_i(k)\}_{k \geq 0}$ $i = 1 \ldots n$ do not have a limit and still the distances $d(x_i(k), x_j(k))$ decrease to zero as k goes to infinity. In other words the agents asymptotically agree on the same value which may be very well variable. However, as stated in the next corollary this is not the case and in fact the states of the agents do converge to the same value.*

Corollary 3.3.1. *Let Assumptions 3.3.1 and 3.3.2 hold for $G(k)$ and let $\varepsilon < 1$ be a positive scalar sufficiently small. If agents update their state according to the scheme*

$$x_i(k+1) \in co_\varepsilon(A_i(k)), \ \forall i, \tag{3.17}$$

then there exists $x^* \in X$ such that

$$\lim_{k \to \infty} d(x_i(k), x^*) = 0, \ \forall i. \tag{3.18}$$

We will give the proofs for both Theorem 3.3.1 and Corollary 3.3.1 in the subsequent section.

Remark 3.3.2. *A procedure for generating points that are guaranteed to belong to $co_\varepsilon(A_i(k))$ is described in Remark 3.2.2. The idea of picking $x_i(k+1)$ from $co_\varepsilon(A_i(k))$ rather than $co(A_i(k))$ is in the same spirit of the assumption imposed on the non-zero consensus weights in [51], [34], [4], i.e. they are assumed lower bounded by a positive, sub-unitary scalar. Setting $x_i(k+1) \in co(A_i(k))$ may not necessarily guarantee asymptotic convergence to consensus. Indeed, consider the case where $X = \mathbb{R}$ with the standard Euclidean distance. A convex structure on $\mathbb{R}$ is given by $\psi(x, y, \lambda) = \lambda x + (1 - \lambda)y$, for*

66

any $x, y \in \mathbb{R}$ and $\lambda \in [0, 1]$. Assume that we have two agents which exchange informa-

tion at all time slots and therefore $A_1(k) = \{x_1(k), x_2(k)\}$, $A_2(k) = \{x_1(k), x_2(k)\}$, $\forall k \geq 0$.

Let $x_1(k + 1) = \lambda(k)x_1(k) + (1 - \lambda(k))x_2(k)$, where $\lambda(k) = 1 - 0.1e^{-k}$ and let $x_2(k + 1) =$

$\mu(k)x_1(k) + (1 - \mu(k))x_2(k)$, where $\mu(k) = 0.1e^{-k}$. Obviously, $x_i(k + 1) \in co(A_i(k))$, $i = 1, 2$

for all $k \geq 0$. It can be easily argued that

$$d(x_1(k + 1), x_2(k + 1)) \leq (\lambda(k)(1 - \mu(k)) + \mu(k)(1 - \lambda(k))) \, d(x_1(k, x_2(k))). \qquad (3.19)$$

We note that $\lim_{K \to \infty} \prod_{k=0}^{K} (\lambda(k)(1 - \mu(k)) + (1 - \lambda(k))\mu(k)) = \lim_{K \to \infty} \prod_{k=0}^{K} (1 - 0.2e^{-k} +$

$0.02e^{-2k}) = 0.73$ and therefore under inequality (3.19) asymptotic convergence to con-

sensus is not guaranteed. In fact it can be explicitly shown that the agents do not reach

consensus. From the dynamic equation governing the evolution of $x_i(k)$, $i = 1, 2$, we can

write

$$\mathbf{x}(k + 1) = \begin{pmatrix} \lambda(k) & 1 - \lambda(k) \\ \mu(k) & 1 - \mu(k) \end{pmatrix} \mathbf{x}(k), \ \mathbf{x}(0) = \mathbf{x}_0,$$

where $\mathbf{x}(k)^T = [x_1(k), x_2(k)]$, and we obtain that

$$\lim_{k \to \infty} \mathbf{x}(k) = \begin{pmatrix} 0.8540 & 0.1451 \\ 0.1451 & 0.8540 \end{pmatrix} \mathbf{x}_0$$

and therefore it can be easily seen that consensus is not reached from any initial states.

3.4   Proof of the main result

This section is divided in three parts. In the first part we use the results of Section

3.2.2 regarding the convex hull of a finite set and show that the entries of the vector

of distances between the states of the agents at time $k + 1$ are upper bounded by linear

combinations of the entries of the same vector but at time k. The coefficients of the linear combinations are the entries of a time varying matrix for which we prove a number of properties (Lemma 3.4.1). In the second part we analyze the properties of the transition matrix of the aforementioned time varying matrix (Lemma 3.4.2). The last part is reserved to the proof of Theorem 3.3.1.

Lemma 3.4.1. *Given a small enough positive scalar $\varepsilon < 1$, assume that agents update their states according to the scheme $x_i(k+1) \in co_\varepsilon(A_i(k))$, for all i. Let $\boldsymbol{d}(k) \triangleq (d(x_i(k), x_j(k)))$ for $i \neq j$ be the N dimensional vector of all distances between the states of the agents, where $N = \frac{n(n-1)}{2}$. Then we obtain that*

$$\boldsymbol{d}(k+1) \leq \boldsymbol{W}(k)\boldsymbol{d}(k), \; \boldsymbol{d}(0) = \boldsymbol{d}_0, \tag{3.20}$$

where the $N \times N$ dimensional matrix $\boldsymbol{W}(k)$ has the following properties:

(a) $\boldsymbol{W}(k)$ is non-negative and there exits a positive scalar $\eta \in (0,1)$ such that

$$[\boldsymbol{W}(k)]_{\bar{i}\bar{i}} \geq \eta, \; \forall \; \bar{i}, k \tag{3.21}$$

$$[\boldsymbol{W}(k)]_{\bar{i}\bar{j}} \geq \eta, \; \forall \; [\boldsymbol{W}(k)]_{\bar{i}\bar{j}} \neq 0, \; \bar{i} \neq \bar{j}, \; \forall \; k. \tag{3.22}$$

(b) If $\mathcal{N}_i(k) \cap \mathcal{N}_j(k) \neq \emptyset$, then the row $\bar{i}$ of matrix $\boldsymbol{W}(k)$, corresponding to the pair of agents (i, j), has the property

$$\sum_{\bar{j}=1}^{N} [\boldsymbol{W}(k)]_{\bar{i}\bar{j}} \leq 1 - \eta, \tag{3.23}$$

where η is the same as in part (a).

(c) If $\mathcal{N}_i(k) \cap \mathcal{N}_j(k) = \emptyset$ then the row $\bar{i}$ corresponding to the pair of agents (i, j) sums up

to one, i.e.

$$\sum_{\bar{j}=1}^{N} [W(k)]_{\bar{i}\bar{j}} = 1. \tag{3.24}$$

In particular if $G(k)$ is completely disconnected (i.e. agents do not send any informa-

tion), then $W(k) = I$.

(d) the rows of $W(k)$ sum up to a value smaller or equal then one, i.e.

$$\sum_{\bar{j}=1}^{N} [W(k)]_{\bar{i}\bar{j}} \leq 1, \ \forall \ \bar{i}, k. \tag{3.25}$$

Proof. Given two agents i and j, by part (b) of Proposition 3.2.2 the distance between their states can be upper bounded by

$$d(x_i(k+1), x_j(k+1)) \leq \sum_{p \in \mathcal{N}_i(k), q \in \mathcal{N}_j(k)} w_{pq}^{ij}(k) d(x_p(k), x_q(k)), \ i \neq j, \tag{3.26}$$

where $w_{pq}^{ij}(k) \geq 0$ and $\sum_{p \in \mathcal{N}_i(k), q \in \mathcal{N}_j(k)} w_{pq}^{ij}(k) = 1$. By defining $W(k) \triangleq (w_{pq}^{ij}(k))$ for $i \neq j$ and

$p \neq q$ (where the pairs (i, j) and (p, q) refer to the rows and columns of $W(k)$, respectively),

inequality (3.20) follows. We continue with proving the properties of matrix $W(k)$.

(a) Since all $w_{pq}^{ij}(k) \geq 0$ for all $i \neq j$, $p \in \mathcal{N}_i(k)$ and $q \in \mathcal{N}_j(k)$ we obtain that $W(k)$ is

non-negative. By part (c) of Proposition 3.2.2, there exists $\eta \triangleq \varepsilon^2$ such that $w_{pq}^{ij}(k) \geq \eta$ for

all non-zero entries of $W(k)$. Also, since $i \in \mathcal{N}_i(k)$ and $j \in \mathcal{N}_j(k)$ for all $k \geq 0$ it follows

that the term $w_{ij}^{ij}(k) d(x_i(k), x_j(k))$, with $w_{ij}^{ij}(k) \geq \eta$ will always be present in the right-hand

side of the inequality (3.26), and therefore $W(k)$ has positive diagonal entries.

(b) Follows from part (d) of Proposition 3.2.2, with $\eta = \varepsilon^2$.

(c) If $\mathcal{N}_i(k) \cap \mathcal{N}_j(k) = \emptyset$ then no terms of the form $w_{pp}^{ij}(k) d(x_p(k), x_p(k))$ will appear

in the sum of the right hand side of inequality (3.26). Hence $\sum_{p \in \mathcal{N}_i(k), q \in \mathcal{N}_j(k)} w_{pq}^{ij}(k) = 1$

69

and therefore

$$\sum_{\bar{j}=1}^{N}[\mathbf{W}(k)]_{\bar{i}\bar{j}} = 1.$$

If $G(k)$ is completely disconnected, then the sum of the right hand side of inequality (3.26) will have only the term $w_{ij}^{ij}(k)d(x_i(k), x_j(k))$ with $w_{ij}^{ij}(k) = 1$, for all $i, j = 1 \ldots n$. Therefore $\mathbf{W}(k)$ is the identity matrix.

(d) The result follows from parts (b) and (c). □

Let $\bar{G}(k) = (\bar{V}, \bar{E}(k))$ be the underlying graph of $\mathbf{W}(k)$ and let $\bar{i}$ and $\bar{j}$ refer to the rows and columns of $\mathbf{W}(k)$, respectively. Note the under this notation, index $\bar{i}$ corresponds to a pair (i, j) of distinct agents. It is not difficult to see that the set of edges of $\bar{G}(k)$ is given by

$$\bar{E}(k) = \{((i, j), (p, q)) \mid (i, p) \in E(k), (j, q) \in E(k), i \neq j, \ p \neq q\}. \tag{3.27}$$

Proposition 3.4.1. *Let Assumptions 3.3.1 and 3.3.2 hold for $G(k)$. Then, similar proper- ties hold for $\bar{G}(k)$ as well, i.e.*

(a) the graph $(\bar{V}, \bar{E}_\infty)$ is connected, where

$$\bar{E}_\infty = \{(\bar{i}, \bar{j}) \mid (\bar{i}, \bar{j}) \in \bar{E}(k) \text{ infinetly many indices } k\};$$

(b) there exists an integer $\bar{B} \geq 1$ such that every $(\bar{i}, \bar{j}) \in \bar{E}_\infty$ appears at least once every $\bar{B}$ consecutive time slots, i.e. at time k or at time $k + 1$ or $\ldots$ or (at latest) at time $k + \bar{B} - 1$ for any $k \geq 0$.

Proof. It is not difficult to observe that similar to (3.27), $\bar{E}_\infty$ is given by

$$\bar{E}_\infty = \{((i, j), (p, q)) \mid (i, p) \in E_\infty, (j, p) \in E_\infty, p \neq q, i \neq j\}. \tag{3.28}$$

(a) Showing that $(\bar{V}, \bar{E}_\infty)$ is connected is equivalently to showing that for any two pairs (i, j) and (p, q) there exits a path connecting them. Since (V, E_∞) is assumed connected, there exits a path $i_0 \to i_1 \to \ldots, \to i_{l-1} \to i_l$, for some $l \le n$, such that $i_0 = p$ and $i_l = i$. From (3.28), it is easily argued that $(i_0, j) \to (i_1, j) \to \ldots \to (i_{l-1}, j) \to (i_l, j)$ represents a path connecting (i, j) with (p, j). Similarly, there exits a path $j_0 \to j_1 \to \ldots \to j_{m-1} \to j_m$ for some $m \le n$, such that $j_0 = q$ and $j_m = j$. Therefore, $(p, j_0) \to (p, j_1) \to \ldots \to (p, j_{m-1}) \to (p, j_m)$ is a path connecting (p, j) with (p, q) and it follows that (i, j) and (p, q) are connected.

(b) Let $((i, j), (p, q))$ be an edge in $\bar{E}_\infty$ or equivalently $(i, p) \in E_\infty$ and $(j, q) \in E_\infty$. By Assumption 3.3.2, we have that for any $k \ge 0$

$$(i, p) \in E(k) \cup E(k+1) \ldots \cup E(k+B-1),$$

$$(j, q) \in E(k) \cup E(k+1) \ldots \cup E(k+B-1),$$

where the scalar B was introduced in Assumption 3.3.2. But this also implies that

$$(\bar{i}, \bar{j}) \in \bar{E}(k) \cup \bar{E}(k+1) \cup \ldots \cup \bar{E}(k+B-1), \; \forall (\bar{i}, \bar{j}) \in \bar{E}_\infty.$$

Choosing $\bar{B} \triangleq B$, the result follows. □

Let $\Phi(k, s) \triangleq \mathbf{W}(k-1)\mathbf{W}(k-2)\cdots\mathbf{W}(s)$, with $\Phi(k, k) = \mathbf{W}(k)$ denote the transition matrix of $\mathbf{W}(k)$ for any $k \ge s$. It should be obvious from the properties of $\mathbf{W}(k)$ that $\Phi(k, s)$ is a non-negative matrix with positive diagonal entries and $\|\Phi(k, s)\|_\infty \le 1$ for any $k \ge s$.

Lemma 3.4.2. *Let $\mathbf{W}(k)$ be the matrix introduced in Lemma 3.4.1. Let Assumptions 3.3.1 and 3.3.2 hold for $G(k)$. Then there exits a row index $\bar{i}^*$ such that*

$$\sum_{\bar{j}=1}^{N} [\Phi(s+m, s)]_{\bar{i}^* \bar{j}} \le 1 - \eta^m \; \forall \; s, m \ge \bar{B} - 1, \tag{3.29}$$

71

where η is the lower bound on the non-zero entries of $\mathbf{W}(k)$ and $\bar{B}$ is the positive integer from the part (b) of the Proposition 3.4.1.

Proof. Let $(i^*, j^*) \in E_\infty$ be a pair of agents. By Assumptions 3.3.1 and 3.3.2, there exits a positive integer $s' \in \{s, s+1, \ldots, s+\bar{B}-1\}$ such that agent j^* sends information to agent i^* at time s'. This implies that $\mathcal{N}_{i^*}(k) \cap \mathcal{N}_{j^*}(k) \neq \varnothing$ and by part (b) of Lemma 3.4.1, we have that

$$\sum_{\bar{j}=1}^{N} [\mathbf{W}(s')]_{\bar{i}^*\bar{j}} \leq 1 - \eta,$$

where $\bar{i}^*$ is the index corresponding to the pair (i^*, j^*). The sum of the $\bar{i}^*$ row of transition matrix $\Phi(s'+1, s)$ can be expressed as

$$\sum_{\bar{j}=1}^{N} [\Phi(s'+1, s)]_{\bar{i}^*\bar{j}} = \sum_{\bar{j}=1}^{N} [\mathbf{W}(s')]_{\bar{i}^*\bar{j}} \sum_{\bar{h}=1}^{N} [\Phi(s', s)]_{\bar{j}\bar{h}}.$$

But since $\|\Phi(k, s)\|_\infty \leq 1$ for any $k \geq s$, we have that $\sum_{\bar{h}=1}^{N} [\Phi(s', s)]_{\bar{j}\bar{h}} \leq 1$ for any $\bar{j}$, and therefore

$$\sum_{\bar{j}=1}^{N} [\Phi(s'+1, s)]_{\bar{i}^*\bar{j}} \leq 1 - \eta. \tag{3.30}$$

We can write $\Phi(s'+2, s) = \mathbf{W}(s'+1)\Phi(s'+1, s)$ and it follows that the $\bar{i}^*$ row sum of $\Phi(s'+2, s)$ can be expressed as

$$\sum_{\bar{j}=1}^{N} [\Phi(s'+2, s)]_{\bar{i}^*\bar{j}} = \sum_{\bar{j}=1}^{N} [\mathbf{W}(s'+1)]_{\bar{i}^*\bar{j}} \sum_{\bar{h}=1}^{N} [\Phi(s'+1, s)]_{\bar{j}\bar{h}}.$$

Since $\sum_{\bar{h}=1}^{N} [\Phi(s'+1, s)]_{\bar{j}\bar{h}} \leq 1$ for any $\bar{j}$ it follows that

$$\sum_{\bar{j}=1}^{N} [\Phi(s'+2, s)]_{\bar{i}^*\bar{j}} \leq [\mathbf{W}(s'+1)]_{\bar{i}^*\bar{i}^*} \sum_{\bar{h}=1}^{N} [\Phi(s'+1, s)]_{\bar{i}^*\bar{h}} + \sum_{\bar{j}=1, \bar{j}\neq\bar{i}^*}^{N} [\mathbf{W}(s'+1)]_{\bar{i}^*\bar{j}} \leq$$

$$\leq [\mathbf{W}(s'+1)]_{\bar{i}^*\bar{i}^*}(1-\eta) + \sum_{\bar{j}=1, \bar{j}\neq\bar{i}^*}^{N} [\mathbf{W}(s'+1)]_{\bar{i}^*\bar{j}} \leq \sum_{\bar{j}=1}^{N} [\mathbf{W}(s'+1)]_{\bar{i}^*\bar{j}} - \eta[\mathbf{W}(s'+1)]_{\bar{i}^*\bar{i}^*} \leq 1 - \eta^2,$$

since $[\mathbf{W}(s'+1)]_{\bar{i}^*\bar{i}^*} \geq \eta$. By induction it can be easily argued that

$$\sum_{\bar{j}=1}^{N}[\Phi(s'+m,s)]_{\bar{i}^*\bar{j}} \leq 1-\eta^m, \ \forall m \geq 0. \tag{3.31}$$

Note that by Assumption 3.3.2, a pair (i,j) can exchange information at $s' = s$ the earliest

or at $s' = s+B-1$ the latest. From (3.31) we obtain that for $s' = s+B-1$

$$\sum_{\bar{j}=1}^{N}[\Phi(s+B-1+m,s)]_{\bar{i}^*\bar{j}} \leq 1-\eta^m, \ \forall m \geq 0, \tag{3.32}$$

and for $s' = s$

$$\sum_{\bar{j}=1}^{N}[\Phi(s+m,s)]_{\bar{i}^*\bar{j}} \leq 1-\eta^m, \ \forall m \geq 0,$$

or

$$\sum_{\bar{j}=1}^{N}[\Phi(s+B-1+m,s)]_{\bar{i}^*\bar{j}} \leq 1-\eta^{m+B-1}, \ \forall m \geq 0, \tag{3.33}$$

From (3.32) and (3.33) we get

$$\sum_{\bar{j}=1}^{N}[\Phi(s+B-1+m,s)]_{\bar{i}^*\bar{j}} \leq 1-\eta^{m+B-1}, \ \forall s, m \geq 0,$$

or equivalently

$$\sum_{\bar{j}=1}^{N}[\Phi(s+m,s)]_{\bar{i}^*\bar{j}} \leq 1-\eta^m, \ \forall m \geq B-1. \tag{3.34}$$

$\square$

Corollary 3.4.1. *Let $\mathbf{W}(k)$ be the matrix introduced in Lemma 3.4.1 and let Assumptions*

3.3.1 and 3.3.2 hold for $G(k)$. We then have

$$[\Phi(s+(N-1)\bar{B}-1,s)]_{ij} \geq \eta^{(N-1)\bar{B}} \ \forall s, i, j, \tag{3.35}$$

where η is the lower bound on the non-zero entries of $\mathbf{W}(k)$ and $\bar{B}$ is the positive integer

from the part (b) of the Proposition 3.4.1.

Proof. By Proposition 3.4.1 and Lemma 3.4.1 all the assumptions of Lemma 2, [34] are satisfied, from which the result follows. □

We are now ready to prove **Theorem 3.3.1** and **Corollary 3.3.1**.

3.4.1   Proof of Theorem 3.3.1

We have that the vector of distances between the states of the agents respects the inequality

$$\mathbf{d}(k+1) \leq \mathbf{W}(k)\mathbf{d}(k),$$

where the properties of $\mathbf{W}(k)$ are described by Lemma 3.4.1.

It immediately follows that

$$\|\mathbf{d}(k+1)\|_\infty \leq \|\mathbf{d}(k)\|_\infty, \text{ for } k \geq 0. \tag{3.36}$$

Let $\bar{B}_0 \triangleq (N-1)\bar{B} - 1$, where $\bar{B}$ is the positive integer from the part (b) of the Proposition 3.4.1. In the following we show that all row sums of $\Phi(s + 2B_0, s)$ are upper-bounded by a positive scalar strictly less than one. Indeed since $\Phi(s + 2\bar{B}_0, s) = \Phi(s + 2\bar{B}_0, s + \bar{B}_0)\Phi(s + \bar{B}_0, s)$ we obtain that

$$\sum_{\bar{j}=1}^N [\Phi(s + 2\bar{B}_0, s)]_{\bar{i}\bar{j}} = \sum_{\bar{j}=1}^N [\Phi(s + 2\bar{B}_0, s + \bar{B}_0)]_{\bar{i}\bar{j}} \sum_{\bar{h}=1}^N [\Phi(s + \bar{B}_0, s)]_{\bar{j}\bar{h}}, \ \forall \bar{i}.$$

By Lemma 3.4.2 we have that there exists a row $\bar{j}^*$ such that

$$\sum_{\bar{h}=1}^N [\Phi(s + \bar{B}_0, s)]_{\bar{j}^*\bar{h}} \leq 1 - \eta^{\bar{B}_0}, \forall s,$$

74

and since $\sum_{\bar{h}=1}^{N}[\Phi(s+\bar{B}_0,s)]_{\bar{j}\bar{h}} \leq 1$ for any $\bar{j}$, we get

$$\sum_{\bar{j}=1}^{N}[\Phi(s+2\bar{B}_0,s)]_{\bar{i}\bar{j}} \leq \sum_{\bar{j}=1,\bar{j}\neq\bar{j}^*}^{N}[\Phi(s+2\bar{B}_0,s+\bar{B}_0)]_{\bar{i}\bar{j}} + [\Phi(s+2\bar{B}_0,s+\bar{B}_0)]_{\bar{i}\bar{j}^*}(1-\eta^{\bar{B}_0}) =$$

$$= \sum_{\bar{j}=1}^{N}[\Phi(s+2\bar{B}_0,s+\bar{B}_0)]_{\bar{i}\bar{j}} - [\Phi(s+2\bar{B}_0,s+\bar{B}_0)]_{\bar{i}\bar{j}^*}\eta^{\bar{B}_0}.$$

By Corollary 3.4.1 it follows that

$$[\Phi(s+2\bar{B}_0,s+\bar{B}_0)]_{\bar{i}\bar{j}} \geq \eta^{\bar{B}_0+1}, \; \forall \bar{i},\bar{j},s,$$

and since $\sum_{\bar{j}=1}^{N}[\Phi(s+2\bar{B}_0,\bar{B}_0)]_{\bar{i}\bar{j}} \leq 1$ we get that

$$\sum_{\bar{j}=1}^{N}[\Phi(s+2\bar{B}_0,s)]_{\bar{i}\bar{j}} \leq 1-\eta^{2\bar{B}_0+1} \; \forall \bar{i},s.$$

Therefore

$$\|\Phi(s+2\bar{B}_0,s)\|_\infty \leq 1-\eta^{2\bar{B}_0+1} \; \forall s. \tag{3.37}$$

It follows that

$$\|\mathbf{d}(t_k)\|_\infty \leq \left(1-\eta^{2\bar{B}_0+1}\right)^k \|\mathbf{d}(0)\|_\infty, \; \forall k \geq 0, \tag{3.38}$$

where $t_k = 2k\bar{B}_0$ which shows that the subsequence $\{\|\mathbf{d}(t_k)\|_\infty\}_{k\geq 0}$ asymptotically converges to zero. Combined with inequality (3.36) we farther obtain that the sequence $\{\|\mathbf{d}(k)\|_\infty\}_{k\geq 0}$ asymptotically converges to zero. Therefore the agents asymptotically reach consensus.

3.4.2 Proof of Corollary 3.3.1

The main idea of the proof consist of showing that the set $co(A(k))$, where $A(k) = \{x_i(k), i = 1\ldots n\}$, converges to a set containing one point.

We first note that since $A_i(k) \subseteq A(k)$ it can be easily argued that $co(A_i(k)) \subseteq co(A(k))$, for all i and k. Also, since $co_\varepsilon(A_i(k)) \subseteq co(A_i(k))$ it follows that $co_\varepsilon(A_i(k)) \subseteq co(A(k))$ and consequently $x_i(k+1) \in co(A(k))$. Therefore, we have that $co(A(k+1)) \subseteq co(A(k))$ for all k and from the theory of limit of sequence of sets, it follows that

$$\liminf co(A(k)) = \limsup co(A(k)) = \lim co(A(k)) = A_\infty,$$

where $A_\infty = \bigcap_{k \geq 0} co(A(k))$. We denote the diameter of the set $A(k)$ by

$$\delta(A(k)) = \sup\{d(x,y) \mid x,y \in A(k)\},$$

and by Proposition 2 of [46] we have that

$$\delta(co(A(k))) = \delta(A(k)).$$

From Theorem 3.3.1 we have that

$$\lim_{k \to \infty} d(x_i(k), x_j(k)) = 0, \ \forall i \neq j,$$

and consequently

$$\lim_{k \to \infty} \delta(A(k)) = \lim_{k \to \infty} \delta(co(A(k))) = 0,$$

which also means that

$$\delta(A_\infty) = 0,$$

i.e. the set A_∞ contains only one point, say $x^* \in X$, or $A_\infty = co(x^*)$, or

$$\lim_{k \to \infty} co(A(k)) = co(x^*).$$

But since $x_i(k+1) \in co_\varepsilon(A_i(k)) \subseteq co(A(k))$ for all i,k it follows that

$$\lim_{k \to \infty} d(x_i(k), x^*) = 0, \forall \ i,$$

i.e. the states of the agents converge to the same point $x^* \in X$.

3.5 Distance between the consensus points and the initial points

In this section we analyze the evolution of the distance between the states of the agents and their initial values under the scheme described by Theorem 3.3.1. This analysis will give us upper bounds on the distance between the consensus point(s) and the initial values of the agents.

Consider distance $d(x_i(k), x_l(0))$ for some i, l and let us assume that $x_i(k+1)$ is chosen according to the scheme described by Theorem 3.3.1, i.e. $x_i(k+1) \in co_\varepsilon(A_i(k))$. By part (a) of Proposition 3.2.2 we can express this distance as

$$d(x_i(k+1), x_l(0)) \leq \sum_{j \in \mathcal{N}_i(k)} \lambda_{ij}(k) d(x_j(k), x_l(0)), \tag{3.39}$$

where $\lambda_{ij}(k) \geq \varepsilon$ and $\sum_{j \in \mathcal{N}_i(k)} \lambda_{ij}(k) = 1$. By defining the n dimensional vector $\mu^l(k) = (d(x_i(k), x_l(0)))$ (where i varies) and the $n \times n$ dimensional matrix $\Lambda(k) = (\lambda_{ij}(k))$, inequality (3.39) can be compactly written as

$$\mu^l(k+1) \leq \Lambda(k)\mu^l(k), \ \mu^l(0) = \mu_0^l. \tag{3.40}$$

where $\Lambda(k)$ is a row stochastic matrix. It is not difficult to note that the underlying graph of $\Lambda(k)$ is $G(k)$ and that in fact inequality (3.40) is valid for any l. In the following proposition we give upper bounds on the distance between the consensus states and the initial values of the states.

Proposition 3.5.1. *Let Assumptions 3.3.1 and 3.3.2 hold for $G(k)$ and let the states of the agents be updates according to the scheme given by Theorem 3.3.1. We then have that*

$$\lim_{k \to \infty} d(x_i(k), x_l(0)) \leq \sum_{j=1}^{n} v_j d(x_j(0), x_l(0)), \ \forall \ i, l, \tag{3.41}$$

77

where $v = (v_j)$ is a vector with positive entries summing up to one satisfying

$$\lim_{k \to \infty} \Lambda(k)\Lambda(k-1)\cdots\Lambda(0) = \mathbf{1}v^T, \tag{3.42}$$

and where $\mathbf{1}$ is the n dimensional vector of all ones and $\Lambda(k)$ is the matrix defined in inequality (3.40).

Proof. Our assumptions fit the assumptions of Lemmas 3 and 4 of [34], from where (3.42) follows. Therefore by inequality (3.40) the result follows. $\qquad\square$

Remark 3.5.1. *If in addition to the assumptions of Proposition 3.5.1 we also assume that $\Lambda(k)$ is doubly stochastic, then by Proposition 1 of [34] we get that*

$$\lim_{k \to \infty} \Lambda(k)\Lambda(k-1)\cdots\Lambda(0) = \frac{1}{n}\mathbf{1}\mathbf{1}^T.$$

Therefore, inequality (3.41) gets simplified to

$$\lim_{k \to \infty} d(x_i(k), x_l(0)) \le \frac{1}{n}\sum_{j=1}^{n} d(x_j(0), x_l(0)), \ \forall i.$$

The assumptions in this remark correspond to the assumptions for the average consensus problem in Euclidean spaces. For the aforementioned case, the consensus point is given by the average of the initial points, i.e. $x_{av} = \frac{1}{n}\sum_{i=1}^{n} x_i(0)$. It can be easily check that indeed x_{av} satisfies

$$\|x_{av} - x_l(0)\| \le \frac{1}{n}\sum_{j=1}^{n} \|x_j(0) - x_l(0)\|,$$

where $\|\cdot\|$ represents the euclidean norm.

3.6 Application - Asymptotic consensus of opinion

Social networks play a central role in the sharing of information and formation of opinions. This is true in the context of advising friends on which movies to see, relaying

information about the abilities and fit of a potential new employee in a firm, debating the merits of politicians. In the following we consider a scenario in which a group of agents try to agree on a common opinion. Assume for example that a group of friends would like to go to see a movie. Different members of the group may suggest different movies. A member of the group discusses with all or just some of his/her friends to find out about their opinions. This member gives some weight (importance) to the opinion of his friends based on the trust in their *expertise*. For instance some members of the group are more informed about the quality of the proposed movies, and therefore there opinions may have a heavier influence on the final decision. By repeatedly discussing among themselves, the group of friends have to choose one of the movies.

In the following we mathematically formalize the scenario described above and show that we can use the framework introduced in the previous sections to give an algorithm which ensures asymptotic consensus on opinions. We model the opinion of a member of the group (agent) as a discrete random variable. Under an appropriate metric and by providing a convex structure we show that the metric space of discrete random variable is convex . In addition, we analyze in more detail the convex hull of a finite set; this analysis is possible since the convex structure is given explicitly. We give an iterative algorithm that ensures agreement of opinion, which is based on Theorem 3.3.1 and provide some numerical simulations.

3.6.1 Geometric framework

Let s be a positive integer, let $S = \{1, 2, \ldots, s\}$ be a finite set and let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. We denote by $\mathcal{X}$ the space of discrete measurable functions (random variable) on $(\Omega, \mathcal{F}, \mathcal{P})$ with values in S.

We introduce the operator $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, defined as

$$d(X, Y) = E[\rho(X, Y)],$$

where $\rho : \mathbb{R} \times \mathbb{R} \to \{0, 1\}$ is the discrete metric, i.e.

$$\rho(x, y) = \begin{cases} 1 & x \neq y \\ 0 & x = y \end{cases}$$

It is not difficult to note that the operator d can also be written as $d(X, Y) = E[\mathbb{1}_{\{X \neq Y\}}] = Pr(X \neq Y)$, where $\mathbb{1}_{\{X \neq Y\}}$ is the indicator function of the event $\{X \neq Y\}$.

We note that the operator d satisfies the following properties

1. For any $X, Y \in \mathcal{X}$, $d(X, Y) = 0$ if and only if $X = Y$ with probability one.

2. For any $X, Y, Z \in \mathcal{X}$, $d(X, Z) + d(Y, Z) \geq d(X, Y)$ with probability one,

and therefore is a metric on $\mathcal{X}$. The set $\mathcal{X}$ together with the operator d define the *metric space* $(\mathcal{X}, d)$.

Let $\theta \in \{1, 2\}$ be an independent random variable, with probability mass function $Pr(\theta = 1) = \lambda$ and $Pr(\theta = 2) = 1 - \lambda$, where $\lambda \in [0, 1]$. We define the mapping $\psi : \mathcal{X} \times \mathcal{X} \times [0, 1] \to \mathcal{X}$ given by

$$\psi(X_1, X_2, \lambda) = \mathbb{1}_{\{\theta=1\}} X_1 + \mathbb{1}_{\{\theta=2\}} X_2, \ \forall X_1, X_2 \in \mathcal{X}, \lambda \in [0, 1]. \tag{3.43}$$

Proposition 3.6.1. *The mapping ψ is a convex structure on $\mathcal{X}$.*

Proof. For any $U, X_1, X_2 \in \mathcal{X}$ and $\lambda \in [0,1]$ we have

$$d(U, \psi(X_1, X_2, \lambda)) = E[\rho(U, \psi(X_1, X_2, \lambda))] = E[E[\rho(U, \psi(X_1, X_2, \lambda))|U, X_1, X_2]] =$$

$$= E[E[\rho(U, \mathbb{1}_{\{\theta=1\}}X_1 + \mathbb{1}_{\{\theta=2\}}X_2)]|U, X_1, X_2] = E[\lambda\rho(U, X_1) + (1-\lambda)\rho(U, X_2)] =$$

$$= \lambda d(U, X_1) + (1-\lambda)d(U, X_2).$$

$\square$

From the above proposition it follows that $(\mathcal{X}, d, \psi)$ is a *convex metric space*.

The next theorem characterizes the convex hull of a finite set in $\mathcal{X}$.

Theorem 3.6.1. *Let n be a positive integer and let $A = \{X_1, \ldots, X_n\}$ be a set of points in $\mathcal{X}$. Consider the independent random variable θ taking values in the finite set $\{1, \ldots, n\}$, with probability measure given by $Pr(\omega : \theta(\omega) = i) = w_i$, for some non-negative scalars w_i, with $\sum_{i=1}^{n} w_i = 1$. Then*

$$co(A) = \left\{ Z \in \mathcal{X} \mid Z = \sum_{i=1}^{n} \mathbb{1}_{\{\theta=i\}}X_i, \ \forall w_i \geq 0, \ \sum_{i=1}^{n} w_i = 1 \right\}. \tag{3.44}$$

Proof. We recall from Proposition 3.2.1 that the convex hull of A is given by

$$co(A) = \lim A_m = \bigcup_{m=1}^{\infty} A_m,$$

where $A_m = \tilde{\psi}(A_{m-1})$, with $A_1 = \tilde{\psi}(A)$. Also, since A_m is an increasing sequence, clearly $A \subset A_m$ for all $m \geq 1$. We define the set

$$\mathcal{K}(A) \triangleq \left\{ Z \in \mathcal{X} \mid Z = \sum_{i=1}^{n} \mathbb{1}_{\{\theta=i\}}X_i, \ \forall w_i \geq 0, \ \sum_{i=1}^{n} w_i = 1 \right\}.$$

The proof is structured in two parts. In the first part we show that any point in $\mathcal{K}(A)$ belongs to the convex hull of A, while in the second part we show that any point in $co(A)$ belongs to $\mathcal{K}(A)$ as well.

Let $Z \in \mathcal{K}(A)$ i.e. $Z = \sum_{i=1}^{n} \mathbb{1}_{\{\theta=i\}} X_i$ where $Pr(\theta = i) = w_i$, for some $w_i \geq 0$, $\sum_{i=1}^{n} w_i = 1$. The random variable θ is defined such that $\theta(\omega_i) = i$ and $Pr(\omega_i) = w_i$. Let $\Omega_i = \{\omega_1^i, \omega_2^i\}$, $i = 1 \ldots n-1$ be a set of independent sample spaces (i.e. the elementary events ω_j^i and ω_p^l are independent for any $l \neq i$ and for any j). We define the probability measure for each of the events in Ω_i as

$$Pr(\omega_1^i) = \frac{w_1 + \ldots + w_{i-1}}{w_1 + \ldots + w_i},$$

$$Pr(\omega_2^i) = \frac{w_i}{w_1 + \ldots + w_i},$$

for $i = 1 \ldots n-1$. We consider the following succession of events from Ω_i

$$
\begin{aligned}
S_1 &= \left\{ \omega_1^1 \omega_1^2 \ldots \omega_1^{n-1} \right\}, \\
S_2 &= \left\{ \omega_2^1 \omega_1^2 \ldots \omega_1^{n-1} \right\}, \\
S_i &= \bigcup_{j_1 \ldots j_{i-2}=1}^{2} \left\{ \omega_{j_1}^1 \ldots \omega_{j_{i-2}}^{i-2} \omega_2^{i-1} \omega_1^i \ldots \omega_1^{n-1} \right\}, i = 3 \ldots n-1, \\
S_n &= \bigcup_{j_1 \ldots j_{n-2}=1}^{2} \left\{ \omega_{j_1}^1 \ldots \omega_{j_{n-2}}^{n-2} \omega_2^{n-1} \right\}.
\end{aligned}
\tag{3.45}
$$

For example, for $n = 4$ (3.45) becomes

$$S_1 = \{\omega_1^1 \omega_1^2 \omega_1^3\},$$

$$S_2 = \{\omega_2^1 \omega_1^2 \omega_1^3\},$$

$$S_3 = \{\omega_1^1 \omega_2^2 \omega_1^3\} \cup \{\omega_2^1 \omega_2^2 \omega_1^3\}$$

$$S_4 = \{\omega_1^1 \omega_1^2 \omega_2^3\} \cup \{\omega_2^1 \omega_1^2 \omega_2^3\} \cup \{\omega_1^1 \omega_2^2 \omega_2^3\} \cup \{\omega_2^1 \omega_2^2 \omega_2^3\}.$$

Using the independence assumption on the events from Ω_i is not difficult no see that

$$Pr(S_i) = w_i, \quad i = 1 \ldots n.$$

Assume that each event ω_i that we observe can be decomposed in a succession of independent events from Ω_i, which are invisible to the observer. In particular let

$$\omega_i = S_i, i = 1\ldots n.$$

The particular decomposition of event ω_i in a set of intermediate, independent events given by S_i makes sense since both ω_i and S_i have the same probability measure. It immediately follows that

$$\mathbb{1}_{\{\omega:\theta(\omega)=i\}} = \mathbb{1}_{\{\omega_i\}} = \mathbb{1}_{\{S_i\}}. \tag{3.46}$$

Let us now define the random variables $\theta_i : \Omega_i \to \{i, i+1\}$, where

$$\theta_i(\omega_1^i) = i, \ \theta_i(\omega_2^i) = i+1,$$

for $i = 1\ldots n-1$. Obviously

$$Pr(\theta_i = i) = \frac{w_1 + \ldots + w_{i-1}}{w_1 + \ldots + w_i}, \ Pr(\theta_i = i+1) = \frac{w_i}{w_1 + \ldots + w_i},$$

and θ_i are independent random variables.

From (3.45) and (3.46) together with the independence of the random variables θ_i the following equalities in terms of the indicator function are satisfied

$$\mathbb{1}_{\{\theta=1\}} = \Pi_{j=1}^{n-1}\mathbb{1}_{\{\theta_j=j\}}$$

$$\mathbb{1}_{\{\theta=i\}} = \mathbb{1}_{\{\theta_{i-1}=i\}}\Pi_{j=i}^{n-1}\mathbb{1}_{\{\theta_j=j\}}, \ i = 2\ldots n-1 \tag{3.47}$$

$$\mathbb{1}_{\{\theta=n\}} = \mathbb{1}_{\{\theta_{n-1}=n\}}.$$

From (3.47) it follows that Z is the result of the nth step of the iteration

$$Y_{i+1} = \mathbb{1}_{\{\theta_i=i\}}Y_i + \mathbb{1}_{\{\theta_i=i+1\}}X_{i+1},$$

for $i = 1 \ldots n$, with $Y_1 = X_1$, i.e. $Z = Y_n$. It can be easily argued that $Y_i \in A_{i-1}$, $i = 2 \ldots n$ and therefore $Z \in A_{n-1}$ or $Z \in co(A)$ which implies that $\mathcal{K}(A) \subset co(A)$.

We now begin the second part of the proof and show that any point in $co(A)$ belongs to $\mathcal{K}(A)$ as well. If $Z \in co(A)$, from Section 3.2.2 we have that there exits a positive integer m such that $Z \in A_m$ and therefore Z is the root of a binary tree of height m with leaves from the set A. Using the same notations as in Section 3.2.2 for each of the leaf nodes X_i, there exists $n_i \geq 1$ paths from Z to X_i, of lengths m_{i_l}, $l = 1 \ldots n_i$ which are denoted by

$$P_{Z,X_i} \triangleq \left\{ \left(\{Y_{i_l,j}\}_{j=0}^{m_{i_l}}, \{\lambda_{i_l,j}\}_{j=1}^{m_{i_l}} \right) \mid l = 1 \ldots n_i \right\},$$

where $Y_{i_l,j-1} = \psi\left(Y_{i_l,j}, *, \lambda_{i_l,j}\right)$ for $j = 1 \ldots m_{i_l}$, $l = 1 \ldots n_i$ and where we denoted by $*$ some intermediate node in the tree. We introduce the independent, random variables $\theta_{i_l,j}$ such that $Pr(\theta_{i_l,j} = i_l, j) = \lambda_{i_l,j}$ and $Pr(\theta_{i_l,j} = *) = 1 - \lambda_{i_l,j}$. It follows that Z can be expressed as

$$Z = \sum_{i=1}^{n} \left(\sum_{l=1}^{n_i} \prod_{j=1}^{m_{i_l}} \mathbb{1}_{\{\omega:\theta_{i_l,j}=i_l,j\}} \right) X_i$$

Using again the independence of $\theta_{i_l,j}$ we have that

$$\sum_{l=1}^{n_i} \prod_{j=1}^{m_{i_l}} \mathbb{1}_{\{\omega:\theta_{i_l,j}=i_l,j\}} = \mathbb{1}_{\{\bigcup_{l=1}^{n_i} \bigcap_{j=1}^{m_{i_l}} \{\omega:\theta_{i_l,j}=i_l,j\}\}}$$

Let $S_i \triangleq \left\{ \bigcup_{l=1}^{n_i} \bigcap_{j=1}^{m_{i_l}} \{\omega : \theta_{i_l,j} = i_l, j\} \right\}$ and let us interpret the events in S_i as the set of underlying *sub − events* generating ω_i i.e. $\omega_i = S_i$. It is not difficult to see that

$$Pr(\omega_i) = Pr(S_i) = \mathcal{W}(P_{Z,X_i}).$$

By defining $w_i \triangleq \mathcal{W}(P_{Z,X_i})$ we get that $\sum_{i=1}^{n} Pr(\omega_i) = 1$. Note that if there exits an i^* such that X_{i^*} is not among the leaves of the binary tree rooted at Z, the measure of the event ω_i is zero. Therefore we have that Z can be expressed as

$$Z = \sum_{i=1}^{n} \mathbb{1}_{\{\omega_i\}} X_i = \sum_{i=1}^{n} \mathbb{1}_{\{\theta=i\}} X_i,$$

84

where $Pr(\theta = i) = w_i$ and hence it follows that $Z \in \mathcal{K}(A)$ and consequently $co(A) \subset \mathcal{K}(A)$.

From part one and part two of our proof, the result follows.

$\square$

Remark 3.6.1. *We say that Z is between X_1 and X_2 if $d(X_1, Z) + d(Z, X_2) = d(X_1, X_2)$. For any two points $X_1, X_2 \in \mathcal{X}$, the set*

$$\{Z \in \mathcal{X} \mid d(X_1, Z) + d(Z, X_2) = d(X_1, X_2)\},$$

is called metric segment *and is denoted by $[X_1, X_2]$. We note that any point $Z \in \mathcal{X}$ belonging to the convex hull of X_1, X_2 is on the metric segment between X_1 and X_2. Indeed, if $Z \in co(\{X_1, X_2\})$ then $Z = \mathbb{1}_{\{\theta=1\}}X_1 + \mathbb{1}_{\{\theta=2\}}X_2$, where $Pr(\theta = 1) = \lambda$ and $Pr(\theta = 2) = 1 - \lambda$, for some $\lambda \in [0, 1]$. It follows that*

$$d(X_1, Z) + d(Z, X_2) = E[\rho(X_1, Z) + \rho(Z, X_2)] = E[E[\rho(X_1, Z) + \rho(Z, X_2)]|X_1, X_2] =$$

$$= E[\lambda\rho(X_1, X_2) + (1 - \lambda)\rho(X_1, X_2)] = d(X_1, X_2).$$

However, not every point belonging to the metric segment $[X_1, X_2]$ belongs to $co(\{X_1, X_2\})$. Indeed, assume for example that $X_1, X_2 \in \{1, 2\}$ and consider a random variable $Z \in \{1, 2\}$ whose probability mass function, conditioned on the values of X_1 and X_2 is given by $Pr(Z = 2|X_1 = 2, X_2 = 1) = \lambda$, $Pr(Z = 1|X_1 = 2, X_2 = 1) = 1 - \lambda$, $Pr(Z = 1|X_1 = 1, X_2 = 2) = \tilde{\lambda}$, $Pr(Z = 1|X_1 = 1, X_2 = 2) = 1 - \tilde{\lambda}$ and $Pr(Z = 2|X_1 = 2, X_2 = 2) = Pr(Z = 1|X_1 = 1, X_2 = 1) = 1$, for some $\lambda \neq \tilde{\lambda} \in (0, 1)$. Since $Pr(Z = 2|X_1 = 2, X_2 = 1) \neq Pr(Z = 1|X_1 = 1, X_2 = 2)$ it follows that $Z \notin co(\{X_1, X_2\})$. However it can be easily checked that $Z \in [X_1, X_2]$. In fact any random variable Z whose probability mass function conditioned on the values of X_1

and X_2 satisfies

$$\sum_{z \neq x_1 \neq x_2} Pr(Z = z | X_1 = x_1, X_2 = x_2) = 0, \sum_{z \neq x} Pr(Z = z | X_1 = x, X_2 = x) = 0,$$

belongs to the metric segment $[X_1, X_2]$.

Corollary 3.6.1. *Let n be a positive integer and let $A = \{X_1, \ldots, X_n\}$ be a set of points in $\mathcal{X}$. Consider the independent random variable θ taking values in the finite set $\{1, \ldots, n\}$, with probability measure given by $Pr(\omega : \theta(\omega) = i) = w_i$, for some non-negative scalars w_i, with $\sum_{i=1}^{n} w_i = 1$. Then*

$$co_\varepsilon(A) = \left\{ Z \in \mathcal{X} \mid Z = \sum_{i=1}^{n} \mathbb{1}_{\{\theta=i\}} X_i, \ \forall w_i \geq \varepsilon, \sum_{i=1}^{n} w_i = 1 \right\}. \tag{3.48}$$

Proof. Follows immediately from Definition 3.2.5 and Theorem 3.6.1. □

Recall the discussion introduced by Remark 3.2.1 on what we understand by a small enough value of ε.

3.6.2 Consensus of Opinion Algorithm

We assume that each agent of a group of n agents has an *initial opinion*. We model the set of opinions by a finite set of distinct integers, say $S = \{1, 2, \ldots, s\}$ for some positive integer s, where each element of S indicates an opinion. The goal of the agents is to reach the same opinion by repeatedly discussing among themselves.

Denoting as before by k the time-index and by $G(k) = (V, E(k))$ the time varying graph modeling the communication network among the n agents, we model the evolution of the opinion of an agent i as a random process $X_i(k)$, where $X_i(k) \in \mathcal{X}$ for all $k \geq 0$. Each agent i has an initial opinion $X_i(0) = x_{il}^0 \in S$ with probability $p_{il} \geq 0$, with $\sum_{l=1}^{s} p_{il} = 1$.

Corollary 3.6.2. *Let Assumptions 3.3.1 and 3.3.2 hold for $G(k)$. Given a small enough, positive scalar $\varepsilon < 1$, assume that at every time-slot each agent i rolls an imaginary dice with $|\mathcal{N}_i(k)|$ facets numbered from 1 to $|\mathcal{N}_i(k)|$, independently of the other agents. The probability that the result of a dice roll is $j \in \mathcal{N}_i(k)$, is $w_{ij}(k)$ with $w_{ij}(k) \geq \varepsilon$ and $\sum_{j \in \mathcal{N}_i(k)} w_{ij}(k) = 1$. The agent i updates its state according to the following scheme. If the result of the dice roll is j then agent i chooses the opinion of agent j. We then have that the agents asymptotically agree on the same opinion, i.e.*

$$\lim_{k \to \infty} d(X_i(k), X_j(k)) = 0, \forall i, j$$

Proof. By modeling the dice of agent i as an i.i.d. random process $\theta_i(k) \in \{1, 2, \ldots, |\mathcal{N}_i(k)|\}$ such that $Pr(\theta_i(k) = j) = w_{ij}(k)$ for all $j \in \mathcal{N}_i(k)$ and for all $i, k \geq 0$, the update scheme of agent i can be formally written as

$$X_i(k+1) = \sum_{j \in \mathcal{N}_i(k)} \mathbb{1}_{\{\theta_i(k)=j\}} X_j(k). \tag{3.49}$$

However this implies that $X_i(k+1) \in co_\varepsilon(A_i(k))$, $\forall i, k$ and the result follows from Theorem 3.3.1. □

3.6.3 Probabilistic analysis of the consensus algorithm

In this section we give a probabilistic analysis of the consensus of opinion algorithm introduced in the previous section. We discuss about the different modes of convergence to agreement (from a probabilistic point of view) and we give an alternative proof of Corollary 3.6.2 using purely probability theory arguments. In addition, we discuss about the convergence in distribution of the states of the agents to a particular random variable and we redefine the notion of *average consensus* from $\mathbb{R}^n$ to fit the metric space $\mathcal{X}$.

Corollary 3.6.2 shows that under the proposed scheme the distances between the states of the agents converge to zero. However, since $\mathcal{X}$ is a space of discrete random variables, we can say more about the modes of convergence of the states of the agents.

Recall that we defined the distance between two points $X_1, X_2 \in \mathcal{X}$ as

$$d(X_1, X_2) = E[\rho(X_1, X_2)] = Pr(X_1 \neq X_2).$$

From Corollary 3.6.2 we have that

$$\lim_{k \to \infty} d(X_i(k), X_j(k)) = 0,$$

or equivalently

$$\lim_{k \to \infty} Pr(X_i(k) \neq X_j(k)) = 0. \tag{3.50}$$

This says that the measure of the set on which $X_i(k)$ and $X_j(k)$ are different converges to zero as k goes to infinity, i.e. *the agents asymptotically agree in probability sense*. In what follows we show that in fact the agents asymptotically agree with probability one (or in almost sure sense).

Given an arbitrary $\epsilon > 0$, we define the event

$$B_k(\epsilon) \triangleq \{\omega : \max_{i \neq j} |X_i(k) - X_j(k)| > \epsilon\}.$$

An upper bound on the probability of the event $B_k(\epsilon)$ is given by

$$
\begin{aligned}
Pr(B_k(\epsilon)) = Pr\left(\bigcup_{i \neq j} \left\{\omega : |X_i(k) - X_j(k)| > \epsilon\right\}\right) \leq \\
\leq \sum_{i \neq j} Pr\left(|X_i(k) - X_j(k)| > \epsilon\right) \leq \sum_{i \neq j} Pr\left(X_i(k) \neq X_j(k)\right).
\end{aligned}
\tag{3.51}
$$

From (3.50) and (3.51) we obtain

$$\lim_{k \to \infty} Pr(B_k(\epsilon)) = 0.$$

88

Recall that by inequality (3.38), $d(X_i(k), X_j(k)) = Pr(X_i(k) \neq X_j(k))$, $\forall i, j$ converge at least geometrically to zero. Therefore

$$\sum_{k \geq 0} Pr(B_k(\epsilon)) < \infty,$$

and by the Borel-Cantelli lemma we have that

$$Pr(B_k(\epsilon) \text{ happens infinetely often}) = 0.$$

Equivalently, this also means that

$$Pr\left(\lim_{k \to \infty} \max_{i \neq j} |X_i(k) - X_j(k)| = 0 \right) = 1,$$

or that the agents asymptotically agree with probability one.

In the following we show that the same result can be obtained by using purely probability theory arguments. For simplicity we assume that the communication network remains constant and connected and that the coefficients w_{ij} from the agreement scheme are constant as well.

Proposition 3.6.2. *Let the graph modeling the communication network be time invariant and connected and let the agents update their state according to the scheme described in Corollary 3.6.2, where $w_{ij} > 0$ are assumed constant for all $k \geq 0$. We then have that the agents asymptotically agree with probability one, i.e.*

$$Pr\left(\lim_{k \to \infty} \max_{i \neq j} |X_i(k) - X_j(k)| = 0 \right) = 1. \tag{3.52}$$

Proof. We define the random process $Z(k) = (X_1(k), X_2(k), \ldots, X_n(k))$ which has a maximum of s^s states and we introduce the *agreement space* as

$$\mathcal{A} \triangleq \{(o, o, \ldots, o) \mid o \in S\}.$$

89

We saw earlier that the state update dynamics is given by

$$X_i(k+1) = \sum_{j \in \mathcal{N}_i} \mathbb{1}_{\{\theta_i(k)=j\}} X_j(k),$$

where $Pr(\theta_i(k) = j) = w_{ij}$, for all $j \in \mathcal{N}_i$ and for all i. The conditional probability of $X_i(k+1)$ conditioned on $X_j(k)$, $j \in \mathcal{N}_i$ is given by

$$Pr(X_i(k+1) = o_i | X_j(k) = o_j, j \in \mathcal{N}_i) = \sum_{j \in \mathcal{N}_i} w_{ij} \mathbb{1}_{\{o_i=o_j\}}. \tag{3.53}$$

It is not difficult to note that $Z(k)$ is a finite state, homogeneous Markov chain. We will show that $Z(k)$ has s absorbing states and all other $s^s - s$ states are transient, where the absorbing states correspond to the states in agreement space $\mathcal{A}$. Using the independence of the random processes $\theta_i(k)$, the entries of the probability transition matrix of $Z(k)$ can be derived from (3.53) and are given by

$$Pr(X_1(k+1) = o_{l_1}, \ldots, X_n(k+1) = o_{l_n} | X_1(k) = o_{p_1}, \ldots, X_n(k) = o_{p_n}) = \tag{3.54}$$

$$= \prod_{i=1}^{n} \sum_{j \in \mathcal{N}_i} w_{ij} \mathbb{1}_{\{l_i=p_j\}}.$$

We note from (3.54) that once the process reaches an agreement state it will stay there indefinitely, i.e.

$$Pr(X_1(k+1) = o, \ldots, X_n(k+1) = o | X_1(k) = o, \ldots, X_n(k) = o) = 1, \ \forall o \in S,$$

and hence the agreement states are absorbing states. We will show next that, under the connectivity assumption, the agreement space $\mathcal{A}$ is reachable from any state, and therefore all other states are transient. We are not saying that all agreement states are reachable from any state, but that from any state at least one agreement state is reachable. Let $(o_1, o_2, \ldots, o_n) \notin \mathcal{A}$, with $o_j \in S$, $j = 1 \ldots n$ be an arbitrary state. We first note that from

90

this state only agreement states of the form $(o_j, o_j, \ldots, o_j)$ can be reached. Given that $X_j(0) = o_j$, we show that with positive probability the agreement vector $(o_j, o_j, \ldots, o_j)$ can be reached. At time slot one, with probability w_{jj} agent j keeps its initial choice, while its neighbors to which it sends information can choose o_j with some positive probability, i.e. $X_i(1) = o_j$ with probability w_{ij}, for all i such that $j \in \mathcal{N}_i$. Due to the connectivity assumption there exits at least one i such that $j \in \mathcal{N}_i$. At the next time-index all the agents which have already chosen o_j keep their opinion with positive probability, while their neighbors will choose o_j with positive probability. Since the communication network is assumed connected, every agent will be able to choose o_j with positive probability in at most $n - 1$ steps, therefore an agreement state can be reached with positive probability. Hence, from any initial state $(o_1, o_2, \ldots, o_n) \notin \mathcal{A}$, all agreement states of the form $(o_j, o_j, \ldots, o_j)$ with $j = 1 \ldots n$ are reachable with positive probability. Since the agreement states are absorbing states, it follows that $(o_1, o_2, \ldots, o_n) \notin \mathcal{A}$ is a transient state. Therefore, the probability for the Markov chain $Z(k)$ to be in a transient state converges asymptotically to zero, while the probability to be in one of the agreement states converges asymptotically to one, i.e.

$$\lim_{k \to \infty} Pr(Z(k) \notin \mathcal{A}) = 0,$$

or equivalently

$$\lim_{k \to \infty} Pr\left(\bigcup_{i \neq j} \left\{ X_i(k) \neq X_j(k) \right\} \right) = 0. \tag{3.55}$$

Given an arbitrary $\epsilon > 0$, we define the event

$$B_k(\epsilon) \triangleq \left\{ \omega : \max_{i \neq j} |X_i(k) - X_j(k)| > \epsilon \right\}.$$

But since

$$B_k(\epsilon) = \bigcup_{i \neq j} \left\{ |X_i(k) - X_j(k)| > \epsilon \right\} \subseteq \bigcup_{i \neq j} \left\{ X_i(k) \neq X_j(k) \right\},$$

from (3.55) it follows that

$$\lim_{k \to \infty} Pr(B_k(\epsilon)) \leq \lim_{k \to \infty} Pr\left(\bigcup_{i \neq j} \left\{ X_i(k) \neq X_j(k) \right\} \right) = 0,$$

and hence the agents asymptotically agree in probability sense. In addition, due to the geometric decay toward zero of the probability $Pr(Z(k) \notin \mathcal{A})$, by the Borel-Cantelli Lemma the result follows. □

We discussed above about the different modes of convergence of the agents to the same opinion, but we said nothing about where the states actually converge. However, from Corollary 3.3.1 we know that there exits a random variable $X^* \in \mathcal{X}$ such that

$$\lim_{k \to \infty} d(X_i(k), X^*) = 0, \forall i,$$

or equivalently

$$\lim_{k \to \infty} Pr(X_i(k) \neq X^*) = 0, \forall i,$$

which implies that the states of the agents $X_i(k)$ converge to X^* in probability. Still, this tells us nothing about the properties of X^*. In what follows we analyze the evolution of the probability with which an agent i chooses between the initial values (opinions) of the other agents in the network. Also, we focus on the convergence in distribution to X^* and more precisely we characterize the distribution of X^*.

By defining the vector $Z(k) \triangleq (X_1(k), X_2(k), \ldots, X_n(k))'$, (3.49) can be compactly written as

$$Z(k+1) = \Theta(k)Z(k), \ Z(0) = Z_0, \tag{3.56}$$

92

where $[\Theta(k)]_{ij} = \mathbb{1}_{\{\theta_i(k)=j\}}$ and where $\theta_i(k)$ are independent random processes with $Pr(\theta_i(k) = j) = w_{ij}(k)$, $w_{ij}(k) \geq \varepsilon$ and $\sum_{j \in \mathcal{N}_i(k)} w_{ij}(k) = 1$. Consequently

$$Z(k) = \Gamma(k)Z(0),$$

where $\Gamma(k) = \Theta(k-1)\Theta(k-2)\cdots\Theta(1)\Theta(0)$ is the transition matrix of (3.56). It can be easily argued that the (i, j) entry of $\Gamma(k)$ can be expressed as

$$[\Gamma(k)]_{ij} = \mathbb{1}_{\{\bar{\theta}_i(k)=j\}}, \tag{3.57}$$

where $\bar{\theta}_i(k)$ are random processes taking values in the discrete set $\{1, 2, \ldots, n\}$. The quantity $\mathbb{1}_{\{\bar{\theta}_i(k)=j\}}$ is updated according to the expression

$$\mathbb{1}_{\{\bar{\theta}_i(k+1)=j\}} = \sum_{l=1}^{n} \mathbb{1}_{\{\theta_i(k)=l\}}\mathbb{1}_{\{\bar{\theta}_l(k)=j\}} = \sum_{l=1}^{n} \mathbb{1}_{\{\theta_i(k)=l,\bar{\theta}_l(k)=j\}}, \tag{3.58}$$

where the second inequality followed from the independence of $\theta_i(k)$ and with $\mathbb{1}_{\{\bar{\theta}_i(0)=j\}} = \mathbb{1}_{\{\theta_i(0)=j\}}$ for all i, j pairs. Since the events $\{\omega : \theta_i(k) = l, \bar{\theta}_l(k) = j\}$ for $l = 1\ldots n$ are mutually exclusive, $\mathbb{1}_{\{\bar{\theta}_i(k+1)=j\}}$ is indeed well defined. The probability mass function of $\bar{\theta}_i(k)$ is given by

$$Pr(\bar{\theta}_i(k) = j) = [W(k)W(k-1)\cdots W(1)W(0)]_{ij},$$

where $[W(k)]_{ij} = w_{ij}(k)$.

It is not difficult to observe that the entries of $\Gamma(k)$ act as selectors between the different entries of the initial vector $Z(0)$, i.e.

$$X_i(k) = \sum_{j=1}^{n} \mathbb{1}_{\{\bar{\theta}_i(k)=j\}} X_j(0).$$

Therefore, the probability for $X_i(k)$ to choose $X_j(0)$ is given by the probability of $\bar{\theta}_i(k)$ to choose j, i.e.

$$Pr(X_i(k) = X_j(0)) = Pr(\bar{\theta}_i(k) = j).$$

93

Under Assumptions 3.3.1 and 3.3.2, we can invoke Lemmas 3 and 4 of [34], and obtain that there exits a vector v with positive entries summing up to one, such that

$$\lim_{k \to \infty} W(k)W(k-1) \cdots W(1)W(0) = \mathbf{1}v',$$

where $\mathbf{1}$ is the vector of all ones. Therefore, as k goes to infinity the agents will pick among the initial values $X_j(0)$ with probability v_j, i.e.

$$\lim_{k \to \infty} Pr(X_i(k) = X_j(0)) = \lim_{k \to \infty} Pr(\bar{\theta}_i(k) = j) = v_j, \tag{3.59}$$

where v_j is the jth entry of vector v. In particular, if the matrix $W(k)$ is doubly stochastic, then by Proposition 1 of [34], $v = \frac{1}{n}\mathbf{1}$ and consequently

$$\lim_{k \to \infty} Pr(X_i(k) = X_j(0)) = \frac{1}{n}. \tag{3.60}$$

This leads us to redefining the *average consensus* concept from $\mathbb{R}^n$ to our particular convex metric space $\mathcal{X}$, i.e. *we can say that the agents reach average consensus if they asymptotically agree on the different initial opinions with the same probability.*

Remarkably, from (3.60) it also follows that $X_i(k)$ converge in distribution to a random variable X^* given by

$$X^* = \sum_{j=1}^{n} \mathbb{1}_{\{\theta^*=j\}} X_j(0),$$

where $Pr(\theta^* = j) = \frac{1}{n}$. Note that X^* is a point in the convex hull of $\{X_1(0), \ldots, X_n(0)\}$ generated by associating equal weights to the initial values $X_j(0)$. Hence, X^* can be interpreted as the *(empirical) average* of the initial values.

Introducing the vector $\mathbf{p}^l(k) = (p_i^l(k))$, where $p_i^l(k) = Pr(X_i(k) = l)$ for some $l \in S$, from (3.49) and from the independence of the random processes $\theta_i(k)$, we obtain that the

94

evolution of $\mathbf{p}^l(k)$ respects the equation

$$\mathbf{p}^l(k+1) = W(k)\mathbf{p}^l(k), \ \mathbf{p}^l(0) = \mathbf{p}_0^l, \qquad (3.61)$$

where $[W(k)]_{ij} = w_{ij}(k)$. Hence we obtain that there exits a vector v with positive entries summing up to one, such that

$$\lim_{k \to \infty} W(k)W(k-1) \cdots W(1)W(0) = \mathbf{1}v'.$$

Therefore, by defining $\pi_l \triangleq \sum_{j=1}^n v_j Pr(X_j(0) = l)$, where v_j is the jth entry of v, we have that

$$\lim_{k \to \infty} Pr(X_i(k) = l) = \pi_l, \ \forall i,$$

or equivalently that $X_i(k)$ converge is distribution to a random variable X^* whose probability mass function is given by $Pr(X^* = l) = \pi_l$, for all i. If in addition we have that $W(k)$ is doubly stochastic, we have that

$$\lim_{k \to \infty} Pr(X_i(k) = l) = \frac{1}{n} \sum_{j=1}^n Pr(X_j(0) = l).$$

3.6.4 Numerical example

In what follows we consider an example where a group of eight agents ($n = 8$) have to choose between two opinions, i.e. $S = \{1, 2\}$. We assume that the agents communication network is given by an undirected circular graph as in Figure 3.2, assumed fixed for all time-slots.

We assume that the agents use the scheme described by Corollary (3.6.2) for updating their states, i.e. the coefficients w_{ij} are constant. In particular we choose $w_{ii} = 7/9$ and $w_{i,i-1} = w_{i,i+1} = 1/9$ and choose as initial values $X_i(0) = 1$ for $i = 1 \ldots 4$ and $X_i(0) = 2$
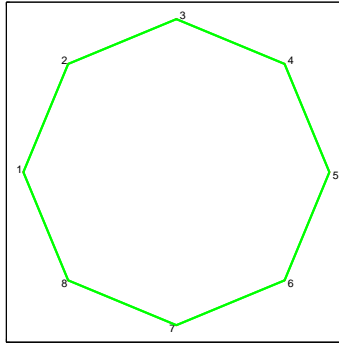
Figure 3.2: Undirected circular graph with eight nodes

for $i = 5 \dots 8$ with probability one. Figure 3.3 presents an execution of our agreement algorithm which indeed shows that the agents agree on the same opinion. The different colors that appear indicates different agents.
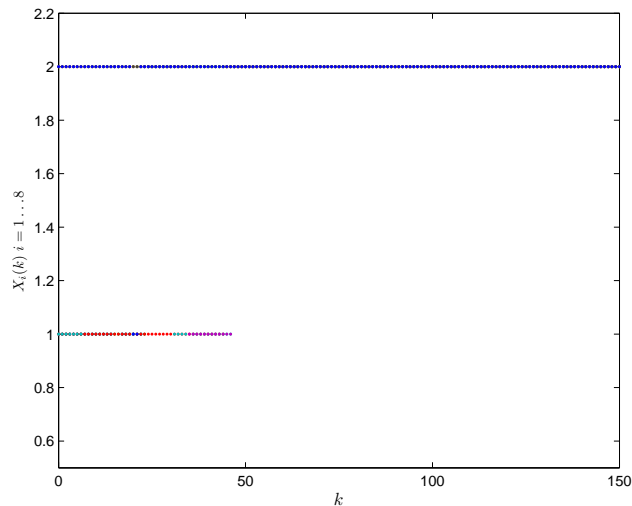
Figure 3.3: Execution of the agreement algorithm

Next we numerically analyze the evolution of the vector of distances $\mathbf{d}(k) = (d(X_i(k), X_j(k)))$, $\forall\, i \neq j$. First we see that under our assumption the entries of matrix $[\mathbf{W}(k)]_{\bar{i}, \bar{j}} = w_{ip} w_{jq}$, where $\bar{i}$ and $\bar{j}$ correspond to the pairs of agents (i, j) and (p, q), respectively, and where

96

w_{ij} define the probability mass function of the random variables $\theta_i(k)$ as described in Corollary 3.6.2. We consider the linear system

$$\tilde{\mathbf{d}}(k+1) = \mathbf{W}(k)\tilde{\mathbf{d}}(k), \ \tilde{\mathbf{d}}(0) = \mathbf{d}(0).$$

By (3.20) of Lemma 3.4.1, we have that $\tilde{\mathbf{d}}(k)$ is an upper bound of $\mathbf{d}(k)$. Figure 3.4 presents the evolution of $\|\tilde{\mathbf{d}}(k)\|_\infty$ with time. It is worth mentioning that since ψ defined in (3.43) satisfies the definition of a convex structure with equality, it can be easily argued that (3.20) holds with equality and therefore the upper bound $\tilde{\mathbf{d}}(k)$ is in fact $\mathbf{d}(k)$.
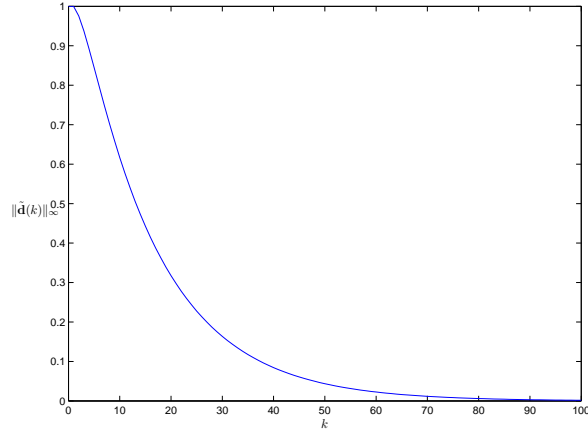
Figure 3.4: Evolution of $\|\tilde{\mathbf{d}}(k)\|_\infty$ with time

We next analyze the distance between the initial points and the consensus point(s). Since ψ respects the definition of a convex structure with equality, we have that

$$d(X_i(k+1), X_l(0)) = \sum_{j \in \mathcal{N}_i} w_{ij} d(X_j(k), X_l(0)),$$

which is basically a consensus algorithm. Since the consensus matrix is doubly stochastic we know that

$$\lim_{k \to \infty} d(X_i(k), X_l(0)) = \frac{1}{n} \sum_{j=1}^{n} d(X_j(0), X_l(0))$$

Figure 3.5 presents the evolution of the distance between $X_i(k)$ and $X_1(0)$ for $i = 1 \ldots n$.

Considering our choice for initial values and the fact that $n = 8$ it is not difficult to see that
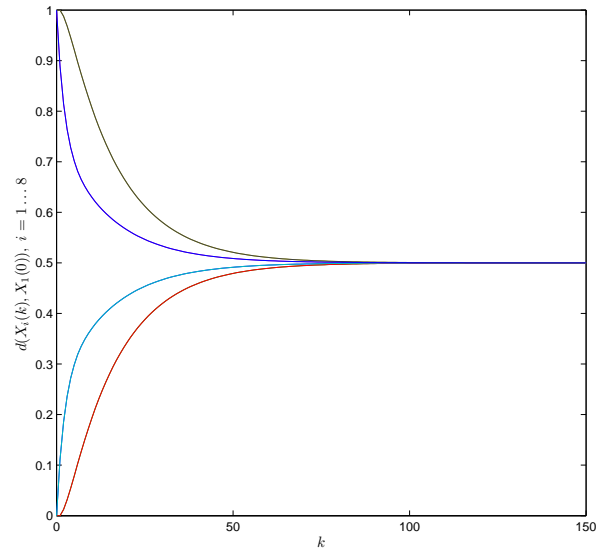
Figure 3.5: The distances between $X_i(k)$ and $X_1(0)$ for $i = 1 \ldots 8$

$$\frac{1}{n} \sum_{j=1}^{n} d(X_j(0), X_l(0)) = \frac{1}{2}.$$

which is also what Figure 3.5 shows.

Chapter 4

Distributed Asymptotic Agreement Problem under Markovian Random Topologies

4.1 Introduction

This chapter deals with the linear consensus problem for a group of dynamic agents. We assume that the communication flow between agents is modeled by a (possibly directed) randomly switching graph. The switching is determined by a homogeneous, finite-state Markov chain, each communication pattern corresponding to a state of the Markov process. We address both the cases where the dynamics of the agents is expressed in continuous and discrete time and, under certain assumptions on the consensus matrices, we give necessary and sufficient conditions to guarantee convergence to average consensus in mean square and in almost sure sense. The Markovian switching model goes beyond the common i.i.d. assumption on the random communication topology and appears in cases where Rayleigh fading channels are considered. One of the goals of this chapter is to show how mathematical techniques used in the stability analysis of Markovian jump linear systems, together with results inspired by matrix and graph theory, can be used to prove (intuitively clear) convergence results for the (linear) stochastic consensus problem.

Basic notations and definitions: We denote by $\mathbb{1}$ the vector of all ones. If the dimension of the vector needs to be emphasized, an index will be added for clarity (for

example, if $\mathbb{1}$ is an n dimensional vector, we will explicitly mark this by using $\mathbb{1}_n$). Let x be a vector in $\mathbb{R}^n$. By $av(x)$ we denote the quantity $av(x) = x'\mathbb{1}/\mathbb{1}'\mathbb{1}$. The symbols $\otimes$ and $\oplus$ represent the Kronecker product and sum, respectively. Given a matrix A, $Null(A)$ designates the nullspace of the considered matrix. If $\mathcal{X}$ is some finite dimensional space, $dim(\mathcal{X})$ gives us the dimension of $\mathcal{X}$. We denote by $col(A)$ a vector containing the columns of matrix A.

Let $\mathcal{M}$ be a set of matrices and let A be some matrix. By $\mathcal{M}'$ we denote the set of the transpose matrices of $\mathcal{M}$, i.e. $\mathcal{M}' = \{M' \mid M \in \mathcal{M}\}$. By $\mathcal{M} \otimes A$ we understand the following matrix set: $\mathcal{M} \otimes A = \{M \otimes A \mid M \in \mathcal{M}\}$. By writing that $A\mathcal{M} = \mathcal{M}$ we understand that $AM \in \mathcal{M}$, for any $M \in \mathcal{M}$.

Let P be a probability transition matrix corresponding to a homogeneous, finite state, Markov chain. We denote by $\mathcal{P}_\infty$ the limit set of the sequence $\{P^k\}_{k \geq 0}$, i.e. all matrices L for which there exists a sequence $\{t_k\}_{k \geq 0}$ in $\mathbb{N}$ such that $\lim_{k \to \infty} P^{t_k} = L$. Note that if the matrix P corresponds to an ergodic Markov chain, the cardinality of $\mathcal{P}_\infty$ is one, with the limit point $\mathbb{1}\pi'$, where π is the stationary distribution. If the Markov chain is periodic with period m, the cardinality of $\mathcal{P}_\infty$ is m. Let $d(M, \mathcal{P}_\infty)$ denote the distance from M to the set $\mathcal{P}_\infty$, that is the smallest distance from M to any matrix in $\mathcal{P}_\infty$:

$$d(M, \mathcal{P}_\infty) = \inf_{L \in \mathcal{P}_\infty} \|L - M\|,$$

where $\|\cdot\|$ is a matrix norm.

Definition 4.1.1. *Let A be a matrix in $\mathbb{R}^{n \times n}$ and let $G = (V, E)$ be a graph of order n. We say that matrix A* corresponds *to graph G or that graph G* corresponds *to matrix A if an edge e_{ij} belongs to E if and only if the (i, j) entry of A is non-zero. The graph*

100

corresponding to A will be denoted by G_A.

Definition 4.1.2. *Let s be a positive integer and let* $\mathcal{A} = \{A_i\}_{i=1}^s$ *be a set of matrices with a corresponding set of graphs* $\mathcal{G} = \{G_{A_i}\}_{i=1}^s$. *We say that the graph* $G_\mathcal{A}$ *corresponds to the set* $\mathcal{A}$ *if it is given by the union of graphs in* $\mathcal{G}$, *i.e.*

$$G_\mathcal{A} \triangleq \bigcup_{i=1}^s G_{A_i}.$$

In this note we will use mainly two type of matrices: *probability transition matrices* (row sum up to one) and *generator* matrices (row sum up to zero). A generator matrix whose both rows and columns sum up to zero will be called *doubly stochastic generator matrix*.

To simplify the exposition we will sometimes characterize a probability transition/generator matrix as being irreducible or strongly connected and by this we understand that the corresponding Markov chain (directed graph) is irreducible (strongly connected).

Definition 4.1.3. *Let* $A \in \mathbb{R}^{n \times n}$ *be a probability transition/generator matrix. We say that A is* block diagonalizable *if there exists a similarity transformation P, encapsulating a number of row permutations, such that* PAP' *is a block diagonal matrix with irreducible blocks on the main diagonal.*

For simplicity, the time index for both the continuous and discrete-time cases is denoted by t.

Chapter organization: In Section 4.2 we present the setup and formulation of the problem and we state our main convergence theorem. In Section 4.3 we derive a number of results which constitute the core of the proof of our main result; proof which is given in Section 4.4. Section 4.5 contains a discussion of our convergence result.

101

4.2 Problem formulation and statement of the convergence result

We assume that a group of n agents, labeled 1 through n, is organized in a communication network whose topology is given by a time varying graph $\mathbf{G}(t) = (V, E(t))$, where V is the set of n vertices and $E(t)$ is the time varying set of edges. The graph $\mathbf{G}(t)$ has an underlying random process governing its evolution, given by a homogeneous, continuous or discrete time Markov chain $\theta(t)$, taking values in the finite set $\{1, \ldots, s\}$, for some positive integer s. In the case $\theta(t)$ is a discrete-time Markov chain, its probability transition matrix is $P = (p_{ij})$ (rows sum up to one), while if $\theta(t)$ is a continuous time Markov chain, its generator matrix is denoted by $\Lambda = (\lambda_{ij})$ (rows sum up to zero). The random graph $\mathbf{G}(t)$ takes values in a finite set of graphs $\mathcal{G} = \{G_i\}_{i=1}^s$ with probability $Pr(\mathbf{G}(t) = G_i) = Pr(\theta(t) = i)$, for $i = 1 \ldots s$. We denote by $q = (q_i)$ the initial distribution of $\theta(t)$.

Letting $x(t)$ denote the state of the n agents, in the case $\theta(t)$ is a discrete-time Markov chain, we model the dynamics of the agents by the following linear stochastic difference equation

$$x(t+1) = \mathbf{D}_{\theta(t)}x(t), \ x(0) = x_0, \tag{4.1}$$

where $\mathbf{D}_{\theta(t)}$ is a random matrix taking values in the finite set $\mathcal{D} = \{D_i\}_{i=1}^s$, with probability distribution $Pr(\mathbf{D}_{\theta(t)} = D_i) = Pr(\theta(t) = i)$. The matrices D_i are stochastic matrices (rows sum up to one) with positive diagonal entries and correspond to the graphs G_i, for $i = 1 \ldots s$.

In the case $\theta(t)$ is a continuous-time Markov chain, we model the dynamics of the agents by the following linear stochastic equation

$$dx(t) = \mathbf{C}_{\theta(t)}x(t)dt, \ x(0) = x_0, \tag{4.2}$$

102

where $\mathbf{C}_{\theta(t)}$ is a random matrix taking values in the finite set $C = \{C_i\}_{i=1}^s$, with probability distribution $Pr(\mathbf{C}_{\theta(t)} = C_i) = Pr(\theta(t) = i)$. The matrices C_i are generator like matrices (rows sum up to zero) and correspond to the graphs G_i, for $i = 1\ldots s$. The initial state $x(0) = x_0$, for both continuous and discrete models, is assumed deterministic. We will sometimes refer to the matrices belonging to the sets $\mathcal{D}$ and C as *consensus matrices*. The underlying probability space (for both models) is denoted by $(\Omega, \mathcal{F}, P)$ and the solution process $x(t, x_0, \omega)$ (or simply, $x(t)$) of (4.1) or (4.2) is a random process defined on $(\Omega, \mathcal{F}, P)$. We note that the stochastic dynamics (4.1) and (4.2) represent *Markovian jump linear systems* for discrete and continuous time, respectively. For a comprehensive study of the theory of (discrete-time) Markovian jump linear systems, the reader can refer to [11] for example.

Assumption 4.2.1. *Throughout this chapter we assume that the matrices belonging to the sets $\mathcal{D}$ and C are* doubly stochastic *(rows and columns sum up to one and zero, respectively) and in the case of the set $\mathcal{D}$ have* positive diagonal entries. *We assume also that the Markov chain $\theta(t)$ is* irreducible.

Remark 4.2.1. *Consensus matrices that satisfy Assumption 4.2.1 can be constructed for instance by using a Laplacian based scheme in the case where the communication graph is undirected or balanced (for every node, the inner degree is equal to the outer degree) and possible weighted. If L_i denotes the Laplacian of the graph G_i, we can choose $A_i = I - \epsilon L_i$ and $C_i = -L_i$, where $\epsilon > 0$ is chosen such that A_i is stochastic.*

Definition 4.2.1. *We say that $x(t)$ converges to average consensus*

    I. in the mean square sense, *if for any $x_0 \in \mathbb{R}^n$ and initial distribution $q = (q_1, \ldots, q_s)$*

of $\theta(t)$,

$$\lim_{t\to\infty} E[\|x(t) - av(x_0)\mathbb{1}\|^2] = 0.$$

II. *in the* almost sure sense, *if for any $x_0 \in \mathbb{R}^n$ and initial distribution $q = (q_1, \ldots, q_s)$ of*

$\theta(t)$,

$$Pr(\lim_{t\to\infty} \|x(t) - av(x_0)\mathbb{1}\|) = 1.$$

Assumption 4.2.1 will guarantee reaching average consensus, *desirable in impor-*

tant distributed computing applications such as distributed estimation [40] or distributed

optimization [34]. Any other scheme can be used as long as it produces matrices with the

properties stated above and it reflects the communication structures among agents.

Problem 4.2.1. *Given the random processes $D(t)$ and $C(t)$, together with Assumption*

4.2.1, we derive necessary and sufficient conditions such that the state vector $x(t)$, evolv-

ing according to (4.1) or (4.2), converges to average consensus in the sense of Definition

4.2.1.

In the following we state the convergence result for the linear consensus problem

under Markovian random communication topology.

Theorem 4.2.1. *The state vector $x(t)$, evolving according to the dynamics (4.1) (or (4.2))*

converges to average consensus in the sense of Definition 4.2.1, if and only if $G_{\mathcal{D}}$ (or G_C)

is strongly connected.

The above theorem formulates an intuitively obvious condition for reaching con-

sensus under the linear scheme (4.1) or (4.2) and under the Markovian assumption on

104

the communication patterns. Namely, it expresses the need for persistent communication paths among all agents. We defer for Section IV the proof of this theorem and provide here an intuitive and non-rigorous interpretation. Since $\theta(t)$ is irreducible, with probability one all states are visited infinitely many times. But since the graph $G_{\mathcal{D}}$ (or G_C) is strongly connected, communication paths between all agents are formed infinitely many times, which allows for consensus to be achieved. Conversely, if the graph $G_{\mathcal{D}}$ (or G_C) is not strongly connected, then there exists at least two agents, such that for any sample path of $\theta(t)$, no communication path among them (direct or indirect) is ever formed. Consequently, consensus can not be reached. Our main contribution is to prove Theorem 4.2.1 using an approach based on the stability theory of Markovian jump linear systems, in conjunction with a set of results based on matrix and graph theory.

4.3   Preliminary results

This section starts with a set of general preliminary results after which it continues with results characteristic to the cases where the dynamics of the agents is expressed in discrete and continuous time. The proof of Theorem 4.2.1 is mainly based on four lemmas (Lemmas 4.3.4 and 4.3.5 for discrete-time case and Lemmas 4.3.6 and 4.3.7 for continuous-time case) which state properties of some matrices that appear in the dynamic equations of the first and second moment of the state vector. The proof of these lemmas are based on results introduced in the next subsection.

4.3.1 General preliminary results

This subsection contains the statement of a number of preliminary results that are needed in the proofs of the auxiliary results corresponding to the discrete and continuous time cases and in the proof of the main theorem.

The next theorem introduces a convergence result for an infinite product of ergodic matrices whose proof can be found in [54].

Theorem 4.3.1. *([54]) Let s be a positive integer and let $\{A_i\}_{i=1}^{s}$ be a finite set of $n \times n$ ergodic matrices. Consider a map $r : \mathbb{N} \to \{1, \ldots, s\}$ such that for any finite sequence $\{r(i)\}_{i=1}^{j}$, the matrix product $\prod_{i=1}^{j} A_{r(i)}$ is ergodic. Then, there exists a vector c with non-negative entries (summing up to one), such that:*

$$\lim_{j \to \infty} \prod_{i=1}^{j} A_{r(i)} = \mathbb{1}c'. \tag{4.3}$$

In the case where the matrices $\{A_i\}_{i=1}^{s}$ are doubly stochastic as well, from the above theorem we can immediately obtain the following corollary.

Corollary 4.3.1. *Under the same assumptions as in* Theorem 4.3.1, *if in addition the matrices in the set $\{A_i\}_{i=1}^{s}$ are doubly stochastic, then*

$$\lim_{j \to \infty} \prod_{i=1}^{j} A_{r(i)} = \frac{1}{n}\mathbb{1}\mathbb{1}'. \tag{4.4}$$

Proof. By *Theorem 4.3.1* we have that

$$\lim_{j \to \infty} \prod_{i=1}^{j} A_{r(i)} = \mathbb{1}c'.$$

Since the matrices considered are doubly stochastic and ergodic their transposes are ergodic as well. Hence, by applying again *Theorem 4.3.1* on the transpose versions of

$\{A_i\}_{i=1}^{s}$, we obtain that there exist a vector d such that

$$\lim_{j \to \infty} \left(\prod_{i=1}^{j} A_{r(i)} \right)' = \mathbb{1} d'.$$

But since the stochastic matrix $\mathbb{1} c'$ must be equal to $d\mathbb{1}'$, the result follows. $\qquad \square$

Remark 4.3.1. *The homogeneous finite state Markov chain corresponding to a doubly stochastic transition matrix P can not have transient states. Indeed, since P is doubly stochastic, the same is true for P^t, for all $t \geq 1$. Assuming that there exist a transient state i, then $\lim_{t \to \infty}(P^t)_{ji} = 0$ for all j, i.e. all entries on column i converge to zero. But this means that there exist some t^* for which $\sum_j (P^{t^*})_{ji} < 1$ which contradicts the fact that P^{t^*} must be doubly stochastic. An important implication is that we can relabel the vertices of the Markov chain such that P is block diagonalizable.*

Remark 4.3.2. *Since the Markov chain corresponding to a doubly stochastic transition/generator matrix can not have transient states, the Markov chain (seen as a graph) has a spanning tree if and only if is irreducible (strongly connected).*

The next lemma gives an upper bound on a finite product of nonnegative matrices in terms of the sum of matrices that appear in the product. The proof of this result can be found in [18].

Lemma 4.3.1. *[18] Let $m \geq 2$ be a positive integer and let $\{A_i\}_{i=1}^{m}$ be a set of nonnegative $n \times n$ matrices with positive diagonal elements, then*

$$\prod_{i=1}^{m} A_i \geq \gamma \sum_{i=1}^{m} A_i,$$

where $\gamma > 0$ depends on the matrices A_i, $i = 1, \ldots, m$.

In the following proposition we study the convergence properties of a particular sequence of matrices.

Proposition 4.3.1. *Consider a matrix $Q \in \mathbb{R}^{n \times n}$ such that $\|Q\|_1 \leq 1$ and a set of matrices $\mathcal{S} = \{S_1, \ldots S_m\}$, for some positive integer $m \leq n$. Assume that there exist a subsequence $\{t_k\} \subset \mathbb{N}$ such that $\mathcal{S}$ is a limit set of the sequence $\{Q^{t_k}\}_{k \geq 0}$ and that for any $S \in \mathcal{S}$, $QS \in \mathcal{S}$, as well. Then, $\mathcal{S}$ is a limit set of the sequence $\{Q^k\}_{k \geq 0}$, i.e.*

$$\lim_{k \to \infty} d(Q^k, \mathcal{S}) = 0, \tag{4.5}$$

where $d(Q, \mathcal{S}) = \min_{S \in \mathcal{S}} \|Q - S\|$ and $\|\cdot\|$ is some arbitrary matrix norm.

Proof. Will will prove (4.5) for the particular case of matrix norm one and the general result will follow from the equivalence of norms. Pick a subsequence $\{t'_k\}_{k \geq 0}$ given by $t'_k = t_k + \delta_k$, where $\delta_k \in \mathbb{N}$. It follows that

$$d(Q^{t'_k}, \mathcal{S}) = \min_{S \in \mathcal{S}} \|Q^{\delta_k} Q^{t_k} - Q^{\delta^k} S\|_1 \leq \|Q^{\delta_k}\|_1 \min_{S \in \mathcal{S}} \|Q^{t_k} - S\|_1 \leq d(Q^{t_k}, \mathcal{S}).$$

Therefore, we get that $\mathcal{S}$ is a limit set for the sequence $\{Q^{t'_k}_{k \geq 0}\}$ and the result follows since we can make $\{t'_k\}_{k \geq 0}$ arbitrary. $\square$

The next lemma states a property of the null spaces of two generator matrices.

Lemma 4.3.2. *Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times n}$ be two block diagonalizable generator matrices. Then*

$$Null(A + B) = Null(A) \cap Null(B).$$

Proof. Obviously, $Null(A) \cap Null(B) \subset Null(A + B)$. In the following we show the opposite inclusion. Since A is block diagonalizable, then there exists a similarity transformation

T such that $\bar{A} = TAT'$ is a block diagonal generator matrix (with irreducible blocks). Let $\bar{A}_i \in \mathbb{R}^{n_i \times n_i}$, $i = 1 \ldots m$ denote the irreducible blocks on the main diagonal of $\bar{A}$, where m is the number of such blocks and $\sum_{i=1}^{m} n_i = n$. The nullspace of $\bar{A}$ can be expressed as

$$Null(\bar{A}) = \left\{ \begin{pmatrix} \alpha_1 \mathbb{1}_{n_1} \\ \vdots \\ \alpha_m \mathbb{1}_{n_m} \end{pmatrix} \mid \alpha_l \in \mathbb{R}, l = 1 \ldots m \right\}.$$

We assumed that B is block diagonalizable, which means that G_B is a union of isolated, strongly connected subgraphs, property which remains valid for the graph corresponding to $\bar{B} = TBT'$, since $G_{\bar{B}}$ is just a relabeled version of G_B. By adding $\bar{B}$ to $\bar{A}$ two phenomena can happen: we can either leave the graph $G_{\bar{A}}$ unchanged or we can create new connections among the vertices of $G_{\bar{A}}$. In the first case, $G_{\bar{B}} \subset G_{\bar{A}}$ and therefore $Null(\bar{A} + \bar{B}) = Null(\bar{A})$. In the second case we create new connections among the blocks of $\bar{A}$. But since all the subgraphs of $\bar{B}$ are strongly connected this means that if $\bar{A}_i$ becomes connected to $\bar{A}_j$, then necessarily $\bar{A}_j$ becomes connected to $\bar{A}_i$, hence $\bar{A}_i$ and $\bar{A}_j$ form an irreducible (strongly connected) new block, whose nullspace is spanned by the vectors of all ones. Assuming that these are the only new connections that are added to $G_{\bar{A}}$, the nullspace of $\bar{A} + \bar{B}$ will have a similar expression to the nullspace of $\bar{A}$ with the main difference that the coefficients α_i and α_j will be equal. Therefore, in this particular case, the nullspace of $\bar{A} + \bar{B}$ can be expressed as

$$Null(\bar{A} + \bar{B}) = \left\{ \begin{pmatrix} \alpha_1 \mathbb{1}_{n_1} \\ \vdots \\ \alpha_m \mathbb{1}_{n_m} \end{pmatrix} \mid \alpha_l \in \mathbb{R}, \ \alpha_i = \alpha_j, \ l = 1 \ldots m \right\}.$$

In general all blocks $\bar{A}_i$ which become interconnected after adding $\bar{B}$ will have equal coefficients in the expression of the nullspace of $\bar{A} + \bar{B}$, compared to the nullspace of $\bar{A}$. Therefore, $Null(\bar{A} + \bar{B}) \subset Null(\bar{A})$, which means also that $Null(A + B) \subset Null(A)$. Therefore, if $(A + B)v = 0$, then $Av = 0$ which implies also that $Bv = 0$ or $v \in Null(B)$. Hence if $v \in Null(A + B)$ then $v \in Null(A) \cap Null(B)$, which concludes the proof. $\qquad\square$

In the next corollary we present a property of the eigenspaces corresponding to the eigenvalue one of a set of probability transition matrices.

Corollary 4.3.2. *Let s be a positive integer and let $\mathcal{A} = \{A_i\}_{i=1}^s$ be a set of doubly stochastic, probability transition matrices. Then,*

$$Null(\sum_{i=1}^s (A_i - I)) = \bigcap_{i=1}^s Null(A_i - I),$$

and $dim(Null(\sum_{i=1}^s (A_i - I))) = 1$ if and only if $G_{\mathcal{A}}$ is strongly connected.

Proof. Since A_i, $i = 1 \ldots s$ are doubly stochastic then $A_i - I$ are block diagonalizable doubly stochastic generator matrices. Therefore, by recursively applying Lemma 4.3.2 $s - 1$ times, the first part of the Corollary follows. For the second part of the Corollary, note that, by Corollary 3.5 of [39], $\frac{1}{N} \sum_{i=1}^s A_i$ has the algebraic multiplicity equal to one, of its eigenvalue $\lambda = 1$ if and only if the graph associated to $\frac{1}{N} \sum_{i=1}^s A_i$ has a spanning tree, or in our case is strongly connected, which in turn implies that $dim(Null(\sum_{i=1}^s (A_i - I))) = 1$ if and only if $G_{\mathcal{A}}$ is strongly connected. $\qquad\square$

The following Corollary is an immediate consequence of Corollary 3.5 of [39].

Corollary 4.3.3. *A generator matrix G has algebraic multiplicity equal to one for its eigenvalue $\lambda = 0$ if and only if the graph associated with the matrix has a spanning tree.*

Proof. Follows immediately from Corollary 3.5 of [39], by forming the probability transition matrix $P = I + \epsilon G$, for some appropriate $\epsilon > 0$, and noting that $Null(P - I) = Null(G)$.

$\square$

The following Corollary is the counterpart of Lemma 3.7 of [39], in the case of generator matrices.

Corollary 4.3.4. *Let $G \in \mathbb{R}^{n \times n}$ be a rate transition matrix. If G has an eigenvalue $\lambda = 0$ with algebraic multiplicity equal to one, then $\lim_{t \to \infty} e^{Gt} = \mathbb{1}v'$, where v is a nonnegative vector satisfying $G'v = 0$ and $v'\mathbb{1} = 1$.*

Proof. Choose $h_1 > 0$ and let $\{t_k^1\}_{k \geq 0}$ be a sequence given by $t_k^1 = h_1 k$, for all $k \geq 0$. Then

$$\lim_{k \to \infty} e^{Gt_k^1} = \lim_{k \to \infty} e^{h_1 kG} = \lim_{k \to \infty} P_{h_1}^k,$$

where we defined $P_{h_1} \triangleq e^{h_1 G}$. From the theory of continuous-time Markov chains we know that P_{h_1} is a stochastic matrix with positive diagonal entries and that, given a vector $x \in \mathbb{R}^n$, $x'P_{h_1} = x'$ if and only if $x'G = 0$. This means that the algebraic multiplicity of the eigenvalue $\lambda = 1$ of P_{h_1} is one. By Lemma 3.7 of [39], we have that $\lim_{k \to \infty} P_{h_1}^k = \mathbb{1}v'_{h_1}$, where v_{h_1} is a nonnegative vector satisfying $P'_{h_1} v_{h_1} = v_{h_1}$ and $v'_{h_1}\mathbb{1} = 1$. Also $G'v_{h_1} = 0$. Choose another $h_2 > 0$ and let $P_{h_2} \triangleq e^{h_2 G}$. Similarly as above, we have that

$$\lim_{k \to \infty} P_{h_2}^k = \mathbb{1}v'_{h_2},$$

where v_{h_2} satisfy similar properties as v_{h_1}. But since both vector belong to the nullspace of G' of dimension one, then they must be equal. Indeed if x is a left eigenvector of G, then v_{h_1} and v_{h_2} can be written as $v_{h_1} = \alpha_1 x$ and $v_{h_2} = \alpha_2 x$. However, since $\mathbb{1}'v_{h_1} = 1$ and

$\mathbb{1}' v_{h_2} = 1$ it follows that $\alpha_1 = \alpha_2$. We have shown that for any choice of $h > 0$,

$$\lim_{k \to \infty} e^{Gt_k} = e^{hkG} = \mathbb{1}v',$$

where v is a nonnegative vector satisfying $G'v = 0$ and $\mathbb{1}'v = 1$, and therefore, the result

follows. □

4.3.2 Preliminary results for the case where the agents' dynamics are expressed in discrete-time

In this subsection we state and prove a set of results used to prove Theorem 4.2.1 in the case where the agents' dynamics are expressed in discrete-time. Basically these results study the convergence properties of a sequence of matrices $\{Q^k\}_{k \geq 0}$, where Q has a particular structure which comes from the analysis of the first and second moment of the state vector $x(t)$.

Lemma 4.3.3. *Let s be a positive integer and let $\{A_{ij}\}_{i,j=1}^s$ be a set of $n \times n$ doubly stochastic, ergodic matrices. Let $P = (p_{ij})$ be a $s \times s$ stochastic matrix corresponding to an irreducible, homogeneous Markov chain and let $\mathcal{P}_\infty$ be the limit set of the sequence $\{P^k\}_{k \geq 0}$. Consider the $ns \times ns$ dimensional matrix Q whose $(i, j)^{th}$ block is defined by $Q_{ij} \triangleq p_{ji}A_{ij}$. Then $\mathcal{P}'_\infty \otimes \left(\frac{1}{n}\mathbb{1}\mathbb{1}'\right)$ is the limit set of the matrix sequence $\{Q^k\}_{k \geq 1}$, i.e.:*

$$\lim_{k \to \infty} d\left(Q^k, \mathcal{P}'_\infty \otimes \left(\frac{1}{n}\mathbb{1}\mathbb{1}'\right)\right) = 0. \tag{4.6}$$

Proof. The proof of this lemma is based on *Corollary 4.3.1*. The $(i, j)^{th}$ block entry of the matrix Q^k can be expressed as

112

$$(Q^k)_{ij} = \sum_{1 \le i_1, \dots i_{k-1} \le s} p_{ji_1} p_{i_1 i_2} \dots p_{i_{k-1}i} A_{ii_1} A_{i_1 i_2} \dots A_{i_{k-1}j}. \tag{4.7}$$

Let p_{ji}^∞ be the (j,i) entry of an arbitrary matrix in $\mathcal{P}_\infty$, i.e. there exist a sequence $\{t_k\}_{k \ge 1} \subset \mathbb{N}$ such that $\lim_{k \to \infty}(P^{t_k})_{ji} = p_{ji}^\infty$.

We have that

$$\left\| (Q^k)_{ij} - p_{ji}^\infty \frac{1}{n} \mathbb{1}\mathbb{1}' \right\| \le \sum_{1 \le i_{t_1}, \dots i_{k-1} \le s} (p_{ji_1} \dots p_{i_{k-1}i}) \left\| A_{ii_1} \dots A_{i_{k-1}j} - \frac{1}{n} \mathbb{1}\mathbb{1}' \right\| +$$

$$+ \sum_{1 \le i_1, \dots i_{k-1} \le s} |p_{ji_1} \dots p_{i_{k-1}i} - p_{ji}^\infty| \left\| \frac{1}{n} \mathbb{1}\mathbb{1}' \right\| \le$$

$$\le \max_{i_1, \dots i_{k-1}} \left\{ \left\| A_{ii_1} \dots A_{i_{k-1}j} - \frac{1}{n} \mathbb{1}\mathbb{1}' \right\| \right\} \sum_{1 \le i_1, \dots i_{k-1} \le s} p_{ji_1} \dots p_{i_{k-1}i} +$$

$$+ \left\| \frac{1}{n} \mathbb{1}\mathbb{1}' \right\| \sum_{1 \le i_1, \dots i_{k-1} \le s} |p_{ji_1} \dots p_{i_{k-1}i} - p_{ji}^\infty|,$$

where $\| \cdot \|$ was used to denote some matrix norm. Consider the limit of the left hand side of the above inequality for the sequence $\{t_k\}_{k \ge 0}$. By Corollary 4.3.1 we know that

$$\lim_{k \to \infty} A_{ii_{t_1}} \dots A_{i_{t_{k-1}}j} = \frac{1}{n} \mathbb{1}\mathbb{1}'$$

for all sequences $i_{t_1}, \dots, i_{t_{k-1}}$ and since obviously,

$$\lim_{k \to \infty} \sum_{1 \le i_{t_1}, \dots i_{t_{k-1}} \le s} p_{ji_{t_1}} \dots p_{i_{t_{k-1}}i} = p_{ji}^\infty,$$

it results

$$\lim_{k \to \infty} (Q^{t_k})_{ij} = p_{ji}^\infty \frac{1}{n} \mathbb{1}\mathbb{1}'.$$

Therefore $\mathcal{P}_\infty' \otimes \left(\frac{1}{n} \mathbb{1}\mathbb{1}' \right)$ is the limit set for the sequence of matrices $\{Q^k\}_{k \ge 1}$. $\qquad \square$

Lemma 4.3.4. *Let s be a positive integer and consider a set of doubly stochastic matrices with positive diagonal entries, $\mathcal{D} = \{D_i\}_{i=1}^{s}$, such that the corresponding graph $G_{\mathcal{D}}$ is strongly connected. Let P be the $s \times s$ dimensional probability transition matrix of an irreducible, homogeneous Markov chain and let $\mathcal{P}_{\infty}$ be the limit set of the sequence $\{P^k\}_{k \geq 0}$. Consider the $ns \times ns$ matrix Q whose blocks are given by $Q_{ij} \triangleq p_{ji}D_j$. Then $\mathcal{P}'_{\infty} \otimes \left(\frac{1}{n}\mathbb{1}\mathbb{1}'\right)$ is the limit set of the sequence of matrices $\{Q^k\}_{k \geq 1}$, i.e.:*

$$\lim_{k \to \infty} d\left(Q^k, \mathcal{P}'_{\infty} \otimes \left(\frac{1}{n}\mathbb{1}\mathbb{1}'\right)\right) = 0. \tag{4.8}$$

Proof. Our strategy consists in showing that there exist a $k \in \mathbb{N}$, such that each $(i,j)^{\text{th}}$ block matrix of Q^k becomes a weighted ergodic matrix, i.e $(Q^k)_{ij} = p_{ji}^{(k)}A_{ij}^{(k)}$, where $A_{ij}^{(k)}$ is ergodic and $p_{ji}^{(k)} = (P^k)_{ji}$. If this is the case, we can apply *Lemma 4.3.3* to obtain (4.8). The $(i,j)^{\text{th}}$ block matrix of Q^k looks as in (4.7), with the difference that in the current case $A_{ij} = D_j$:

$$(Q^k)_{ij} = \sum_{1 \leq i_1, \ldots i_{k-1} \leq s} p_{ji_1}p_{i_1 i_2} \ldots p_{i_{k-1}i} D_j D_{i_1} \ldots D_{i_{k-1}} = p_{ji}^{(k)} A_{ij}^{(k)} \tag{4.9}$$

where

$$A_{ij}^{(k)} \triangleq \sum_{1 \leq i_1, \ldots i_{k-1} \leq s} \alpha_{i_1, \ldots i_{k-1}} D_j D_{i_1} \ldots D_{i_{k-1}},$$

with

$$\alpha_{i_1, \ldots i_{k-1}} \triangleq \begin{cases} p_{ji_1}p_{i_1 i_2} \cdots p_{i_{k-1}i}/p_{ji}^{(k)}, & p_{ji}^{(k)} > 0 \\ \\ 0, & \text{otherwise} \end{cases}$$

Note that each of the matrix product $D_j D_{i_1} \ldots D_{i_{k-1}}$ appearing in $A_{ij}^{(k)}$, corresponds to a path from node j to node i in $k-1$ steps. Therefore, by the irreducibility assumption of P, there exists a k such that each matrix in the set $\mathcal{D}$ appears at least once in one of the terms of the sum (4.9), i.e. $\{1, \ldots, s\} \subseteq \{i_1, \ldots i_{k-1}\}$. Using a similar idea as in Lemma 1 in

[18] or Lemma 3.9 in [39], by Lemma 4.3.1, we upper bound such term

$$D_j D_{i_1} \ldots D_{i_{k-1}} \geq \gamma \sum_{l=1}^{s} D_l = \gamma s \bar{D}, \tag{4.10}$$

where $\gamma > 0$ depends on the matrices in $\mathcal{D}$ and $\bar{D}$ is a doubly stochastic matrix with positive entries

$$\bar{D} = \frac{1}{s} \sum_{i}^{s} D_i.$$

Since $G_\mathcal{D}$ is strongly connected, the same is true for $G_{\bar{D}}$. Therefore, $\bar{D}$ corresponds to an irreducible, aperiodic ($\bar{D}$ has positive diagonal entries) and hence ergodic, Markov chain. By inequality (4.10), it follows that the matrix product $D_j D_{i_1} \ldots D_{i_{k-1}}$ is ergodic. This is enough to infer that $A_{ij}^{(k)}$ is ergodic as well, since is a result of a convex combination of (doubly) stochastic matrices with at least one ergodic matrix in the combination. Choose a k^* large enough such that for all non-zero $p_{ij}^{(k^*)}$, the matrices $A_{ij}^{(k^*)}$ are ergodic $\forall i, j$. Such k^* always exists due to irreducibility assumption on P. Then according to Lemma 4.3.3, we have that for the subsequence $\{t_m\}_{m \geq 0}$, with $t_m = mk^*$

$$\lim_{m \to \infty} d\left(Q^{t_m}, \mathcal{P}'_\infty \otimes \left(\frac{1}{n} \mathbb{1} \mathbb{1}'\right)\right) = 0. \tag{4.11}$$

The result follows by Proposition 4.3.1 since $\|Q\|_1 \leq 1$ and since $Q\left(\mathcal{P}'_\infty \otimes \left(\frac{1}{n} \mathbb{1} \mathbb{1}'\right)\right) = \mathcal{P}'_\infty \otimes \left(\frac{1}{n} \mathbb{1} \mathbb{1}'\right).$

$\square$

Lemma 4.3.5. *Under the same assumptions as in Lemma 4.3.4, if we define the matrix blocks of Q as $Q_{ij} \triangleq p_{ji} D_j \otimes D_j$, then $\mathcal{P}'_\infty \otimes \left(\frac{1}{n^2} \mathbb{1} \mathbb{1}'\right)$ is the limit set of the sequence $\{Q^k\}_{k \geq 1}$, i.e.*

$$\lim_{k \to \infty} d\left(Q^k, \mathcal{P}'_\infty \otimes \left(\frac{1}{n^2} \mathbb{1} \mathbb{1}'\right)\right),$$

where the vector $\mathbb{1}$ above has dimension n^2.

Proof. In the current setup (4.9) becomes:

$$(Q^k)_{ij} = \sum_{1 \leq i_1,...i_{k-1} \leq s} p_{ji_1} p_{i_1 i_2} \ldots p_{i_{k-1} i} (D_j \otimes D_j)(D_{i_1} \otimes D_{i_1}) \ldots (D_{i_{k-1}} \otimes D_{i_{k-1}}). \qquad (4.12)$$

The result follows from the same arguments used in Lemma 4.3.4 together with the fact that the matrix products in (4.12) can be written as $(D_j \otimes D_j)(D_{i_1} \otimes D_{i_1}) \ldots (D_{i_{k-1}} \otimes D_{i_{k-1}}) = (D_j D_{i_1} \ldots D_{i_{k-1}}) \otimes (D_j D_{i_1} \ldots D_{i_{k-1}})$ and with the observation that the Kronecker product of an ergodic matrix with itself produces an ergodic matrix as well. $\qquad \square$

4.3.3 Preliminary results for the case where the agents' dynamics are expressed in continuous-time

The following two lemmas emphasize geometric properties of two matrices arising from the linear dynamics of the first and second moment of the state vector, in the continuous-time case.

Lemma 4.3.6. *Let s be a positive integer and let $C = \{C_i\}_{i=1}^s$ be a set of $n \times n$ doubly stochastic matrices such that G_C is strongly connected. Consider also a $s \times s$ generator matrix $\Lambda = (\lambda_{ij})$ corresponding to an irreducible Markov chain with stationary distribution $\pi = (\pi_i)$. Define the matrices $A \triangleq diag(C_i', i = 1 \ldots s)$ and $B \triangleq \Lambda \otimes I$. Then $A + B$ has an eigenvalue $\lambda = 0$ with algebraic multiplicity one and with corresponding right and left eigenvectors given by $\mathbb{1}_{ns}$ and $(\pi_1 \mathbb{1}_n', \pi_2 \mathbb{1}_n', \ldots, \pi_s \mathbb{1}_n')$, respectively.*

Proof. We first note that $A + B$ is a generator matrix and that both A and B are block diagonalizable (indeed A has doubly stochastic matrices on its main diagonal and B contains

116

n copies of the irreducible Markov chain corresponding to Λ). Therefore, $A + B$ has an eigenvalue $\lambda = 0$ with algebraic multiplicity at least one.

Let v be a vector in the null space of $A + B$. By Lemma 4.3.2, we have that $v \in Null(A)$ and $v \in Null(B)$. Given the structure of B, v must respect the following pattern $v' = \{(\underbrace{u'\ u'\ldots u'}_{s\ times}) \mid u \in \mathbb{R}^n\}$. But since $v \in Null(A)$, we have that $C'_i u = 0$, $i = 1\ldots s$, or $\mathbf{C}u = 0$, where $\mathbf{C} = \sum_{i=1}^{s} C'_i$. Since G_C was assumed strongly connected, $\mathbf{C}$ corresponds to an irreducible Markov chain, and it follows that u must be of the form $u = \alpha \mathbb{1}$, for some $\alpha \in \mathbb{R}$. By backtracking, we get that $v = \alpha \mathbb{1}$, for some $\alpha \in \mathbb{R}$ and consequently $Null(A + B) = span(\mathbb{1})$. Therefore, $\lambda = 0$ has algebraic multiplicity one, with right eigenvector given by $\mathbb{1}$. By simple verification we note that $(\pi_1 \mathbb{1}', \pi_2 \mathbb{1}', \ldots, \pi_s \mathbb{1}')$ is a left eigenvector corresponding to the eigenvalue $\lambda = 0$. $\square$

Lemma 4.3.7. *Let s be a positive integer and let $C = \{C_i\}_{i=1}^{s}$ be a set of $n \times n$ doubly stochastic matrices such that G_C is strongly connected. Consider also a $s \times s$ generator matrix $\Lambda = (\lambda_{ij})$ corresponding to an irreducible Markov chain with stationary distribution $\pi = (\pi_i)$. Define the matrices $A \triangleq diag(C'_i \oplus C'_i, i = 1\ldots s)$ and $B \triangleq \Lambda \otimes I$. Then $A + B$ has an eigenvalue $\lambda = 0$ with algebraic multiplicity one, with corresponding right and left eigenvectors given by $\mathbb{1}_{n^2 s}$ and $(\pi_1 \mathbb{1}'_{n^2}, \pi_2 \mathbb{1}'_{n^2}, \ldots, \pi_s \mathbb{1}'_{n^2})$, respectively.*

Proof. It is not difficult to check that $A + B$ is a generator matrix. Also we note that $C'_i \oplus C'_i = C'_i \otimes I + I \otimes C'_i$ is block diagonalizable since both $C'_i \otimes I$ and $I \otimes C'_i$ are block diagonalizable. Indeed, since C_i is doubly stochastic then it is block diagonalizable. The matrix $C'_i \otimes I$ contains n isolated copies of C'_i and therefore it is block diagonalizable. Also, $I \otimes C'_i$ it has a number of n block on its diagonal, each block being given by C'_i, and

117

it follows is block diagonalizable as well.

Let v be a vector in the nullspace of $A + B$. By Lemma 4.3.2, $v \in Null(A)$ and $v \in Null(B)$. From the structure of B we note that v must be of the form $v' = \underbrace{(u', \ldots, u')'}_{s \ times} \mid u \in \mathbb{R}^{n^2}$. Consequently we have that $(C_i' \oplus C_i')u = 0$, $i = 1, \ldots s$, or $(\mathbf{C} \otimes \mathbf{C})u = 0$, where $\mathbf{C} = \sum_{i=1}^{s} C_i'$. Since, $G_{\mathcal{G}}$ is strongly connected, $\mathbf{C}$ is a generator matrix corresponding to an irreducible Markov chain. By applying again Lemma 4.3.2 for the matrix $\mathbf{C} \oplus \mathbf{C} = I \otimes \mathbf{C} + \mathbf{C} \otimes I$, we get that u must have the form $u' = \underbrace{(\bar{u}', \ldots, \bar{u}')'}_{n \ times}$, where $\bar{u} \in \mathbb{R}^n$ and $\mathbf{C}\bar{u} = 0$. But $\mathbf{C}$ is irreducible and therefore $\bar{u} = \alpha \mathbb{1}_n$, or $u = \alpha \mathbb{1}_{n^2}$, or finally $v = \alpha \mathbb{1}_{n^2 s}$, where $\alpha \in \mathbb{R}$. Consequently, $Null(A + B) = span(\mathbb{1})$ which means the eigenvalue $\lambda = 0$ has algebraic multiplicity one. By simple verification, we note that $(\pi_1 \mathbb{1}'_{n^2}, \pi_2 \mathbb{1}'_{n^2}, \ldots, \pi_s \mathbb{1}'_{n^2})$ is a left eigenvector corresponding to the zero eigenvalue. □

4.4 Proof of the convergence theorem

The proof will focus on showing that the state vector $x(t)$ converges in mean square sense to average consensus. Equivalently, by making the change of variable $z(t) = x(t) - av(x_0)\mathbb{1}$, we will actually show that $z(t)$ is mean square stable for the initial condition $z(0) = x_0 - av(x_0)\mathbb{1}$, where $z(t)$ respects the same dynamic equation as $x(t)$. Using results for the stability theory of Markovian jump linear systems, mean square stability also imply stability in the almost sure sense (see for instance Corollary 3.46 of [11] for discrete-time case or Theorem 2.1 of [15] for continuous-time case, with the remark that we are interested for the stability property to be satisfied for a specific initial condition, rather then for any initial condition), which for us imply that $x(t)$ converges almost surely

to average consensus.

We first prove the discrete-time case after which we continue with the proof for the continuous-time case.

4.4.1 Discrete-time case - Sufficiency

Proof. Let $V(t)$ denote the second moment of the state vector

$$V(t) \triangleq E[x(t)x(t)^T],$$

where we used E to denote the expectation operator. The matrix $V(t)$ can be expressed as

$$V(t) = \sum_{i=1}^{s} V_i(t), \tag{4.13}$$

where $V_i(t)$ is given by

$$V_i(t) \triangleq E[x(t)x(t)^T \chi_{\{\theta(t)=i\}}] \ i = 1 \ldots s, \tag{4.14}$$

with $\chi_{\{\theta(t)=i\}}$ being the indicator function of the event $\{\theta(t) = i\}$.

The set of discrete coupled Lyapunov equations governing the evolution of the matrices $V_i(t)$ are given by

$$V_i(t+1) = \sum_{j=1}^{s} p_{ji} D_j V_j(t) D_j^T, \ i = 1 \ldots s, \tag{4.15}$$

with initial conditions $V_i(0) = q_i x_0 x_0^T$. By defining $\eta(t) \triangleq col(V_i(t), i = 1 \ldots s)$, we obtain a vectorized form of equations (4.15)

$$\eta(t+1) = \Gamma_d \eta(t), \tag{4.16}$$

119

where Γ_d is an $n^2 s \times n^2 s$ matrix given by

$$\Gamma_d = \begin{pmatrix} p_{11} D_1 \otimes D_1 & \dots & p_{s1} D_s \otimes D_s \\ \vdots & \ddots & \vdots \\ p_{1s} D_1 \otimes D_1 & \dots & p_{ss} D_s \otimes D_s \end{pmatrix} \text{ and } \eta_0 = \begin{pmatrix} q_1 col(x_0 x_0') \\ \vdots \\ q_s col(x_0 x_0') \end{pmatrix}. \qquad (4.17)$$

We note that Γ_d satisfies all the assumptions of *Lemma 4.3.5* and hence we get

$$\lim_{k \to \infty} d\left(\Gamma_d^k, \mathcal{P}_\infty' \otimes \left(\frac{1}{n^2} \mathbb{1}\mathbb{1}' \right) \right) = 0,$$

where $\mathcal{P}_\infty$ is the limit set of the matrix sequence $\{P^k\}_{k \geq 0}$. Using the observation that

$$\frac{1}{n^2} \mathbb{1}\mathbb{1}' col(x_0 x_0') = av(x_0)^2 \mathbb{1},$$

the limit of the sequence $\{\eta(t_k)\}_{k \geq 0}$, where $\{t_k\}_{k \geq 0}$ is such that $\lim_{k \to \infty} (P^{t_k})_{ij} = p_{ij}^\infty$, is

$$\lim_{k \to \infty} \eta(t_k)' = av(x_0)^2 \begin{pmatrix} \sum_{j=1}^s p_{j1}^\infty q_j \mathbb{1} \\ \vdots \\ \sum_{j=1}^s p_{js}^\infty q_j \mathbb{1}' \end{pmatrix}.$$

By collecting the entries of $\lim_{k \to \infty} \eta(t_k)$ we obtain

$$\lim_{k \to \infty} V_i(t_k) = av(x_0)^2 \left(\sum_{j=1}^s p_{ji}^\infty q_j \right) \mathbb{1}\mathbb{1}',$$

and from (4.13) we get

$$\lim_{k \to \infty} V(t_k) = av(x_0)^2 \mathbb{1}\mathbb{1}' \qquad (4.18)$$

since $\sum_{i,j=1}^s p_{ji}^\infty q_j = 1$. By repeating the previous steps for all subsequences generating limit points for $\{P^k\}_{k \geq 0}$ we obtain that (4.18) holds for any sequence in $\mathbb{N}$.

Through a similar process as in the case of the second moment (in stead of Lemma 4.3.5 we use Lemma 4.3.4), we show that:

$$\lim_{k \to \infty} E[x(t)] = av(x_0)\mathbb{1}. \tag{4.19}$$

From (4.18) and (4.19) we have that

$$\lim_{t \to \infty} E[\|x(t) - av(x_0)\mathbb{1}\|^2] = \lim_{t \to \infty} trace(E[(x(t) - av(x_0)\mathbb{1})(x(t) - av(x_0)\mathbb{1})']) =$$

$$= \lim_{t \to \infty} trace(E[x(t)x(t)'] - av(x_0)\mathbb{1}E[x(t)'] - av(x_0)E[x(t)]\mathbb{1}' + av(x_0)^2\mathbb{1}\mathbb{1}') = 0.$$

Therefore, $x(t)$ converges to average consensus in the mean square sense, and consequently in the almost sure sense, as well. □

4.4.2 Discrete-time case - Necessity

Proof. If $G_{\mathcal{A}}$ is not strongly connected then by Corollary 4.3.2, $dim(\bigcap_{i=1}^{s} Null(A_i - I)) > 1$. Consequently, there exist a vector $v \in \bigcap_{i=1}^{s} Null(A_i - I))$ such that $v \notin span(\mathbb{1})$. If we choose v as initial condition, for every realization of $\theta(t)$, we have that

$$x(t) = v, \text{ for all } t \geq 0,$$

and therefore consensus can not be reached in the sense of Definition 4.2.1. □

4.4.3 Continuous time - Sufficiency

Using the same notations as in the discrete-time case, the dynamic equations describing the evolution of the second moment of $x(t)$ are given by

$$\frac{d}{dt}V_i(t) = C_i V_i(t) + V_i(t)C_i' + \sum_{j=1}^{s} \lambda_{ji}V_j(t), \ i = 1 \ldots s, \tag{4.20}$$

121

equations whose derivation is treated in [16]. By defining the vector $\eta(t) \triangleq col(V_i(t), i = 1 \ldots s)$, the vectorized equivalent of equations (4.20) is given by

$$\frac{d}{dt}\eta(t) = \Gamma_c \eta(t), \tag{4.21}$$

where

$$\Gamma_c = \begin{pmatrix} C_1 \oplus C_1 & 0 & \cdots & 0 \\ 0 & C_2 \oplus C_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_s \oplus C_s \end{pmatrix} + \Lambda' \otimes I \text{ and } \eta_0 = \begin{pmatrix} q_1 col(x_0 x_0') \\ q_2 col(x_0 x_0') \\ \vdots \\ q_s col(x_0 x_0') \end{pmatrix}.$$

By Lemma 4.3.7, the eigenspace corresponding to the zero eigenvalue of Γ_c has dimension one, with unique (up to the multiplication by a scalar) left and right eigenvectors given by $\mathbb{1}_{n^2 s}$ and $\frac{1}{n^2}(\pi_1 \mathbb{1}'_{n^2}, \pi_2 \mathbb{1}'_{n^2}, \ldots, \pi_s \mathbb{1}'_{n^2})$, respectively. Since Γ'_c is a generator matrix with an eigenvalue zero of algebraic multiplicity one, by Corollary 4.3.4 we have that $\lim_{t \to \infty} e^{\Gamma'_c t} = v \mathbb{1}'$, where $v' = \frac{1}{n^2}(\pi_1 \mathbb{1}', \pi_2 \mathbb{1}', \ldots, \pi_s \mathbb{1}')$. Therefore, as t goes to infinity, we have that

$$\lim_{t \to \infty} \eta(t) = \begin{pmatrix} \pi_1 \frac{\mathbb{1}\mathbb{1}'}{n^2} & \cdots & \pi_1 \frac{\mathbb{1}\mathbb{1}'}{n^2} \\ \vdots & \ddots & \vdots \\ \pi_s \frac{\mathbb{1}\mathbb{1}'}{n^2} & \cdots & \pi_s \frac{\mathbb{1}\mathbb{1}'}{n^2} \end{pmatrix} \begin{pmatrix} q_1 col(x_0 x_0') \\ \vdots \\ q_s col(x_0 x_0') \end{pmatrix}.$$

By noting that

$$\frac{\mathbb{1}\mathbb{1}'}{n^2} col(x_0 x_0') = av(x_0)^2 \mathbb{1}_{n^2},$$

we farther get

$$\lim_{t \to \infty} \eta(t) = av(x_0)^2 \begin{pmatrix} \pi_1 \mathbb{1}_{n^2} \\ \vdots \\ \pi_s \mathbb{1}_{n^2} \end{pmatrix}.$$

122

Rearranging the columns of $\lim_{t\to\infty}\eta(t)$, we get

$$\lim_{t\to\infty} V_i(t) = av(x_0)^2 \pi_i \mathbb{1}\mathbb{1}',$$

or

$$\lim_{t\to\infty} V(t) = av(x_0)^2 \mathbb{1}\mathbb{1}'.$$

Through a similar process (using this time Lemma 4.3.6), we can show that

$$\lim_{t\to\infty} E[x(t)] = av(x_0)\mathbb{1}.$$

Therefore, $x(t)$ converges to average consensus in the mean square sense and consequently in the almost surely sense.

4.4.4   Continuous time - Necessity

Follows the same lines as in the discrete-time case.

4.5   Discussion

In the previous sections we proved a convergence result for the stochastic, linear consensus problem, for the cases where the dynamics of the agents were expressed in both discrete and continuous time. Our main contributions consist of considering a Markovian process, not necessarily ergodic, as underlying process for the random communication graph and of using a Markovian jump system theory inspired approach to prove this result. In what we have shown, we assumed that the Markov process $\theta(t)$ was irreducible and that the matrices D_i and C_i were doubly stochastic. We can assume for instance that $\theta(t)$ is not irreducible (i.e. $\theta(k)$ may have transient states). We treated this case in [23] (only for

discrete-time dynamics), and we showed that convergence in the sense of Definition 4.2.1 is achieved if and only if the union of graphs corresponding to each of the irreducible closed sets of states of the Markov chain produces a strongly connected graph. This should be intuitively clear since the probability to return to a transient state converges to zero as time goes to infinity, and therefore the influence of the matrices D_i (or C_i), corresponding to the transient states, is canceled. We can also assume that D_i and C_i are not necessarily doubly stochastic. We treated this case (again only for the discrete-time dynamics and without being completely rigorous) in [26] and we showed that the state converges in mean square sense and in almost sure sense to consensus, and not necessarily average consensus. From a technical point view, the difference lies in the fact that the $n^2 \times n^2$ block matrices of $\{\Gamma_d^t\}_{t \geq 0}$ (or $\{e^{t\Gamma_c}\}_{t \geq 0}$) no longer converge to $\pi_i \frac{1}{n^2} \mathbb{1}\mathbb{1}'$ but to $\pi_i \mathbb{1} c'$, for some vector $c \in \mathbb{R}^{n^2}$ with non-negative entries summing up to one; vector c which in general can not be a priori determined. In relevant distributed computation application (such as distributed state estimation or distributed optimization) however, convergence to average consensus is desired, and therefore the assumption, that D_i or C_i are doubly stochastic, makes sense.

The proof of Theorem 4.2.1 was based on the analysis of two matrix sequences $\{e^{\Gamma_c t}\}_{t \geq 0}$ and $\{\Gamma_d^t\}_{t \geq 0}$ arising from the dynamic equations of the state's second moment, for the continuous and discrete time, respectively. The reader may have noted that we approached differently the analysis of the two sequences. In the case of continuous-time dynamics, our approach was based on showing that the left and right eigenspaces induced by the zero eigenvalue of Γ_c have dimension one, and we provided the left and right eigenvectors (bases of the respective subspaces). The convergence of $\{e^{\Gamma_c t}\}_{t \geq 0}$ followed

from Corollary 4.3.4. In the case of the discrete-time dynamics, we analyzed the sequence $\{\Gamma_d^t\}_{t\geq 0}$, by looking at how the matrix blocks of Γ_d^t evolve as t goes to infinity. Although, similar to the continuous-time case, we could have proved properties of Γ_d related to the left and right eigenspaces induced by the eigenvalue one, this would not have been enough in the discrete-time case. This is because, through $\theta(t)$, Γ_d can be periodic, in which case the sequence $\{\Gamma_d^t\}_{t\geq 0}$ does not converge (remember that in the discrete-time consensus problems, the stochastic matrices are assumed to have positive diagonal entries, to avoid the possibility of being periodic).

In the case of i.i.d. random graphs [44], or more general, in the case of strictly stationary, ergodic random graphs [45] , a necessary and sufficient condition for reaching consensus almost surely (in the discrete-time case) is $|\lambda_2(E[\mathbf{D}_{\theta(t)}])| < 1$, where λ_2 denotes the eigenvalue with second largest modulus. In the case of Markovian random topology a similar condition, does not necessarily hold, neither for each time t, nor in the limit. Take, for instance, two (symmetric) stochastic matrices D_1 and D_2 such that each of the graphs G_{D_1} and G_{D_2}, respectively, are not strongly connected but their union is. If the two state Markov chain $\theta(t)$ is periodic, with transitions given by $p_{11} = p_{22} = 0$ and $p_{12} = p_{21} = 1$, we note that $\lambda_2(E[\mathbf{D}_{\theta(t)}]) = 1$, for all $t \geq 0$. Also note that $\lambda_2(\lim_{t\to\infty} E[\mathbf{D}_{\theta(t)}])$ does not exist since the sequence $\{E[\mathbf{D}_{\theta(t)}]\}_{t\geq 0}$ does not have a limit. Yet, consensus is reached. The assumption that allowed for the aforementioned necessary and sufficient condition to hold, was that $\theta(t)$ is a stationary process (which in turn implies that $E[D_{\theta(t)}]$ is constant for all $t \geq 0$). However, this is not necessarily true if $\theta(t)$ is a (homogeneous) irreducible Markov chain, *unless* the initial distribution is the stationary distribution.

For the discrete-time case we can formulate a result involving the second largest

eigenvalue of the time average expectation of $\mathbf{D}_{\theta(t)}$, i.e. $\lim_{N\to\infty} \frac{\sum_{t=1}^{N} E[\mathbf{D}_{\theta(t)}]}{N}$, which reflects the proportion of time $\mathbf{D}_{\theta(t)}$ spends in each state of the Markov chain.

Proposition 4.5.1. *Consider the stochastic system (4.1). Then, under Assumption 4.2.1, the state vector $x(t)$ converges to average consensus in the sense of Definition 4.2.1, if and only if*

$$\left| \lambda_2 \left(\lim_{N\to\infty} \frac{\sum_{t=0}^{N} E[\mathbf{D}_{\theta(t)}]}{N} \right) \right| < 1.$$

Proof. The time average of $E[\mathbf{D}_{\theta(t)}]$ can be explicitly written as

$$\lim_{N\to\infty} \frac{\sum_{t=0}^{N} E[\mathbf{D}_{\theta(t)}]}{N} = \sum_{i=1}^{s} \pi_i D_i = \bar{D},$$

where $\pi = (\pi_i)$ is the stationary distribution of $\theta(t)$. By Corollary 3.5 in [39], $|\lambda_2(\bar{D})| < 1$ if and only if the graph corresponding to $\bar{D}$ has a spanning tree, or in our case, is strongly connected. But the graph corresponding to $\bar{D}$ is the same as $G_{\mathcal{D}}$, and the result follows from Theorem 4.2.1. □

Unlike the discrete-time, in the case of continuous time dynamics, we know that if there exists a stationary distribution π (under the irreducibility assumption), the probability distribution of $\theta(t)$ converges to π, hence the time averaging is not necessary. In the following we introduce (without proof since basically it is similar to the proof of Proposition 4.5.1) a necessary and sufficient condition for reaching average consensus, involving the expected value of the second largest eigenvalue of $\mathbf{C}_{\theta(t)}$, for the continuous-time dynamics.

Proposition 4.5.2. *Consider the stochastic system (4.2). Then, under Assumption 4.2.1, the state vector $x(t)$ converges to average consensus in the sense of Definition 4.2.1, if*

and only if

$$Re\left(\lambda_2\left(\lim_{t\to\infty} E[\mathbf{C}_{\theta(t)}]\right)\right) < 0.$$

Our analysis provides also estimates on the rate of convergence to average consensus in the mean square sense. From linear dynamic equations of the state's second moment we notice that the eigenvalues of Γ_d and Γ_c dictates how fast the second moment converges to average consensus. Since Γ'_d is a probability transition matrix and since Γ'_c is a generator matrix, an estimate of the rate of convergence of the second moment of $x(t)$ to average consensus is given by the second largest eigenvalue (in modulus) of Γ_d, for the discrete-time case, and by the real part of the second largest eigenvalue of Γ_c, for the continuous time case.

Chapter 5

Distributed Consensus-Based Linear Filtering

5.1 Introduction

In this chapter we address the consensus-based distributed linear filtering problem as well. We assume that each agent updates its (local) estimate in two steps. In the first step, an update is produced using a Luenberger observer type of filter. In the second step, called *consensus step*, every sensor computes a convex combination between its local update and the updates received from the neighboring sensors. Our focus is *not* on designing the consensus weights, but on designing the *filter gains*. For given consensus weights, we will first give sufficient conditions for the existence of filter gains such that the dynamics of the estimation errors (without noise) are asymptotically stable. These sufficient conditions are also expressible in terms of the feasibility of a set of linear matrix inequalities. Next, we present a distributed (in the sense that each sensor uses only information available within its neighborhood), sub-optimal filtering algorithm, valid for time varying topologies as well, resulting from minimizing an upper bound on a quadratic cost expressed in terms of the covariances matrices of the estimation errors. In the case where the matrices defining the stochastic process and the consensus weights are time invariant, we present sufficient conditions such that the aforementioned distributed algorithm produces filter gains which converge and ensure the stability of the dynamics of the covariances matrices of the estimation errors. We will also present a connection

between the consensus-based linear filter and the linear filtering of an appropriately defined Markovian jump linear system. More precisely, we show that if the aforementioned Markovian jump linear system is (mean square) detectable then the stochastic process is detectable as well under the consensus-based distributed linear filtering scheme. Finally we show that the optimal gains of a linear filter for the state estimation of the Markovian jump linear system can be used to approximate the optimal gains of the consensus-based distributed linear filtering strategy.

Chapter structure: In Section 5.2 we describe the problems addressed in this chapter. Section 5.3 introduces the sufficient conditions for detectability under the consensus-based linear filtering scheme together with a test expressed in terms of the feasibility of a set of linear matrix inequalities. In Section 5.4 we present a sub-optimal distributed consensus based linear filtering scheme with quantifiable performance. Section 5.5 makes a connection between the consensus-based distributed linear filtering algorithm and the linear filtering scheme for a Markovian jump linear system.

Notations and Abbreviations: We represent the property of positive (semi-positive) definiteness of a symmetric matrix A, by $A > 0$ ($A \geq 0$). By convention, we say that a symmetric matrix A is *negative definite* (*semi-definite*) if $-A > 0$ ($-A \geq 0$) and we denote this by $A < 0$ ($A \leq 0$). By $A > B$ we understand that $A - B$ is positive definite. Given a set of square matrices $\{A_i\}_{i=1}^{N}$, by $diag(A_i, i = 1 \ldots N)$ we understand the block diagonal matrix which contains the matrices A_i's on the main diagonal. We use the abbreviations CBDLF, MJLS and LMI for consensus-based linear filter(ing), Markovian jump linear system and linear matrix inequality, respectively.

Remark 5.1.1. *Given a positive integer N, a set of vectors $\{x_i\}_{i=1}^N$, a set of non-negative scalars $\{p_i\}_{i=1}^N$ summing up to one and a positive definite matrix Q, the following holds*

$$\left(\sum_{i=1}^N p_i x_i\right)' Q \left(\sum_{i=1}^N p_i x_i\right) \leq \sum_{i=1}^N p_i x_i' Q x_i.$$

Remark 5.1.2. *Given a positive integer N, a set of vectors $\{x_i\}_{i=1}^N$, a set of matrices $\{A_i\}_{i=1}^N$ and a set of non-negative scalars $\{p_i\}_{i=1}^N$ summing up to one, the following holds*

$$\left(\sum_{i=1}^N p_i A_i x_i\right)\left(\sum_{i=1}^N p_i A_i x_i\right)' \leq \sum_{i=1}^N p_i A_i x_i x_i' A_i'. \tag{5.1}$$

5.2 Problem formulation

We consider a stochastic process modeled by a discrete-time linear dynamic equation

$$x(k+1) = A(k)x(k) + w(k), \ x(0) = x_0, \tag{5.2}$$

where $x(k) \in \mathbb{R}^n$ is the state vector and $w(k) \in \mathbb{R}^n$ is a driving noise, assumed Gaussian with zero mean and (possibly time varying) covariance matrix $\Sigma_w(k)$. The initial condition x_0 is assumed to be Gaussian with mean μ_0 and covariance matrix Σ_0. The state of the process is observed by a network of N sensors indexed by i, whose sensing models are given by

$$y_i(k) = C_i(k)x(k) + v_i(k), \ i = 1\ldots N, \tag{5.3}$$

where $y_i(k) \in \mathbb{R}^{r_i}$ is the observation made by sensor i and $v_i(k) \in \mathbb{R}^{r_i}$ is the measurement noise, assumed Gaussian with zero mean and (possibly time varying) covariance matrix $\Sigma_{v_i}(k)$. We assume that the matrices $\{\Sigma_{v_i}(k)\}_{i=1}^N$ and $\Sigma_w(k)$ are positive definite for $k \geq 0$ and that the initial state x_0, the noises $v_i(k)$ and $w(k)$ are independent for all $k \geq 0$. For

later reference we also define $\Sigma_{v_i}^{1/2}(k)$, $\Sigma_w^{1/2}(k)$, where $\Sigma_{v_i}(k) \triangleq \Sigma_{v_i}^{1/2}(k)\Sigma_{v_i}^{1/2}(k)'$ and $\Sigma_w(k) \triangleq$
$\Sigma_w^{1/2}(k)\Sigma_w^{1/2}(k)'$.

The set of sensors form a communication network whose topology is modeled by a directed graph that describes the information exchanged among agents. The goal of the agents is to (locally) compute estimates of the state of the process (5.2).

Let $\hat{x}_i(k)$ denote the state estimate computed by sensor i at time k and let $\epsilon_i(k)$ denote the estimation error, i.e. $\epsilon_i(k) \triangleq x(k) - \hat{x}_i(k)$. The covariance matrix of the estimation error of sensor i is denoted by $\Sigma_i(k) \triangleq E[\epsilon_i(k)\epsilon_i(k)']$, with $\Sigma_i(0) = \Sigma_0$.

The sensors update their estimates in two steps. In the first step, an intermediate estimate, denoted by $\varphi_i(k)$, is produced using a Luenberger observer filter

$$\varphi_i(k) = A(k)\hat{x}_i(k) + L_i(k)(y_i(k) - C_i(k)\hat{x}_i(k)), \ i = 1 \dots N, \tag{5.4}$$

where $L_i(k)$ is the *filter gain*.

In the second step, the new state estimate of sensor i is generated by a convex combination between $\varphi_i(k)$ and all other intermediate estimates within its communication neighborhood, i.e.

$$\hat{x}_i(k+1) = \sum_{j=1}^{N} p_{ij}(k)\varphi_j(k), \ i = 1 \dots N, \tag{5.5}$$

where $p_{ij}(k)$ are non-negative scalars summing up to one ($\sum_{j=1}^{N} p_{ij}(k) = 1$), and $p_{ij}(k) = 0$ if no link from j to i exists at time k. Having $p_{ij}(k)$ dependent on time accounts for a possibly time varying communication topology.

Combining (5.4) and (5.5) we obtain the dynamic equations for the consensus based

distributed filter:

$$\hat{x}_i(k+1) = \sum_{j=1}^{N} p_{ij}(k) \left[A(k)\hat{x}_j(k) + L_j(k) \left(y_j(k) - C_j(k)\hat{x}_j(k) \right) \right], \ i = 1 \dots N. \tag{5.6}$$

From (5.6) the estimation errors evolve according to

$$\epsilon_i(k+1) = \sum_{j=1}^{N} p_{ij}(k) \left[\left(A(k) - L_j(k)C_j(k) \right) \epsilon_j(k) + w(k) - L_j(k)v_j(k) \right], \ i = 1 \dots N. \tag{5.7}$$

We define the aggregate vectors of estimates, measurements, estimation errors, driving noise and measurements noise, respectively

$$\hat{\mathbf{x}}(k)' \triangleq (\hat{x}_1(k)', \dots, \hat{x}_N(k)'),$$

$$\mathbf{y}(k)' \triangleq (y_1(k)', \dots, y_N(k)'),$$

$$\epsilon(k)' \triangleq (\epsilon_1(k)', \dots, \epsilon_N(k)'),$$

$$\mathbf{w}(k)' \triangleq (w(k)', \dots, w(k)'),$$

$$\mathbf{v}(k)' \triangleq (v_1(k)', \dots, v_N(k)'),$$

and the following block matrices

$$\mathcal{A}(k) \triangleq \begin{pmatrix} A(k) & O_{n \times n} & \cdots & O_{n \times n} \\ O_{n \times n} & A(k) & \cdots & O_{n \times n} \\ \vdots & \vdots & \ddots & \vdots \\ O_{n \times n} & O_{n \times n} & \cdots & A(k) \end{pmatrix} \in \mathbb{R}^{nN \times nN},$$

$$C(k) \triangleq \begin{pmatrix} C_1(k) & O_{r_2 \times n} & \cdots & O_{r_N \times n} \\ O_{r_1 \times n} & C_2(k) & \cdots & O_{r_N \times n} \\ \vdots & \vdots & \ddots & \vdots \\ O_{r_1 \times n} & O_{r_2 \times n} & \cdots & C_N(k) \end{pmatrix} \in \mathbb{R}^{r \times nN}, \ \mathcal{L}(k) \triangleq \begin{pmatrix} L_1(k) & O_{n \times r_1} & \cdots & O_{n \times r_N} \\ O_{n \times r_1} & L_2(k) & \cdots & O_{n \times r_N} \\ \vdots & \vdots & \ddots & \vdots \\ O_{n \times r_1} & O_{n \times r_2} & \cdots & L_N(k) \end{pmatrix} \in \mathbb{R}^{nN \times r},$$

where $r = \sum_{i=1}^{N} r_i$. The dynamics (5.6) and (5.7) can be compactly written as

$$\hat{\mathbf{x}}(k+1) = \mathcal{P}(k)\mathcal{A}(k)\hat{\mathbf{x}}(k) + \mathcal{P}(k)\mathcal{L}(k)[\mathbf{y}(k) - C(k)\hat{\mathbf{x}}(k)], \tag{5.8}$$

$$\epsilon(k+1) = \mathcal{P}(k)[\mathcal{A}(k) - \mathcal{L}(k)C(k)]\epsilon(k) + \mathbf{w}(k) - \mathcal{P}(k)\mathcal{L}(k)\mathbf{v}((k), \tag{5.9}$$

where $\mathcal{P}(k) = P(k) \otimes I$ and $P(k) = (p_{ij}(k))$ is a stochastic matrix, with rows summing up to one.

Definition 5.2.1. *(distributed detectability) Assuming that $A(k)$, $C(k) \triangleq \{C_i(k)\}_{i=1}^{N}$ and $p(k) \triangleq \{p_{ij}(k)\}_{i,j=1}^{N}$ are time invariant, we say that the linear system (5.2) is detectable using the CBDLF scheme (5.6), if there exist a set of matrices $\mathbf{L} \triangleq \{L_i\}_{i=1}^{N}$ such that the system (5.7), without the noise inputs, is asymptotically stable.*

We introduce the following finite horizon quadratic filtering cost function

$$J_K(\mathbf{L}(\cdot)) = \sum_{k=0}^{K} \sum_{i=1}^{N} E[\|\epsilon_i(k)\|^2], \tag{5.10}$$

where by $\mathbf{L}(\cdot)$ we understand the set of matrices $\mathbf{L}(\cdot) \triangleq \{L_i(k), k = 0 \ldots K - 1\}_{i=1}^{N}$. The optimal filtering gains represent the solution of the following optimization problem

$$\mathbf{L}^*(\cdot) = \arg\min_{\mathbf{L}(\cdot)} J^K(\mathbf{L}(\cdot)). \tag{5.11}$$

Assuming that $A(k)$, $\mathbf{C}(k) \triangleq \{C_i(k)\}_{i=1}^{N}$, $\Sigma_w(k)$, $\Sigma_v(k) \triangleq \{\Sigma_{v_i}(k)\}$ and $\mathbf{p}(k) \triangleq \{p_{ij}(k)\}_{i,j=1}^{N}$ are time invariant, we can also define the infinite horizon filtering cost function

$$J_\infty(\mathbf{L}) = \lim_{K\to\infty} \frac{1}{K} J_K(\mathbf{L}) = \lim_{k\to\infty} \sum_{i=1}^{N} E[\|\epsilon_i(k)\|^2], \tag{5.12}$$

where $\mathbf{L} \triangleq \{L_i\}_{i=1}^{N}$ is the set of steady state filtering gains. By solving the optimization problem

$$\mathbf{L}^* = \arg\min_{\mathbf{L}} J^\infty(\mathbf{L}), \tag{5.13}$$

133

we obtain the optimal steady-state filter gains.

In the next sections we will address the following problems:

Problem 5.2.1. *(Detectability conditions) Under the above setup, we want to find conditions under which the system (5.2) is detectable in the sense of Definition 5.2.1.*

Problem 5.2.2. *(Sub-optimal scheme for consensus based distributed filtering) Ideally, we would like to obtain the optimal filter gains by solving the optimization problems (5.11) and (5.13), respectively. Due to the complexity of these problems, we will not provide the optimal filtering gains but rather focus on providing a sub-optimal scheme with quantifiable performance.*

Problem 5.2.3. *(Connection with the linear filtering of a Markovian jump linear system) We make a parallel between the consensus-based distributed linear filtering scheme and the linear filtering of a particular Markovian jump linear system.*

5.3 Distributed detectability

We start with a toy example motivating our interest in the distributed detectability problem under the CBDLF scheme. Let us assume that no single pair (A, C_i) is detectable in the classical sense, but the pair (A, C) is detectable, where $C' = (C'_1, \ldots, C'_N)$. In this case, we can design a stable (centralized) Luenberger observer filter. The question is, can we obtain a stable consensus-based distributed filter? As the following example will show, in general this is not true. That is why it is important to find conditions under which the CBDLF can produce stable estimates.

Example 5.3.1. *(Centralized detectable but not distributed detectable) Consider a linear*

dynamics as in (5.2-5.3), with two sensors, where

$$A = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}, \; C_1 = (\; 1 \quad 0 \;) \; and \; C_2 = (\; 0 \quad 1 \;).$$

Obviously, the pairs (A, C_1) and (A, C_2) are not detectable while the pair (A, C) is,

where $C' = (C'_1 \; C'_2)$ is. Let $L'_1 = (l_1 \; l_2)$ and $L'_2 = (l_3 \; l_4)$. For this example, the matrix that

dictates the stability property of (5.9) is given by

$$\mathcal{A} = \begin{pmatrix} p_{11}(10 - l_1) & 0 & 10p_{12} & -p_{12}l_3 \\ -p_{11}l_2 & 10p_{11} & 0 & p_{12}(10 - l_4) \\ p_{21}(10 - l_1) & 0 & 10p_{22} & -p_{22}l_3 \\ -p_{21}l_2 & 10p_{21} & 0 & p_{22}(10 - l_4) \end{pmatrix}$$

For $p_{11} = 0.9$, $p_{12} = 0.1$, $p_{21} = 0.7$ and $p_{22} = 0.3$, the characteristic polynomial of the

above matrix is given by

$$q(s) = s^4 + q_3(l_1, l_3)s^3 + q_2(l_1, l_4, l_2l_3)s^2 + q_1(l_1, l_4) + q_0(l_1, l_4),$$

where

$$q_3(l_1, l_3) = -24 + 0.9l_1 + 0.3l_4,$$

$$q_2(l_1, l_4, l_2l_3) = -0.07l_2l_3 - 5.6l_4 + 184 - 12.8l_1 + 0.27l_1l_4,$$

$$q_1(l_1, l_4) = 30l_4 - 480 - 2.4l_1l_4 + 42l_1,$$

$$q_0(l_1, l_4) = -40l_1 - 40l_4 + 4l_1l_4 + 400.$$

Let $\lambda_i(l_1, l_4, l_2l_3)$ denote the eigenvalues of $\mathcal{A}$. We define $\lambda_{max}(l_1, l_4, l_2l_3) = \max_i |\lambda_i(l_1, l_4, l_2l_3)|$.

The system (5.2-5.3) is not detectable in the sense of Definition 5.2.1 if $\lambda_{max}(l_1, l_4, l_2l_3) > 1$

135

for all values of l_1, l_2 and of the product l_2l_3. We introduce also the quantity $\lambda_{max}^{23}(l_2l_3) = \min_{l_1,l_4} \lambda_{max}(l_1, l_4, l_2l_3)$.
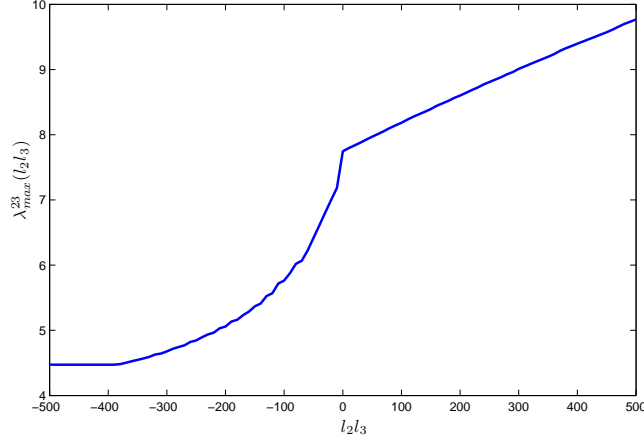

Figure 5.1: The evolution of $\lambda_{max}^{23}(l_2l_3)$

From Figure 5.1, we note that $\min_{l_2l_3} \lambda_{max}^{23}(l_2l_3) = 4.498$, which shows that, for the given consensus weights, and matrices A, C_1 and C_2, there are no values for l_1, l_2, l_3 and l_4, such that (5.9) can be made asymptotically stable.

The CBDLF (5.8) uses only one consensus step and we have seen, through Example 5.3.1, that in general this does not guarantee stable estimates, even in the case where the pair (A, C) is detectable. However, as the next proposition suggests, stable estimates might be achieved if a large enough number of consensus steps is used, i.e. we set $\mathcal{P}(k) = P(k)^\eta \otimes I$, for some positive integer value η, large enough.

Proposition 5.3.1. *Consider the linear dynamics (5.2)-(5.3). Assume that in the CBDLF scheme (5.6), we have $p_{ij} = \frac{1}{N}$ and that $\hat{x}_i(0) = x_0$, for all $i, j = 1 \dots N$. If the pair (A, C) is detectable, then the system (5.2) is detectable as well, in the sense of Definition 5.2.1.*

Proof. Rewrite the matrix C as

$$C = \sum_{i=1}^{N} \bar{C}_i,$$

where $\bar{C}_i' = (O_{n \times r_1} \ldots O_{n \times r_{i-1}} \ C_i' \ O_{n \times r_{i+1}} \ldots O_{n \times r_N})$. Ignoring the noise, we define the measurements

$$\bar{y}_i(k) = \bar{C}_i x(k),$$

which are equivalent to the ones in (5.3). Under the assumption that $p_{ij} = \frac{1}{N}$ and $\hat{x}_i = x_0$ for all $i, j = 1 \ldots N$, it follows that the estimation errors respect the dynamics

$$\epsilon(k+1) = \frac{1}{N} \sum_{i=1}^{N} (A - L_i \bar{C}_i) \epsilon(k). \tag{5.14}$$

Setting $L_i = NL$ for $i = 1 \ldots N$, it follows that

$$\epsilon(k+1) = (A - LC)\epsilon(k).$$

Since the pair (A, C) is detectable, there exists a matrix L such that $A - LC$ has all eigenvalues within the unit circle and therefore the dynamics (5.14) is asymptotically stable, which implies that (5.2) is detectable in the sense of Definition 5.2.1. □

The previous proposition tells us that if we achieve (average) consensus between the state estimates at each time instant, and if the pair (A, C) is detectable (in the classical sense), then the system (5.2) is detectable in the sense of Definition 5.2.1. However, achieving consensus at each time instant can be time and numerically costly and that is why it is important to find (testable) conditions under which the CBDLF produces stable estimates.

Lemma 5.3.1. *(sufficient conditions for distributed detectability) If there exists a set of*

symmetric, positive definite matrices $\{Q_i\}_{i=1}^N$ *and a set of matrices* $\{L_i\}_{i=1}^N$ *such that*

$$Q_i = \sum_{j=1}^N p_{ji}(A - L_jC_j)'Q_j(A - L_jC_j) + S_i, \ i = 1 \dots N, \tag{5.15}$$

for some positive definite matrices $\{S_i\}_{i=1}^N$, *then the system (5.2) is detectable in the sense of Definition 5.2.1.*

Proof. The dynamics of the estimation error without noise is given by

$$\epsilon_i(k+1) = \sum_{j=1}^N p_{ij}(A - L_jC_j)\epsilon_j(k), \ i = 1 \dots N. \tag{5.16}$$

In order to prove the stated result we have to show that (5.16) is asymptotically stable. We define the Lyapunov function

$$V(k) = \sum_i^N x_i(k)'Q_i x_i(k),$$

and our goal is to show that $V(k+1) - V(k) < 0$ for all $k \geq 0$. The Lyapunov difference is given by

$$V(k+1) - V(k) = \sum_{i=1}^N \left(\sum_{j=1}^N p_{ij}(A - L_jC_j)\epsilon_j(k) \right)' Q_i \left(\sum_{j=1}^N p_{ij}(A - L_jC_j)\epsilon_j(k) \right) - \epsilon_i(k)'TQ_i\epsilon_i(k) \leq$$

$$\leq \sum_{i=1}^N \left(\sum_{j=1}^N p_{ij}\epsilon_j(k)'(A - L_jC_j)'Q_i(A - L_jC_j)\epsilon_j(k) \right) - \epsilon_i(k)'Q_i\epsilon_i(k), \tag{5.17}$$

where the inequality followed from Remark 5.1.1. By changing the summation order we can further write

$$V(k+1) - V(k) \leq \sum_{i=1}^N \epsilon_i(k)' \left(\sum_{j=1}^N p_{ji}(A - L_jC_j)'Q_j(A - L_jC_j) - Q_i \right) \epsilon_i(k).$$

Using (5.15) yields

$$V(k+1) - V(k) \leq -\sum_{i=1}^N \epsilon_i(k)'S_i\epsilon_i(k)$$

138

From the fact that $\{S_j\}_{j=1}^N$ are positive definite matrices, we get

$$V(k+1) - V(k) < 0,$$

which implies that (5.16) is asymptotically stable. □

The following result relates the existence of the sets of matrices $\{Q_i\}_{i=1}^N$ and $\{L_i\}_{i=1}^N$ such that (5.15) is satisfied, with the feasibility of a set of linear matrix inequalities (LMI).

Proposition 5.3.2. *(distributed detectability test) The linear system (5.2) is detectable in the sense of Definition 5.2.1 if the following linear matrix inequalities, in the variables $\{X_i\}_{i=1}^N$ and $\{Y_i\}_{i=1}^N$, are feasible*

$$\begin{pmatrix} X_i & \sqrt{p_{1i}}(A'X_1 - C_1'Y_1') & \sqrt{p_{2i}}(A'X_2 - C_2'Y_2') & \cdots & \sqrt{p_{Ni}}(A'X_N - C_N'Y_N') \\ \sqrt{p_{1i}}(X_1A - Y_1C1) & X_1 & 0 & \cdots & 0 \\ \sqrt{p_{2i}}(X_2A - Y_2C2) & 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sqrt{p_{Ni}}(X_NA - Y_NC_N) & 0 & 0 & \cdots & X_N \end{pmatrix} > 0,$$

(5.18)

for $i = 1\ldots N$ and where $\{X_i\}_{i=1}^N$ are symmetric. Moreover, a stable CBDLF is obtained by choosing the filter gains as $L_i = X_i^{-1}Y_i$ for $i = 1\ldots N$.

Proof. First we note that, by the Schur complements Lemma, the linear matrix inequalities (5.18) are feasible if and only if there exist a set a symmetric matrices $\{X_i\}_{i=1}^N$ and a set of matrices $\{Y_i\}_{i=1}^N$, such that

$$X_i - \sum_{j=1}^N (X_jA - Y_jC_j)'X_j^{-1}(X_jA - Y_jC_j) > 0, \ X_i > 0$$

for all $i = 1\ldots N$. We further have that,

$$X_i - \sum_{j=1}^N (A - X_j^{-1}Y_jC_j)'X_j(X_jA - X_j^{-1}Y_jC_j) > 0, \ X_i > 0$$

139

By defining $L_i \triangleq X_i^{-1} Y_i$, it follows that

$$X_i - \sum_{j=1}^{N} (A - L_j C_j)' X_j (A - L_j C_j) > 0, \ X_i > 0.$$

Therefore, if the matrix inequalities (5.18) are feasible, there exists a set of positive definite matrices $\{X_i\}_{i=1}^{N}$ and a set of positive matrices $\{S_i\}_{i=1}^{N}$, such that

$$X_i = \sum_{j=1}^{N} (A - L_j C_j)' X_j (A - L_j C_j) + S_i.$$

By Lemma 5.3.1, it follows that the linear dynamics (5.7), without noise, is asymptotically stable, and therefore the system (5.2 is detectable in the sense of Definition 5.2.1. □

5.4 Sub-Optimal Consensus-Based Distributed linear Filtering

Obtaining the closed form solution of the optimization problem (5.11) is a challenging problem, which is in the same spirit as the decentralized optimal control problem. In this section we provide a sub-optimal algorithm for computing the filter gains of the CB-DLF, with quantifiable performance in the sense that we compute a set of filtering gains which guarantee a certain level of performance with respect the quadratic cost (5.10).

5.4.1 Finite Horizon Sub-Optimal Consensus-Based Distributed Linear Filtering

The sub-optimal scheme for computing the CBDLF gains results from minimizing an upper bound of the quadratic filtering cost (5.10). The following proposition gives upper-bounds for the covariance matrices of the estimation errors.

Proposition 5.4.1. *Consider the following coupled difference equations*

$$Q_i(k+1) = \sum_{i=1}^{N} p_{ij}(k) \Big[\big(A(k) - L_j(k)C_j(k)\big) Q_j(k) \big(A(k) - L_j(k)C_j(k)\big)' +$$

$$+ L_j(k)\Sigma_{v_j}(k)L_j(k) \Big] + \Sigma_w(k), \tag{5.19}$$

with $Q_i(0) = \Sigma_i(0)$, for $i = 1\ldots N$. The following inequality holds

$$\Sigma_i(k) \preceq Q_i(k), \tag{5.20}$$

for $i = 1\ldots N$ and for all $k \geq 0$.

Proof. Using (5.7), the matrix $\Sigma_i(k+1)$ can be explicitly written as

$$\Sigma_i(k+1) = E[\epsilon_i(k+1)'\epsilon_i(k+1)] =$$

$$= E\left[\left(\sum_{j=1}^{N} p_{ij}(k)\big(A(k) - L_j(k)C_j(k)\big)\epsilon_j(k) + w(k) - \sum_{j=1}^{N} p_{ij}(k)L_j(k)v_j(k) \right)'\right.$$

$$\left. \left(\sum_{j=1}^{N} p_{ij}(k)\big(A(k) - L_j(k)C_j(k)\big)\epsilon_j(k) + w(k) - \sum_{j=1}^{N} p_{ij}(k)L_j(k)v_j(k) \right) \right].$$

Using the fact that the noises $w(k)$ and $v_i(k)$ have zero mean, and they are independent with respect to themselves and the initial state, for every time instant, we can further write

$$\Sigma_i(k+1) = E\left[\left(\sum_{j=1}^{N} p_{ij}(k)\big(A(k) - L_j(k)C_j(k)\big)\epsilon_j(k) \right)' \left(\sum_{j=1}^{N} p_{ij}(k)\big(A(k) - L_j(k)C_j(k)\big)\epsilon_j(k) \right) \right] +$$

$$+ E\left[\left(\sum_{j=1}^{N} p_{ij}(k)L_j(k)v_j(k) \right)' \left(\sum_{j=1}^{N} p_{ij}(k)L_j(k)v_j(k) \right) \right] + \Sigma_w(k).$$

By Remark 5.1.2, it follows that

$$E\left[\left(\sum_{j=1}^{N} p_{ij}(k)\big(A(k) - L_j(k)C_j(k)\big)\epsilon_j(k) \right)' \left(\sum_{j=1}^{N} p_{ij}(k)\big(A(k) - L_j(k)C_j(k)\big)\epsilon_j(k) \right) \right] \leq$$

$$\leq \sum_{j=1}^{N} p_{ij}(k)\left(A(k) - L_j(k)C_j(k)\right)\Sigma_j(k)\left(A(k) - L_j(k)C_j(k)\right)'$$

and

$$E\left[\left(\sum_{j=1}^{N} p_{ij}(k)L_j(k)v_j(k)\right)'\left(\sum_{j=1}^{N} p_{ij}(k)L_j(k)v_j(k)\right)\right] \leq \sum_{j=1}^{N} p_{ij}(k)L_j(k)\Sigma_{v_j}(k)L_j(k)', \ i = 1\ldots N.$$

From the previous two expressions, we obtain that

$$\Sigma_i(k+1) \leq \sum_{j=1}^{N} p_{ij}(k)\left(A(k) - L_j(k)C_j(k)\right)\Sigma_j(k)\left(A(k) - L_j(k)C_j(k)\right)' +$$

$$+ \sum_{j=1}^{N} p_{ij}(k)L_j(k)\Sigma_{v_j}(k)L_j(k) + \Sigma_w(k)$$

We prove (5.20) by induction. Assume that $\Sigma_i(k) \leq Q_i(k)$ for all $i = 1\ldots N$. Then

$$(A(k) - L_i(k)C_i(k))\Sigma_i(k)(A(k) - L_i(k)C_i(k))' \leq (A(k) - L_i(k)C_i(k))Q_i(k)(A(k) - L_i(k)C_i(k))',$$

and

$$L_i(k)\Sigma_i(k)L_i(k)' \leq L_i(k)Q_i(k)L_i(k)', \ i = 1\ldots N.$$

and therefore

$$\Sigma_i(k+1) \leq Q_i(k+1), \ i = 1\ldots N.$$

$\square$

Defining the finite horizon quadratic cost function

$$\bar{J}_K(\mathbf{L}(\cdot)) = \sum_{k=1}^{K}\sum_{i=1}^{N} tr(Q_i(k)), \tag{5.21}$$

the next Corollary follows immediately.

Corollary 5.4.1. *The following inequalities hold*

$$J^K(\mathbf{L}(\cdot)) \leq \bar{J}^K(\mathbf{L}(\cdot)), \tag{5.22}$$

and

$$\limsup_{K \to \infty} \frac{1}{K} J^K(\mathbf{L}) \le \limsup_{K \to \infty} \frac{1}{K} \bar{J}^K(\mathbf{L}) \tag{5.23}$$

Proof. Follows immediately from Proposition 5.4.1. □

In the previous Corollary we obtained an upper bound on the filtering cost function. Our sub-optimal consensus based distributed filtering scheme will result from minimizing this upper bound in terms of the filtering gains $\{L_i(k)\}_{i=1}^N$:

$$\min_{\mathbf{L}(\cdot)} \bar{J}_K(\mathbf{L}(\cdot)). \tag{5.24}$$

Proposition 5.4.2. *The optimal solution for the optimization problem (5.24) is*

$$L_i^*(k) = A(k)Q_i^*(k)C_i(k)' \left[\Sigma_{v_i}(k) + C_i(k)Q_i^*(k)C_i(k)' \right]^{-1}, \tag{5.25}$$

and the optimal value is given by

$$\bar{J}_K^*(\mathbf{L}^*(\cdot)) = \sum_{k=1}^K \sum_{i=1}^N tr(Q_i^*(k)),$$

where $Q_i^*(k)$ is computed using

$$\begin{aligned} Q_i^*(k+1) = \sum_{j=1}^N p_{ij}(k) \Big[A(k)Q_j^*(k)A(k)' + \Sigma_w(k) - A(k)Q_j^*(k)C_j(k)' \cdot \\ \cdot \left(\Sigma_{v_j}(k) + C_j(k)Q_j^*(k)C_j(k)' \right)^{-1} C_j(k)Q_j^*(k)A(k)' \Big], \end{aligned} \tag{5.26}$$

with $Q_i^*(0) = \Sigma_i(0)$ and for $i = 1 \dots N$.

Proof. Let $\bar{J}_K(\mathbf{L}(\cdot))$ be the cost function when an arbitrary set of filtering gains $\mathbf{L}(\cdot) \triangleq \{L_i(k), k = 0 \dots K-1\}_{i=1}^N$ is used in (5.19). We will show that $\bar{J}_K^*(\mathbf{L}^*(\cdot)) \le \bar{J}_K(\mathbf{L}(\cdot))$, which in turn will show that $\mathbf{L}^*(\cdot) \triangleq \{L_i(k)^*, k = 0 \dots K-1\}_{i=1}^N$ is the optimal solution of the optimization problem (5.24). Let $\{Q_i^*(k)\}_{i=1}^N$ and $\{Q_i(k)\}_{i=1}^N$ be the matrices obtained when $\mathbf{L}^*(\cdot)$ and

$\mathbf{L}(\cdot)$, respectively are substituted in (5.19). In what follows we will show by induction that $Q_i^*(k) \le Q_i(k)$ for $k \ge 0$ and $i = 1\dots N$, which basically proves that $\bar{J}_K^*(\mathbf{L}^*(\cdot)) \le \bar{J}_K(\mathbf{L}(\cdot))$, for any $\mathbf{L}(\cdot)$. For simplifying the proof, we will omit in what follows the time index for some matrices and for the consensus weights.

Substituting $\{L_i^*(k), k \ge 0\}_{i=1}^N$ in (5.19), after some matrix manipulations we get

$$Q_i^*(k+1) = \sum_{j=1}^N p_{ij} \Big[AQ_j^*(k)A' + \Sigma_w - AQ_j^*(k)C_j'(\Sigma_{v_j} + $$

$$+ C_j Q_j^*(k)C_j')^{-1} C_j Q_j^*(k)A' \Big], \ Q_i^*(0) = \Sigma_i(0), \ i = 1\dots N.$$

We can derive the following matrix identity (for simplicity we will give up the time index):

$$(A + L_i C_i)Q_i(A_i + L_i C_i)' + L_i \Sigma_{v_i} L_i' = (A + L_i^* C_i)Q_i(A_i + L_i^* C_i)' + L_i^* \Sigma_{v_i} L_i^{*'} + $$

$$+(L_i - L_i^*)(\Sigma_{v_i} + C_i Q_i C_i')(L_i - L_i^*). \tag{5.27}$$

Assume that $Q_i^*(k) \le Q_i(k)$ for $i = 1\dots N$. Using identity (5.27), the dynamics of $Q_i(k)^*$ becomes

$$Q_i^*(k+1) = \sum_{j=1}^N p_{ij} \Big((A + L_j(k)C_j)Q_j(k)(A + L_j(k)C_j)' + L_j(k)\Sigma_{v_j} L_j(k)' - $$

$$- (L_j(k) - L_j^*(k))(\Sigma_{v_j} + C_j Q_j(k)C_j')(L_j(k) - L_j^*(k))' + \Sigma_w \Big).$$

The difference $Q_i^*(k+1) - Q_i(k+1)$ can be written as

$$Q_i(k+1)^* - Q_i(k+1) = \sum_{j=1}^N p_{ij} \Big((A + L_j(k)C_j)(Q_j^*(k) - Q_j(k))(A + L_j(k)C_j)' - $$

$$- (L_j(k) - L_j^*(k))(\Sigma_{v_j} + C_j Q_j(k)C_j')(L_j(k) - L_j^*(k))' \Big).$$

Since $\Sigma_{v_i} + C_i Q_i(k) C_i'$ is positive definite for all $k \geq 0$ and $i = 1 \dots N$, and since we assumed that $Q_i^*(k) \preceq Q_i(k)$, it follows that $Q_i^*(k+1) \preceq Q_i(k+1)$. Hence we obtain that

$$\bar{J}_K^*(\mathbf{L}^*(\cdot)) \leq \bar{J}_K(\mathbf{L}(\cdot)),$$

for any set of filtering gains $\mathbf{L}(\cdot) = \{L_i(k), k = 0 \dots K - 1\}_{i=1}^{N}$, which concludes the proof. $\square$

We summarize in the following algorithm the sub-optimal CBDLF scheme resulting from Proposition 5.4.2.

Algorithm 1: Consensus Based Distributed Linear Filtering Algorithm

---

Input*: μ_0, P_0*

Initialization: $\hat{x}_i(0) = \mu_0$, $Y_i(0) = \Sigma_0$

while *new data exists* **do**

    Compute the filter gains:

$$L_i \leftarrow A Y_i C_i' (\Sigma_{v_i} + C_i Y_i C_i')^{-1}$$

    Update the state estimates:

$$\varphi_i \leftarrow A\hat{x}_i + L_i(y_i - C - i\hat{x}_i)$$

$$\hat{x}_i \leftarrow \sum_{j} p_{ij} \varphi_j$$

    Update the matrices Y_i:

$$Y_i \leftarrow \sum_{j=1}^{N} p_{ij}\Big((A - L_j C_j) Y_j (A - L_j C_j)' + L_j \Sigma_{v_j} L_j'\Big) + \Sigma_w$$

    end

5.4.2 Infinite Horizon Consensus Based Distributed Filtering

We now assume that the matrices $A(k)$, $\{C_i(k)\}_{i=1}^N$, $\{\Sigma_{v_i}(k)\}_{i=1}^N$ and $\Sigma_w(k)$ and the weights $\{p_{ij}(k)_{i,j=1}^N\}$ are time invariant. We are interested in finding out under what conditions Algorithm 1 converges and if the filtering gains produce stable estimates. From the previous section we note that the optimal infinite horizon cost can be written as

$$\bar{J}_\infty^* = \lim_{k\to\infty} \sum_{i=1}^N tr(Q_i^*(k)),$$

where the dynamics of $Q_i(k)^*$ is given by

$$Q_i^*(k+1) = \sum_{j=1}^N p_{ij}\left[AQ_j^*(k)A' + \Sigma_w - AQ_j^*(k)C_j'\left(\Sigma_{v_j} + C_jQ_j^*(k)C_j'\right)^{-1}C_jQ_j^*(k)A'\right], \quad (5.28)$$

and the optimal filtering gains are given by

$$L_i^*(k) = AQ_i^*(k)C_i'\left[\Sigma_{v_i} + C_iQ_i^*(k)C_i'\right]^{-1},$$

for $i = 1\ldots N$. Assuming that (5.28), converges, the optimal value of the cost $\bar{J}_\infty^*$ is given by

$$\bar{J}_\infty^* = \sum_{i=1}^N tr(\bar{Q}_i),$$

where $\{\bar{Q}_i\}_{i=1}^N$ satisfy

$$\bar{Q}_i = \sum_{j=1}^N p_{ij}\left[A\bar{Q}_jA' + \Sigma_w - A\bar{Q}_jC_j'(\Sigma_{v_j} + C_j\bar{Q}_jC_j')^{-1}C_j\bar{Q}_jA'\right]. \quad (5.29)$$

Sufficient conditions under which there exists a unique solution of (5.29) are provided by Proposition A.2.1, which says that if $(\mathbf{p}, \mathbf{L}, \mathbf{A})$ is detectable and $(\mathbf{A}, \Sigma_v^{1/2}, \mathbf{p})$ is stabilizable in the sense of Definitions A.1.1 and A.1.2, respectively, then there is a unique solution of (5.29) and $\lim_{k\to\infty} Q_i^*(k) = \bar{Q}_i$.

146

Mimicking Theorem A.12 of [11], it can be shown that a numerical approach to solve (5.29) (if it has a solution) can be obtained by (numerically) solving the following convex programming optimization problem

$$\max \quad tr\left(\sum_{i=1}^{N} Q_i\right) \qquad (5.30)$$

$$\begin{pmatrix} -Q_i + \sum_{j=1}^{N} p_{ij} A Q_j A' + \Sigma_w & \sqrt{p_{i1}} C_1 Q_1 A' & \dots & \sqrt{p_{i1}} C_N Q_N A' \\ \sqrt{p_{i1}} A Q_1 C_1' & \Sigma_{v_1} + C_1 Q_1 C_1' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{p_{iN}} A Q_N C_N' & 0 & \dots & \Sigma_{v_N} + C_N Q_N C_N' \end{pmatrix} \geq 0$$

$$Q_i > 0, \ i = 1 \dots N.$$

5.5 Connection with Markovian Jump Linear System state estimation

In this section we present a connection between the detectability of (5.2) in the sense of Definition 5.2.1 and the detectability property of a MJLS, which is defined in what follows. We also show that the optimal gains of a linear filter for the state estimation of the aforementioned MJLS can be used to approximate the solution of the optimization problem (5.11), which gives the optimal CBDLF. We assume that the matrix $P(k)$ describing the communication topology of the sensors is *irreducible* and *doubly stochastic* and we assume, without loss of generality, that the matrices $\{C_i(k), k \geq 0\}_{i=1}^{N}$ in the sensing model (5.3), have the same dimensions. We define the following Markovian jump linear

system

$$\xi(k+1) = \tilde{A}_{\theta(k)}(k)\xi(k) + \tilde{B}_{\theta(k)}(k)\tilde{w}(k)$$

$$z(k) = \tilde{C}_{\theta(k)}(k)\xi(k) + \tilde{D}_{\theta(k)}(k)\tilde{v}(k), \ \xi(0) = \xi_0,$$

(5.31)

where $\xi(k)$ is the state, $z(k)$ is the output, $\theta(k) \in \{1, \ldots, N\}$ is a Markov chain with probability transition matrix $P(k)'$, $\tilde{w}(k)$ and $\tilde{v}(k)$ are independent Gaussian random variables with zero mean and identity covariance matrices. Also, ξ_0 is a Gaussian noise with mean μ_0 and covariance matrix Σ_0. We denote by $\pi_i(k)$ the probability distribution of $\theta(k)$ ($Pr(\theta(k) = i) = \pi_i(k)$) and we assume that $\pi_i(0) > 0$. We have that $\tilde{A}_{\theta(k)}(k) \in \{\tilde{A}_i(k)\}_{i=1}^N$, $\tilde{B}_{\theta(k)}(k) \in \{\tilde{B}_i(k)\}_{i=1}^N$, $\tilde{C}_{\theta(k)}(k) \in \{\tilde{C}_i(k)\}_{i=1}^N$ and $\tilde{D}_{\theta(k)}(k) \in \{\tilde{D}_i(k)\}_{i=1}^N$, where the index i refers to the state i of $\theta(k)$. We set

$$\tilde{A}_i(k) = A(k), \qquad \tilde{B}_i(k) = \frac{\sqrt{\pi_i(0)}}{\sqrt{\pi_i(k)}}\Sigma_w^{1/2}(k),$$

$$\tilde{C}_i(k) = \frac{1}{\sqrt{\pi_i(0)}}C_i(k), \quad \tilde{D}_i(k) = \frac{1}{\sqrt{\pi_i(k)}}\Sigma_{v_i}^{1/2}(k),$$

(5.32)

for all $i, k \geq 0$ (note that since $P(k)$ is assumed doubly stochastic and irreducible and $\pi_i(0) > 0$, we have that $\pi_i(k) > 0$ for all $i, k \geq 0$). In addition, ξ_0, $\theta(k)$, $\tilde{w}(k)$ and $\tilde{v}(k)$ are assumed independent for all $k \geq 0$. The random process $\theta(k)$ is also called *mode*. Assuming that the mode is directly observed, a linear filter for the state estimation is given by

$$\hat{\xi}(k+1) = \tilde{A}_{\theta(k)}(k)\hat{\xi}(k) + M_{\theta(k)}(k)(z(k) - \tilde{C}_{\theta(k)}(k)\hat{\xi}(k)),$$

(5.33)

where we assume that the filter gain $M_{\theta(k)}$ depends only on the current mode. The dynamics of the estimation error $e(k) \triangleq \xi(k) - \hat{\xi}(k)$ is given by

$$e(k+1) = \left(\tilde{A}_{\theta k}(k) - M_{\theta(k)}(k)\tilde{C}_{\theta(k)}(k)\right)e(k) +$$

$$+ \tilde{B}_{\theta(k)}(k)w(k) - M_{\theta(k)}(k)\tilde{D}_{\theta(k)}(k)v(k).$$

(5.34)

Let $\mu(k)$ and $Y(k)$ denote the mean and the covariance matrix of $e(k)$, i.e. $\mu(k) \triangleq$ $E[e(k)]$ and $Y(k) \triangleq E[e(k)e(k)']$, respectively. We define also the mean and the covariance matrix of $e(k)$, when the system is in mode i, i.e. $\mu_i(k) \triangleq E[e(k)\mathbb{1}_{\{\theta(k)=i\}}]$ and $Y_i(k) \triangleq$ $E[e(k)e(k)'\mathbb{1}_{\{\theta(k)=i\}}]$, where $\mathbb{1}_{\{\theta(k)=i\}}$ is the indicator function. It follows immediately that $\mu(k) = \sum_{i=1}^{N} \mu_i(k)$ and $Y(k) = \sum_{i=1}^{N} Y_i(k)$.

Definition 5.5.1. *The optimal linear filter (5.33) is obtain by minimizing the following quadratic finite horizon cost function*

$$\tilde{J}_K(\mathbf{M}(\cdot)) = \sum_{k=1}^{K} tr(Y(k)) = \sum_{k=1}^{K} \sum_{i=1}^{N} tr(Y_i(k)), \tag{5.35}$$

where $\mathbf{M}(\cdot) \triangleq \{M_i(k), k = 0 \ldots K - 1\}_{i=1}^{N}$ are the filter gains and where $M_i(k)$ corresponds to $M_{\theta(k)}(k)$ when $\theta(k)$ is in mode i. We can give a similar definition for an optimal steady state filter using the infinite horizon quadratic cost function.

Definition 5.5.2. *Assume that the matrices $\tilde{A}_i(k)$, $\tilde{C}_i(k)$ and $P(k)$ are constant for all $k \geq 0$. We say that the Markovian jump linear system (5.31) is mean square detectable if there exits $\{M_i\}_{i=1}^{N}$ such that $\lim_{k \to \infty} E[\|e(k)\|^2] = 0$, when the noises $\tilde{w}(k)$ and $\tilde{v}(k)$ are set to zero.*

The next result makes the connection between the detectability of the MJLS defined above and the distributed detectability of the process (5.2).

Proposition 5.5.1. *If the Markovian jump linear system (5.31) is mean square detectable, then the linear stochastic system (5.2)-(5.3) is detectable in the sense of Definition 5.2.1.*

Proof. In the context of this proposition, the dynamics of the estimation error for the

MJLS (5.31) becomes

$$e(k+1) = (A - M_{\theta(k)}\tilde{C}_{\theta(k)})e(k), \ e(0) = e_0,$$

where $\tilde{C}_i = C_i$. It is not difficult to check that the dynamic equations for the covariance matrices $\{Y_i(k)\}_{i=1}^N$ and the mean vectors $\{\mu_i(k)\}_{i=1}^N$ are given by

$$Y_i(k+1) = \sum_{j=1}^N p_{ij}(A - M_j \frac{1}{\sqrt{\pi_i(0)}}C_j)Y_j(k)(A - M_j \frac{1}{\sqrt{\pi_i(0)}}C_j)', \qquad (5.36)$$

with $Y_i(0) = Y_i^0$ and

$$\mu_i(k+1) = \sum_{j=1}^N p_{ij}(A - M_j \frac{1}{\sqrt{\pi_i(0)}}C_j)\mu_j(k), \mu_i(0) = \mu_i^0, \qquad (5.37)$$

for $i = 1\ldots N$. Since the MJLS is assumed mean square detectable it follows that there exists a set of matrices $\{M_i\}_{i=1}^N$ such that (5.36) is asymptotically stable. But this also implies (see for instance Proposition 3.6 of [11]) that (5.37) is asymptotically stable as well. Setting $L_i = \pi_i(0)M_i$, we see that (5.37) is identical to equation (5.7) and therefore (5.7) is asymptotically stable (when ignoring the noise). Hence, (5.2) is detectable in the sense of Definition 5.2.1. □

The next result establishes that the optimal gains of the filter (5.33) can be used to approximate the solution of the optimzation problem (5.11).

Proposition 5.5.2. *Let $M^*(\cdot) \triangleq \{M_i^*(k), k = 0,\ldots,K-1\}_{i=1}^N$ be the optimal gains of the linear filter (5.33). If we set $L_i(k) = \frac{1}{\sqrt{\pi_i(0)}}M_i^*(k)$ as filtering gains in the CBDLF scheme, then the filter cost function (5.10) is guaranteed to be upper bounded by*

$$J_K(\boldsymbol{L}(\cdot)) \leq \sum_{k=0}^K \sum_{i=1}^N \frac{1}{\pi_i(0)} tr(Y_i^*(k)), \qquad (5.38)$$

where $Y_i^*(k)$ are the covariance matrices resulting from minimizing (5.35).

Proof. By Theorem 5.5 of [11], the filtering gains that minimize (5.35) are given by

$$M_i^*(k) = \tilde{A}_i(k)Y_i^*(k)\tilde{C}_i(k)' \left[\pi_i(k)\tilde{D}_j(k)\tilde{D}_j(k)' + \tilde{C}_i(k)Y_i^*(k)\tilde{C}_i(k)' \right]^{-1}, \qquad (5.39)$$

for $i = 1 \ldots N$, where $Y_i^*(k)$ satisfies

$$Y_i^*(k+1) = \sum_{j=1}^N p_{ij}(k) \Big[\tilde{A}_j(k)Y_j^*(k)\tilde{A}_j(k)' + \pi_j(k)\tilde{B}_j(k)\tilde{B}_j(k)' -$$
$$-\tilde{A}_j(k)Y_j^*(k)\tilde{C}_j(k)' \left(\pi_j(k)\tilde{D}_j(k)\tilde{D}_j(k)' + \tilde{C}_j(k)Y_j^*(k)\tilde{C}_j(k)' \right)^{-1} \tilde{C}_j(k)Y_j^*(k)\tilde{A}_j(k)' \Big]. \qquad (5.40)$$

In what follows we will show by induction that $Y_i^*(k) = \pi_i(0)Q_i^*(k)$ for all $i, k \geq 0$, where $Q_i^*(k)$ satisfies (5.26). For $k = 0$ we have $Y_i^*(0) = \pi_i(0)Y^*(0) = \pi_i(0)\Sigma_0 = \pi_i(0)Q_i^*(0)$. Let us assume that $Y_i^*(k) = \pi_i(0)Q_i^*(k)$. Then, from (5.32) we have

$$\pi_j(k)\tilde{B}_j(k)\tilde{B}_j(k)' = \pi_i(0)\Sigma_w(k), \quad \pi_j(k)\tilde{D}_j(k)\tilde{D}_j(k)' = \Sigma_{v_i}(k),$$
$$\pi_j(k)\tilde{D}_j(k)\tilde{D}_j(k)' + \tilde{C}_j(k)Y_j^*(k)\tilde{C}_j(k)' = \Sigma_{v_j}(k) + C_j(k)Q_j^*(k)C_j(k)'. \qquad (5.41)$$

Also,

$$M_i^*(k) = \pi_i(0)A(k)Q_i^*(k)C_i(k)' \left[\Sigma_{v_j}(k) + C_j(k)Q_j^*(k)C_j(k)' \right]^{-1}, \qquad (5.42)$$

and from (5.25) we get that $M_i^*(k) = \sqrt{\pi_i(0)}L_i^*(k)$. From (5.40) and (5.41) it can be easily argued that $Y_i^*(k+1) = \pi_i(0)Q_i^*(k+1)$. By Corollary 5.4.1 we have that

$$J_K(\mathbf{L}(\cdot)) \leq \bar{J}_K(\mathbf{L}(\cdot)),$$

for any set of filtering gains $\mathbf{L}(\cdot)$ and in particular for $L_i(k) = \frac{1}{\pi_i(0)}M_i^*(k) = L_i^*(k)$, for all i and k. But since

$$\bar{J}_K(\mathbf{L}^*(\cdot)) = \sum_{k=0}^K \sum_{i=1}^N \frac{1}{\pi_i(0)}Y_i^*(k),$$

the result follows.

$\square$

151

Chapter 6

Conclusions

In Chapter 2 we studied a multi-agent subgradient method under random communication topology. Under an i.i.d. assumption on the random process governing the evolution of the topology, we derived upper bounds on two performance metrics related to the CBMASM. The first metric reflects how close each agent can get to the optimal value. The second metric reflects how close and fast the agents' estimates of the decision vector can get to the minimizer of the objective function, and it was analyzed for a particular class of convex functions. All the aforementioned performance measures were expressed in terms of the probability distribution of the random communication topology. In addition we showed how the distributed optimization algorithm can be used to perform collaborative system identification, application which can be useful in collaborative tracking

In Chapter 3 we emphasized the importance of the convexity concept and in particular the importance of the convex hull notion for reaching consensus. We did this by generalizing the asymptotic consensus problem to the case of convex metric spaces. For a group of agents taking values in a convex metric space, we introduced an iterative algorithm which ensures asymptotic convergence to agreement under some minimal assumptions for the communication graph. As an application, we provided an iterative algorithm which guarantees convergence to consensus of opinion.

In Chapter 4 we analyzed the convergence properties of the linear consensus prob-

lem, when the communication topology is modeled as a directed random graph with an underlying Markovian process. We addressed both the cases where the dynamics of the agents are expressed in continuous and discrete time. Under some assumptions on the communication topologies, we provided a rigorous mathematical proof for the intuitive necessary and sufficient conditions for reaching average consensus in the mean square and almost sure sense. These conditions are expressed in terms of connectivity properties of the union of graphs corresponding to the states of the Markov process. The aim of this work has been to show how mathematical techniques from the stability theory of the Markovian jump systems, in conjunction with results from the matrix and graph theory can be used to prove convergence results for consensus problems under a stochastic framework.

In Chapter 5 we first provided (testable) sufficient conditions under which stable consensus-based distributed linear filters can be obtained. Second, we gave a sub-optimal, linear filtering scheme, which can be implemented in a distributed manner and is valid for time varying communication topologies as well, and which guarantees a quantifiable level of performance. Third, under the assumption that the stochastic matrix used in the consensus step is doubly stochastic we showed that if an appropriately defined Markovian jump linear system is detectable, then the stochastic process of our interest is detectable as well. We also showed that the optimal gains of the consensus-based distributed linear filter scheme can be approximated by using the optimal linear filter for the state estimation of a particular Markovian jump linear system.

As future directions, an immediate extension of the results of Chapter 2 is the generalization of the convergence analysis to case where the communication topology is mod-

eled by a Markovian random graph. The results introduced in Chapter 4 provide the appropriate framework to this end. In Chapter 5 we proposed a distributed algorithm for the state estimation of a process observed by a network of sensors. When considering wireless networks, another relevant problem is designing network architectures aimed at ensuring good estimation performance and network longevity. The problem increases in complexity if we impose the solution to be obtained in a distributed manner. Due to the communication costs inherent to a wireless network, the network architecture should be a result of a tradeoff between the need for rich communication neighborhoods for obtaining accurate and stable estimates and the need for small communication neighborhoods for energy conservation. Our approach will consist in formulating the network architecture design problem as a constraint optimization problem which is solved in a distributed manner by the sensors. The main cost should reflect the relevance of the sensor measurements for the estimation process, while the constraints should reflect the limited energy available for communication and the need to ensure rich enough local neighborhoods for computing the state estimates.

As we showed in Chapters 2 and 5, the consensus problem represents a tool for localizing algorithms in distributed computing. Important optimization problems go beyond the realm of $\mathbb{R}^n$. For example, as we have mentioned in the introduction chapter,the trusted routing problem is formulated on the *Max-plus semiring*, while the design of network topology can be formulated on a *Hamming space*. We plan to continue the analysis started in Chapter 3, and formulate the consensus problem on semirings, and in particular on the *Max-plus* algebra. One of our goals is to explore the feasibility of using consensus to localize the algorithms used for solving optimization problems on spaces where the

operations and relations are described by the *Max-plus* algebra, for example. A simple model for a graph link is obtained by assigning to the link a boolean value. By stacking all possible links, we obtain a vector whose entries can take zero/one values (corresponding to the existence or non-existence of links), and which lives in a Hamming space. As we have previously commented, designing communication topologies is an important problem in distributed optimization, estimation and control applications, in particular in the case of wireless networks for which usually the resources are scarce. Another goal of ours is to study the possibility of using the consensus problem formulated on Hamming spaces for solving distributed optimization problems whose result should provide a network architecture, specifically designed for a particular task, such as estimation or optimization.

Appendix A

Discrete-Time Coupled Matrix Equations

A.1   Properties of a special class of difference matrix equations

Given a positive integer N, a sequence of positive numbers $\mathbf{p} = \{p_{ij}\}_{i,j=1}^{N}$ and a set of matrices $\mathbf{F} = \{F_i\}_{i=1}^{N}$, we consider the following matrix difference equations

$$W_i(k+1) = \sum_{j=1}^{N} p_{ij} F_j W_j(k) F_j', \ W_i(0) = W_i^0, \ i = 1 \ldots N. \tag{A.1}$$

Additionally, consider a similar set of matrix difference equations

$$W_i(k+1) = \sum_{j=1}^{N} p_{ji} F_j' W_j(k) F_j, \ W_i(0) = W_i^0, \ i = 1 \ldots N. \tag{A.2}$$

Proposition A.1.1. *[9] The dynamics (A.1)are asymptotically stable if and only if the dynamics (A.2) are asymptotically stable.*

Related to the above dynamic equations, we introduce the following stabilizability and detectability definitions.

Definition A.1.1. *[10] Given a set of matrices $\boldsymbol{C} = \{C_i\}_{i=1}^{N}$, we say that $(\boldsymbol{p}, \boldsymbol{L}, \boldsymbol{A})$ is detectable if there exists a set of matrices $\boldsymbol{L} = \{L_i\}_{i=1}^{N}$ such that the dynamics (A.1) is asymptotically stable, where $F_i = A_i - L_i C_i$, for $i = 1 \ldots N$.*

Definition A.1.2. *[10] Given a set of matrices $\boldsymbol{C} = \{C_i\}_{i=1}^{N}$, we say that $(\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{p})$ is stabilizable, if there exists a set of matrices $\boldsymbol{L} = \{L_i\}_{i=1}^{N}$ such that the dynamics (A.1) is asymptotically stable, where $F_i = A_i - C_i L_i$, for $i = 1 \ldots N$.*

Remark A.1.1. *Given a semipositive definite matrix X and a positive definite matrix Y, the following holds:*

$$\min_{i=1,\ldots,n} \lambda_i(Y) tr(X) \leq tr(YX) \leq \max_{i=1,\ldots,n} \lambda_i(Y) tr(X)$$

Proposition A.1.2. *If there exists a set of symmetric positive definite matrices $\{V_i\}_{i=1}^{N}$ such that*

$$V_i = \sum_{j=1}^{N} p_{ji} F'_i V_j F_i + S_i, \tag{A.3}$$

for some set of symmetric positive definite matrices $\{S_i\}_{i=1}^{N}$, then the dynamics (A.1) are asymptotically stable.

Proof. We use the same idea as in the proof of Theorem 3.19 of [11] and define the following Lyapunov function

$$\Phi(k) = \sum_{i=1}^{N} tr(W_i(k) V_i).$$

In the following we show that the difference $\Phi(k+1) - \Phi(k)$ is negative for all $k \geq 0$, from which we infer the asymptotic stability of (A.1). We get that

$$\Phi(k+1) - \Phi(k) = tr\left(\sum_{i=1}^{N}\left(\sum_{j=1}^{N} p_{ij} F_j W_j(k) F'_j\right) V_i - W_i(k) V_i\right) =$$

$$= tr\left(\sum_{i=1}^{N} W_i(k) \left(\sum_{j=1}^{N} p_{ji} F_i V_j(k) F'_i - V_i\right)\right) = \sum_{i=1}^{N} tr(W_i(k) S_i).$$

Since $\{W_i(k)\}_{i=1}^{N}$ are positive semi-definite matrices for $k \geq 0$ and $\{S_i\}_{i=1}^{N}$ are positive definite, by Remark A.1.1, it follows that

$$\Phi(k+1) - \Phi(k) < 0, \ k \geq 0.$$

$\square$

Proposition A.1.3. *If there exists a set of symmetric positive definite matrices $\{V_i\}_{i=1}^{N}$ such that*

$$V_i = \sum_{j=1}^{N} p_{ij} F_i' V_j F_i + S_i, \tag{A.4}$$

for some set of symmetric positive definite matrices $\{S_i\}_{i=1}^{N}$, then the dynamics (A.1) are asymptotically stable.

Proof. Using the same approach as in the previous proposition, we prove the asymptotic stability of the dynamics (A.2). Using Proposition A.1.1, the result follows. □

Proposition A.1.4. *If the following linear matrix inequalities are feasible*

$$\begin{pmatrix} X_i & \sqrt{p_{1i}} X_i F_i & \sqrt{p_{2i}} F_i' X_i & \cdots & \sqrt{p_{Ni}} F_i' X_i \\ \sqrt{p_{1i}} X_i F_i & X_1 & 0 & \cdots & 0 \\ \sqrt{p_{2i}} X_i F_i & 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sqrt{p_{Ni}} X_i F_i & 0 & 0 & \cdots & X_N \end{pmatrix} > 0, \tag{A.5}$$

for $i = 1 \ldots N$, where $\{X_i\}_{i=1}^{N}$ are the unknown variables, then the dynamics (A.1) are asymptotically stable.

Proof. By the Schur complement lemma, (A.5) are feasible if and only if

$$X_i - \sum_{j=1}^{N} p_{ji} X_i F_i X_j^{-1} F_i' X_i > 0, \; X_i > 0, \; i = 1 \ldots N. \tag{A.6}$$

By defining $V_i \triangleq X_i^{-1}$, $i = 1 \ldots N$, (A.6), becomes

$$V_i - \sum_{j=1}^{N} p_{ji} F_i V_j F_i' > 0, \; V_i > 0, \; i = 1 \ldots N.$$

By Proposition A.1.2, (A.1) is asymptotically stable. □

158

Inspired by Proposition A.1.4, detectability and stabilizability tests, in the sense of Definitions A.1.1 and A.1.2, respectively, can be formulated in terms of the feasibility of a set of linear matrix inequalities.

Proposition A.1.5 (detectability test)**.** *If the following matrix inequalities are feasible*

$$
\begin{pmatrix}
X_i & \sqrt{p_{i1}}(X_iA_i - Y_iC_i) & \sqrt{p_{i2}}(X_iA_i - Y_iC_i) & \cdots & \sqrt{p_{iN}}(X_iA_i - Y_iC_i) \\
\sqrt{p_{i1}}(X_iA_i - Y_iC_i)' & X_1 & 0 & \cdots & 0 \\
\sqrt{p_{i2}}(X_iA_i - Y_iC_i)' & 0 & X_2 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\sqrt{p_{iN}}(X_iA_i - Y_iC_i)' & 0 & 0 & \cdots & X_N
\end{pmatrix} > 0,
$$

$$\text{(A.7)}$$

for $i = 1\ldots N$, where $\{X_i\}_{i=1}^N$ and $\{Y_i\}_{i=1}^N$ are the unknown variables, then $(\boldsymbol{p},\boldsymbol{L},\boldsymbol{A})$ is detectable in the sense of Definition A.1.1. Moreover chosing $L_i = X_i^{-1}Y_i$, for $i = 1\ldots N$, the dynamics (A.1) are asymptotically stable.

Proof. By the Schur complement lemma, (A.7) are feasible if and only if

$$
X_i - \sum_{j=1}^N p_{ij}(X_iA_i - Y_iC_i)X_j^{-1}(X_iA_i - Y_iC_i)' > 0, \ X_i > 0, \ i = 1\ldots N. \qquad \text{(A.8)}
$$

By defining $L_i \triangleq X_i^{-1}Y_i$ and $V_i \triangleq X_i^{-1}$, $i = 1\ldots N$, (A.8), becomes

$$
V_i - \sum_{j=1}^N p_{ij}F_iV_jF_i' > 0, \ V_i > 0, \ i = 1\ldots N.
$$

By Proposition A.1.3, $(\boldsymbol{p},\boldsymbol{L},\boldsymbol{A})$ is detectable in the sense of Definition A.1.1. $\qquad\square$

Proposition A.1.6 (stabilizability test)**.** *If the following matrix inequalities are feasible*

$$
\left(
\begin{array}{ccccc}
X_i & \sqrt{p_{1i}}(X_iA_i - C_iY_i)' & \sqrt{p_{2i}}(X_iA_i - C_iY_i)' & \cdots & \sqrt{p_{Ni}}((X_iA_i - C_iY_i)' \\
\sqrt{p_{1i}}(X_iA_i - C_iY_i) & X_1 & 0 & \cdots & 0 \\
\sqrt{p_{2i}}(X_iA_i - C_iY_i) & 0 & X_2 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\sqrt{p_{Ni}}(X_iA_i - C_iY_i) & 0 & 0 & \cdots & X_N
\end{array}
\right) > 0,
$$

(A.9)

for $i = 1\ldots N$, where $\{X_i\}_{i=1}^N$ and $\{Y_i\}_{i=1}^N$ are the unknown variables, then $(\mathbf{A},\mathbf{L},\mathbf{p})$ is stabilizable in the sense of Definition A.1.2. Moreover choosing $L_i = Y_iX_i^{-1}$, for $i = 1\ldots N$, the dynamics (A.1) are asymptotically stable.

Proof. By the Schur complement lemma, (A.9) are feasible if and only if

$$
X_i - \sum_{j=1}^N p_{ji}(X_iA_i - Y_iC_i)'X_j^{-1}(X_iA_i - Y_iC_i) > 0, \ X_i > 0, \ i = 1\ldots N.
$$

(A.10)

By defining $L_i \triangleq X_i^{-1}Y_i$ and $V_i \triangleq X_i^{-1}$, $i = 1\ldots N$, (A.10), becomes

$$
V_i - \sum_{j=1}^N p_{ji}F_i'V_jF_i > 0, \ V_i > 0, \ i = 1\ldots N.
$$

By Proposition A.1.2, $(\mathbf{p},\mathbf{L},\mathbf{A})$ is stabilizable in the sense of Definition A.1.2. $\qquad\square$

A.2   Discrete-time coupled Riccati equations

Consider the following coupled Riccati difference equations

$$
Q_i(k+1) = \sum_{i=1}^N p_{ij}\Big(A_jQ_j(k)A_j' - A_jQ_j(k)C_j'(C_jQ_j(k)C_j' + \Sigma_{v_j})^{-1}C_jQ_j(k)A_j' + \Sigma_w\Big),
$$

(A.11)

$Q_i(0) = Q_i^0 > 0$, $i = 1\ldots N$, where $\{\Sigma_{v_i}\}_{i=1}^N$ and Σ_w are symmetric positive definite matrices.

Proposition A.2.1. *Let* $\Sigma_v^{1/2} = \{\Sigma_{v_i}^{1/2}\}_{i=1}^N$, *where* $\Sigma_{v_i} = \Sigma_{v_i}^{1/2'}\Sigma_{v_i}^{1/2}$. *Suppose that* $(\boldsymbol{p}, \boldsymbol{C}, \boldsymbol{A})$ *is*

detectable and that $(\boldsymbol{A}, \Sigma_v^{1/2}, \boldsymbol{p})$ *is stabilizable in the sense of Definitions A.1.1 and A.1.2,*

respectively. Then there exists a unique set of symmetric positive definite matrices $\bar{\boldsymbol{Q}} =$

$\{\bar{Q}_i\}_{i=1}^N$ *satisfying*

$$\bar{Q}_i = \sum_{i=1}^N p_{ij}\left(A_j\bar{Q}_jA_j' - A_j\bar{Q}_jC_j'(C_j\bar{Q}_jC_j' + \Sigma_{v_j})^{-1}C_j\bar{Q}_jA_j' + \Sigma_w\right), \ i = 1\dots N. \quad \text{(A.12)}$$

Moreover, for any initial conditions $Q_i^0 > 0$, *we have that* $\lim_{k\to\infty} Q_i(k) = \bar{Q}_i$.

Proof. The proof can be mimicked after the proof of Theorem 1 of [10]. Compared to our

case, in Theorem 1 of [10], scalar terms, taking values between zero and one, multiply

the matrices Σ_{v_j} in (A.12). However it is not difficult to note that the result holds even in

the case where these scalar terms take the value one, which corresponds to our setup. □

Bibliography

[1] T. Ayasal, M. Yildiz, A. Sarwate, and A. Scaglione. Broadcast gossip algorithms for consensus. *IEEE Transaction on Signal Processing*, 57(7):2748–2761, July 2009.

[2] M. Rabi B. Johansson and K.H. Johansson. A randomized incremental subgradient method for distributed optimization in networked systems. *SIAM Journal on Optimization*, 20(3):1157–1170, 2009.

[3] J.S. Baras and P. Hovareshti. Effects of topology in networked systems: Stochastic methods and small worlds. *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 2973–2978, Dec 2008.

[4] V.D. Blondel, J.M. Hendrickx, A. Olshevsky, and J.N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 2996–3000, Dec 2005.

[5] V. Borkar and P. Varaya. Asymptotic agreement in distributed estimation. *IEEE Trans. Autom. Control*, AC-27(3):650–655, Jun 1982.

[6] S. Boyd, P. Diaconis, and L. Xiao. Fastest mixing markov chain on a graph. *SIAM Review*, 46(4):667–689, Dec 2004.

[7] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE/ACM Trans. Netw.*, 14(SI):2508–2530, 2006.

[8] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri. Distributed kalman filtering based on consensus strategies. *IEEE Journal on Selected Area in Communication*, 26(4):622–633, May 2008.

[9] O.L.V. Costa and M.D. Fragoso. Stability results for discrete-time linear systems with markovian jumping parameters. *Journal of Mathematical Analysis and Application*, 179:154–178, 1993.

[10] O.L.V. Costa and M.D. Fragoso. Discrete-time coupled riccati equations for systems with markov switching parameters. *Journal of Mathematical Analysis and Application*, 194:197–216, 1995.

[11] O.L.V. Costa, M.D. Fragoso, and R.P. Marques. *Discrete-Time Markov Jump Linear Systems*. Springer-Verlag, London, 2005.

[12] A.G. Dimakis, S. Kar, J.M.F. Moura, M.G. Rabbat, and A. Scaglione. Gossip algorithms for distributed signal processing. *arXiv:1003.5309v1 [cs.DC]*, March 2010.

[13] J.C. Duchi, A. Agarwal, and M.J. Wainwright. Distributed dual averaging in netwroks. *Proceedings of the 2010 Neural Information System Foundation Conference*, December 2010.

[14] J.A. Fax and R.M. Murray. Information flow and cooperative control of vehicles formations. *IEEE Trans. Autom. Control*, 49(9):1456–1476, Sept 2004.

[15] X. Feng and K.A. Loparo. Stability of linear markovian systems. *Proceedings of the 29th IEEE Conference on Decision and Control*, pages 1408–1413, Dec 1990.

[16] M.D. Fragoso and O.L.V. Costa. A unified approach for stochastic and mean-square stability of continuous-time linear systems with markovian jumping parameters and additive disturbances. *SIAM Journal on Control and Optimization*, 44(4):1165–1191, 2005.

[17] Y. Hatano and M. Mesbahi. Agreement over random networks. *IEEE Trans. Autom. Control*, 50(11):1867–1872, 2005.

[18] A. Jadbabaie, J. Lin, and A.S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor. *IEEE Trans. Autom. Control*, 48(6):998–1001, Jun 2004.

[19] B. Johansson. On distributed optimization in network systems. *Ph.D. dissertation in Telecommunication*, 2008.

[20] B. Johansson, T. Keviczky, M. Johansson, and K.H. Johansson. Subgradient methods and consensus algorithms for solving convex optimization problems. *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 4185–4190, Dec 2008.

[21] S. Kandukuri and S. Boyd. Optimal power control in interference-limited fading wireless channels with outage-probability specifications. *IEEE Transactions on Wireless Communications*, 1(1), Jan 2002.

[22] I. Lobel and A. Ozdalgar. Convergence analysis of distributed subgradient methods over random networks. *Proceedings of 46th Allerton Conference on Communication, Control, and Computing*, pages 353–360, Sept 2008.

[23] I. Matei, N. Martins, and J. Baras. Almost sure convergence to consensus in markovian random graphs. *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 3535–3540, Dec 2008.

[24] I. Matei, N. Martins, and J. Baras. Optimal state estimation for discrete-time markovian jump linear systems, in the presence of delayed mode observations. *Proceedings of the 2008 IEEE American Control Conference*, pages 3560–3565, Jun 2008.

[25] I. Matei, N. Martins, and J. Baras. Optimal state estimation for discrete-time markovian jump linear systems, in the presence of delayed output observations. *Proceedings of the IEEE 2008 Information Theory Workshop*, pages 237–242, May 2008.

[26] I. Matei, N. Martins, and J. Baras. Consensus problems with directed markovian communication patterns. *Proceedings of the 2009 IEEE American Control Conference*, pages 1298–1303, Jun 2009.

[27] A. Moraglio, C. Di Chio, J. Togelius, and R. Poli. Geometric particle swarm optimization. *Journal of Artificial Evolution and Applications, Article ID 143624, 14 pages*, 2008.

[28] A. Moraglio and J. Togelius. Geometric particle swarm optimization for the sudoku puzzle. *Proceedings of the 9th Conference on Genetic and Evolutionary Computation*, pages 118–125, 2007.

[29] L. Moreau. Stability of multi-agents systems with time-dependent communication links. *IEEE Trans. Autom. Control*, 50(2):169–182, Feb 2005.

[30] A. Nedic and D. Bertsekas. Convergence rate of incremental subgradient algorithm. *Stochastic Optimization: Algorithms and Applications*, pages 263–304, 2000.

[31] A. Nedic and D. Bertsekas. Incremental subgradient methods for nondifferential optimization. *SIAM Journal on Optimization*, 12(1):109–138, 2001.

[32] A. Nedic and A. Ozdaglar. Convergence rate for consensus with delays. *Journal of Global Optimization*, 47(3):437–456, Jul 2010.

[33] A. Nedic, A. Ozdaglar, and P.A. Parrilo. Constrained consensus and optimization in multi-agent networks. *IEEE Trans. Autom. Control*, 55(4):922–938, Apr 2010.

[34] A. Nedic and A. Ozdalgar. Distributed subgradient methods for multi-agent optimization. *IEEE Trans. Autom. Control*, 54(1):48–61, Jan 2009.

[35] C.T.K. Ng, M. Medard, and A. Ozdaglar. Completion time minimization and robust power control in wireless packet networks. *arXiv:0812.3447v1 [cs.IT]*, Dec 2008.

[36] A. Olshevsky and J.N. Tsitsiklis. Convergence speed in distributed consensus and averaging. *SIAM Journal on Control and Optimization*, 48(1):33–55, Sept 2009.

[37] B.T. Polyak. *Introduction to Optimization*. Optimization Software, Inc, Publications Division, New York, 1987.

[38] M. Porfiri and D.J. Stilwell. Consensus seeking over random directed weighted graphs. *IEEE Trans. Autom. Control*, 52(9):1767–1773, Sept 2007.

[39] W. Ren and R.W. Beard. Consensus seeking in multi-agents systems under dynamically changing interaction topologies. *IEEE Trans. Autom. Control*, 50(5):655–661, May 2005.

[40] R.O. Saber. Distributed kalman filter with embedded consensus filters. *Proceedings of the 44th IEEE Conference on Decision and Control*, 2005.

[41] R.O. Saber. Distributed kalman filtering for sensor networks. *Proceedings of the 46th IEEE Conference on Decision and Control*, pages 5492–5498, 2007.

[42] R.O. Saber and R.M. Murray. Consensus protocols for networks of dynamic agents. *Proceedings of the 2003 IEEE American Control Conference*, pages 951–956, 2003.

[43] R.O. Saber and R.M. Murray. Consensus problem in networks of agents with switching topology and time-delays. *IEEE Trans. Autom. Control*, 49(9):1520–1533, Sept 2004.

[44] A. Tahbaz Salehi and A. Jadbabaie. Necessary and sufficient conditions for consensus over random networks. *IEEE Trans. Autom. Control*, 53(3):791–795, Apr 2008.

[45] A. Tahbaz Salehi and A. Jadbabaie. Consensus over ergodic stationary graph processes. *IEEE Trans. Autom. Control*, 55(1):225–230, Jan 2010.

[46] B.K. Sharma and C.L. Dewangan. Fixed point theorem in convex metric space. *Novi Sad Journal of Mathematics*, 25(1):9–18, 1995.

[47] K.K. Somasundaram and J.S. Baras. *Path Optimization Techniques for Trusted Routing in Tactical Mobile Ad-Hoc Networks*. ISR Technical Report 2009-3, 2009.

[48] A. Speranzon, C. Fischione, K. H. Johansson, and A. Sangiovanni-Vincentelli. A distributed minimum variance estimator for sensor networks. *IEEE Journal on Select Areas in Communication*, 26(4):609–621, May 2008.

[49] W. Takahashi. A convexity in metric mpace and non-expansive mappings i. *Kodai Math. Sem. Rep.*, 22:142–149, 1970.

[50] D. Teneketzis and P. Varaiya. Consensus in distributed estimation. *Advances in Statistical Signal Processing*, pages 361–386, Jan 1988.

[51] J.N. Tsitsiklis. Problems in decentralized decision making and computation. *Ph.D. dissertation, Dept. Electr. Eng.*, Nov 1984.

[52] J.N. Tsitsiklis, D.P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Trans. Autom. Control*, 31(9):803–812, Sept 1986.

[53] D.J. Watts and S.H Strogatz. Collective dynamics of small world networks. *Nature*, 393:440–442, 1998.

[54] J. Wolfowitz. Product of indecomposable aperiodic stochastic matrices. *Proc. Am. Math. Soc.*, 15:733–736, 1963.