CONVERGENCE RATE ANALYSIS AND OPTIMIZATION OF DISTRIBUTED CONSENSUS ALGORITHMS

by

Onur Cihan

B.S., Control Engineering, İstanbul Technical University, 2007

B.S., Electrical Engineering, İstanbul Technical University, 2007

M.S., Electrical and Electronics Engineering, Boğaziçi University, 2009

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Electrical and Electronics Engineering

Boğaziçi University

2014

ACKNOWLEDGEMENTS

Undertaking this Ph.D. has been a truly life–changing experience for me and it would not have been possible to do without the support and guidance of a great many people. I still offer my deepest thanks to those individuals I neglect to mention here by name.

First and foremost, I would like to thank my thesis supervisor Assoc. Prof. Mehmet Akar for his invaluable guidance, patience, motivation, advice and contributions throughout this study. Without his support, it would be impossible for me to complete this thesis and it was really a privilege and pleasure for me to work with him.

I would like to thank Prof. Yağmur Denizhan and Asst. Prof. Gülay Öke for serving as a member of my thesis committee and their academic guidance. I am also grateful to Prof. Kadri Özçaldıran and Prof. Veysel Gazi for kindly taking part in my thesis jury.

I would also like to extend my sincerest gratitude to my colleagues, Özlem Feyza Erkan, Huzeyfe Esen, Ali Haseltalab and the former Networked and Embedded Control System Laboratory (NECSLAB) members. I feel very lucky to have worked with these wonderful people who made the university a warm place for me.

Last but not the least, there is no word to describe my gratitude toward my parents, Sevil and Esat Cihan and my brother Ahmet Cihan for their endless support and love during my life.

This thesis is partially supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under BİDEB 2211 (National Ph.D. Scholarship Program).

ABSTRACT

CONVERGENCE RATE ANALYSIS AND OPTIMIZATION OF DISTRIBUTED CONSENSUS ALGORITHMS

The problem of achieving a common value in a distributed information sharing context, referred to as distributed consensus or agreement, is an important topic that has drawn significant research attention of late. Consensus algorithms find applications in many areas including network clock synchronization, sensor fusion and load balancing where achieving consensus as fast as possible is important. In this dissertation, we study the analysis and optimization of convergence rate of averaging based distributed consensus algorithms evolving on graphs. By relating the convergence speed of the algorithm to the mixing rate of a Markov chain, we propose two semi–definite programming (SDP) methods of assigning transition probabilities to a Markov chain in order to optimize its mixing rate. In the first SDP formulation, there is a single transition probability parameter to be optimized (the holding probability of vertices) which leads to easier and faster computation as opposed to the more general reversible Markov chain formulation corresponding to a stationary distribution that is proportional to the degree of vertices. By deriving exact analytical results, it is shown that both the single parameter and the degree proportional reversible fastest mixing Markov chain formulations yield better results than the symmetric SDP formulation for a path and some well–known edge–transitive and orbit graphs. The convergence rate of the averaging based distributed consensus algorithm is also analyzed for networks where delay exists in data receptions, which is unavoidable in practice. After introducing the delayed version of the consensus algorithm, it is analytically shown that bounded non–uniform delay does not adversely affect its convergence rate for directed acyclic graphs.

ÖZET

DAĞITIK ONAYLAŞIM ALGORİTMALARININ YAKINSAMA HIZI ANALİZİ VE EN İYİLEMESİ

Dağıtık onaylaşım ya da anlaşma olarak anılan, dağıtık bilgi paylaşımı bağlamında ortak bir değere ulaşma problemi, son zamanlarda kayda değer araştırma ilgisi çekmiş önemli bir konudur. Onaylaşım algoritmaları, onaylaşıma mümkün olabildiğince en hızlı şekilde ulaşılmasının önemli olduğu, ağlardaki saat eş zamanlaması, algılayıcı tümleştirmesi, yük dengelemesi gibi birçok uygulamada kullanılmaktadır. Bu tezde, çizgeler üzerinde tanımlanan ortalama tabanlı dağıtık onaylaşım algoritmalarının yakınsama hızı analizi ve en iyilemesi konusunda çalışılmıştır. Algoritmanın yakınsama hızı Markov zincirlerinin karışım hızı ile ilişkilendirilerek, bir Markov zincirine geçiş olasılıklarını atayarak karışım hızını en iyilemeyi amaçlayan iki Yarı–Tanımlı Programlama (YTP) yöntemi önerilmiştir. İlk YTP formülasyonunda, en iyilenen tek bir geçiş olasılığı parametresi (köşelerde kalma olasılığı) vardır ve bu, köşe dereceleriyle orantılı bir kalıcı dağılıma karşılık gelen daha genel tersinir Markov zinciri formülasyonuna nazaran daha kolay ve hızlı hesaplama sağlar. Kesin çözümler elde edilerek, hem tek parametreli hem de dereceyle oransal tersinir en hızlı karışan Markov zincirlerinin bazı tanınmış kenar geçişli ve yörünge çizgeleri için bakışımlı YTP formülasyonundan daha iyi sonuçlar sağladığı gösterilmiştir. Ortalama tabanlı dağıtık onaylaşım algoritmalarının yakınsama hızı, pratikte kaçınılmaz olan veri alışlarında gecikmenin olduğu ağlar için de analiz edilmiştir. Onaylaşım algoritmasının gecikmeli sürümünün tanıtılmasından sonra, sınırlandırılmış bir örnek olmayan gecikmenin yönlendirilmiş çevrimsiz çizgeler için yakınsama hızını olumsuz etkilemediği analitik olarak gösterilmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

a	The identical holding probability of vertices
a^*	The optimal identical holding probability of vertices
B_{n_1,n_2}	The Markov chain on a complete bipartite graph with n_1 vertices in the first part and n_2 vertices in the second part when all vertices have the identical holding probabilities
C_n	The Markov chain on a cycle with n vertices when all vertices have the identical holding probabilities
d_k	The difference between the maximum and minimum components of the vector $x(k)$ at time k
$\mathcal{E}$	The set of edges
$\mathcal{F}$	The unitary Fourier matrix
$\mathcal{G}$	An undirected graph
I_n	The identity matrix of size n
$K_{n,n}$	The Markov chain on the $K_n - K_n$ graph when all vertices have the identical holding probabilities
$L_{n,n}$	The Markov chain on a path with equal transition probabilities
L_{n,n_1}	The $n_1 \times n_1$ submatrix of $L_{n,n}$ formed by deleting the last $n - n_1$ rows and columns
n	The number of vertices of a graph
N_i	The number of neighbors of vertex i
P	The transition probability matrix
$P_{n,n}(a)$	The Markov chain on a path with n vertices when all vertices have the identical holding probabilities
$P_{n,n_1}(a)$	The $n_1 \times n_1$ submatrix of $P_{n,n}(a)$ formed by deleting the last $n - n_1$ rows and columns
q_i	The basis vector that has 1 as its i–th component and zeros elsewhere
S_n	The Markov chain on a star with n vertices when all vertices have the identical holding probabilities

S_n^*	The fastest mixing symmetric Markov chain on a path with n vertices
U_n	The Markov chain on a wheel graph with a central hub and a cycle of n nodes that are connected to the hub when all vertices have the identical holding probabilities
$\mathcal{V}$	The set of vertices
w_{ij}	The (i, j)–th element of the matrix W
W	The system matrix of the consensus algorithm
$\hat{W}$	The augmented system matrix
$\mathbb{W}$	The set of system matrices
W_D	The diagonal part of W
x	The state vector
$\hat{x}$	The augmented state vector
ϵ	A chosen consensus accuracy level
$\lambda_i(W)$	The i–th largest eigenvalue of the matrix W
$\mu(W)$	The mixing rate of W
$\mu_i^*(W)$	The optimal mixing rate of the Markov chain W when all vertices have the identical holding probabilities
$\mu_s^*(W)$	The optimal mixing rate of the symmetric Markov chain W
$\mu_r^*(W)$	The optimal mixing rate of the Markov chain W with degree proportional stationary distributions
π_k	The set that consist of all possible k matrix products of a given matrix set
$\pi(k)$	The probability distribution at time k
$\sigma(W)$	The spectrum of W
τ_{ij}	The amount of delay in data transmission between nodes j and i
τ_{max}	The maximum amount of delay in data transmission
$\tau(A)$	The coefficient of ergodicity of A

LIST OF ACRONYMS/ABBREVIATIONS

FMMC	Fastest Mixing Markov Chain
LMI	Linear Matrix Inequality
SDP	Semi Definite Programming
SLEM	Second Largest Eigenvalue Modulus

1. INTRODUCTION

The problem of achieving a common value in a distributed information sharing context, referred to as distributed consensus or agreement, is an important topic that has drawn significant research attention of late [1–25]. There are numerous applications employing a consensus algorithm. In the vehicle formation control problem, there are multiple dynamically decoupled vehicles in the network that are trying to perform a shared task. Each vehicle must decide on its formation with the limited information gathered from its neighbors [19–22]. Rendezvous problem is another application of the consensus algorithm, where agents in the network are required to arrive at a location at the same time [16, 17]. In the flocking problem, the agents in the network move as a group with common speed with predefined distances between agents and without any collision [6–8]. Another application of the consensus algorithm is the clock synchronization problem where the nodes in the network share their local clock values so as to decide on a common notion of time. The clocks can easily drift seconds in a day, unless a proper consensus algorithm is adopted. Furthermore, the clocks may not stay synchronized due to different ticking rates of the clocks although they are initially synchronized. The divergence of the clock values is depicted in Figure 1.1.

The rate of convergence of averaging based distributed consensus algorithms is related to the averaging coefficients of the algorithm and a proper choice is required to provide faster convergence. This problem is studied not only by control and computer engineers, but also by mathematicians under the topic *mixing rates of Markov chains* since the averaging coefficients of the consensus algorithms form a matrix which corresponds to the probability distribution matrix of a Markov chain. Optimization of the mixing rates of Markov chains on different network topologies has been an appealing research area in recent years [26–31]. Computing or bounding the mixing rate of Markov chains is very important in Markov chain Monte Carlo simulation, that finds applications in many fields including statistics, physics, chemistry, biology, and computer science [26].

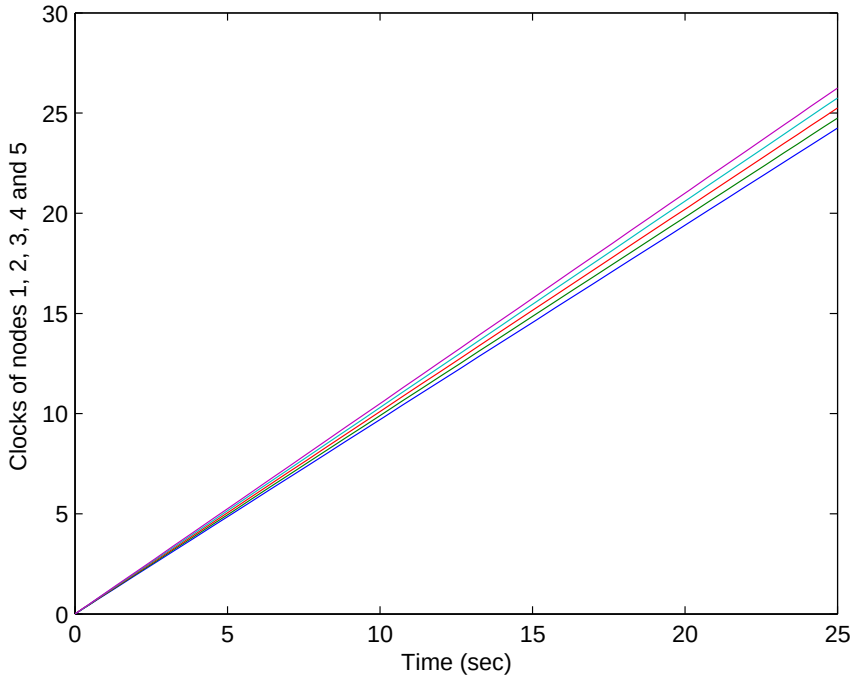


Figure 1.1. Clocks of nodes in the network when no consensus algorithm is used.

Among many applications of the Markov chains is the distributed consensus algorithm implementation where each node updates own value based on the information received from the neighboring nodes in the network so as to decide on a common state. The consensus algorithm is based on a row–stochastic averaging matrix whose elements correspond to the transition probabilities of a Markov chain. In order to achieve the fastest convergence for a given network topology, the optimal averaging coefficients must be determined. Other applications where faster mixing Markov chains are desired are gossip algorithms in sensor networks, and load balancing in computer networks [27].

In [26], the fastest mixing Markov chain (FMMC) problem is cast as convex optimization under the assumption that the Markov chain is symmetric (i.e., the transition probabilities from state j to state i and from i to j are equal to each other). In [27], the FMMC on a path is considered and it is shown that the fastest mixing is achieved when all edges have the transition probability of 1/2. Graph symmetries are exploited in order to reduce the number of variables in the optimization problem in [28]. In [29], the authors derive majorization relations for doubly–stochastic matrices which can be used to solve the FMMC problem. An analytical solution to this problem is given for

networks with K–partite topology in [30]. Some open problems on the mixing rates of Markov chains on acyclic graphs are listed in [31]. Fastest mixing Markov chains on the graphs which can be defined as union of two cliques are investigated in [32].

In most of the above [26–31], the FMMC problem is considered for symmetric matrices. The authors of [26] not only consider a symmetric formulation but also introduce the reversible Markov chain problem where the chain is not necessarily symmetric. In this context, a stationary distribution has to be chosen initially which is subsequently used in formulating a convex Semi Definite Programming (SDP) optimization problem that can be solved by SDP solvers [33, 34]. However, the authors neither provide any analytical results, nor any comparison is done with respect to their own symmetric formulation.

Although there exists some relevant research on the mixing rates of Markov chains that correspond to the convergence speed of consensus algorithms [1, 10], only a few studies were focused on the performance of such algorithms under time delays. In [13], the authors proposed consensus protocols so that agents with integrator dynamics could achieve *average consensus* (i.e., nodes converge to the arithmetic mean of the initial values) under a constant amount of delay. Later, based on the results of [13], the effect of time–varying delays on convergence was examined in [4]. Some sufficient conditions for achieving consensus in directed networks in the presence of non–uniform delays were given in [11]. In [23], the authors used contraction theory and a simplified wave variable design in the stability analysis of interacting nonlinear systems with time delayed communications.

While the above work has been for continuous–time networks, the focus of this thesis is discrete–time consensus which has been addressed by [3, 5, 9, 14, 18, 25]. Conditions for achieving consensus in discrete–time were first given in [18] and were later re–evaluated in [5] for bounded delays and conditions on the connectivity of the network. The results of [5, 18] were extended in [3] by relaxing the convexity of the allowed regions for the state transition map of each agent. In [25], the authors studied the convergence properties of discrete–time consensus algorithm under the assumptions

of bounded delay and connectivity of the network using the concept of spanning tree which was utilized by [14] in proposing a condition for achieving consensus. In [9], the author considered a network of interacting agents and showed that the convergence to a common value would be achieved, provided that old information was uniformly purged from the system. The authors of [12] provided bounds on the time required to reach consensus. Naturally, these bounds turn out to be geometric for the conditions under which consensus is reached.

1.1. Motivation of the Thesis

In most real–time applications, it is desired to achieve consensus as fast as possible, and therefore the parameter selection of distributed consensus algorithm is an important task. For an ordinary Markov chain, the problem of achieving the fastest convergence is not a convex problem. In order to cast it as an Semi Definite Programming problem, the Markov chain must be either symmetric or similar to a symmetric matrix. The authors in [28] derived mathematical expressions of the results of the symmetric formulation for some edge transitive graphs (e.g., cycle, complete bipartite graph), for some distance transitive graphs (e.g., complete graph, Hamming graph, Petersen graph) and for some orbit graphs (e.g., the wheel graph and the graph $K_n - K_n$). However, they do not evaluate the performance of the reversible formulation for a stationary distribution. One of the motivations of this thesis is formulating optimization problems that lead to faster mixing than the the results of the existing symmetric formulation and evaluating its performance by deriving analytical expressions related to the mixing rates. While convergence rates of the consensus algorithm that are related to the mixing rates of Markov chains are extensively investigated in the literature, the possible (non) adverse effect of delay on the performance of averaging based consensus algorithms is not studied. Since communication delay is unavoidable in the networks, this is an important topic to be examined as well.

1.2. The Contributions of the Thesis

The major contributions of the thesis are listed as follows:

- Two optimization problems for fastest mixing Markov chains are proposed throughout the thesis. In the first semi definite programming formulation, there is a single transition probability parameter to be optimized (the holding probability of vertices) which leads to easier and faster computation as opposed to the more general reversible Markov chain formulation corresponding to a stationary distribution that is proportional to the degree of vertices. The solutions of these optimization problems are derived analytically for many important topologies and it is shown that the results of proposed optimization problems tend to outperform the existing solutions in the literature.
- The fastest mixing Markov chain problem is discussed for a path, and two transition probability assignment scenarios (equal weighting and identical holding probability) are examined. The performance of both scenarios are derived as mathematical expressions and compared with the solution of the fastest mixing symmetric Markov chain problem.
- The non–adverse effect of delay on the convergence rate of consensus algorithm is investigated and it is mathematically proven that bounded delay may not necessarily decrease convergence rate for directed acyclic graphs.

1.3. The Organization of the Thesis

The rest of this thesis is organized as follows: The consensus algorithm, some graph theory notions and necessary mathematical definitions are given in Chapter 2.

In Chapter 3, the mixing rates of the Markov chains on undirected graphs which correspond to the convergence speed of the averaging matrix of the consensus algorithm are discussed in Chapter 2. Since the mixing rate of an irreducible and aperiodic Markov chain depends on its second largest eigenvalue modulus, the fastest mixing Markov chain problem can be cast as a convex optimization problem for matrices with real eigenvalues. Two alternative optimization problems of achieving fastest mixing on undirected graphs, namely the fastest mixing reversible Markov chains with degree proportional stationary distributions and the FMMC with identical holding probabilities, are studied and exact analytical results are derived so as to compare their performance

with the existing ones in the literature.

In Chapter 4, the FMMC problem is considered for the path graph, which has many applications in computer science. Two different transition probability assignment scenarios are reviewed and important analytical results in each scenario are obtained. For ease of readability, long proofs are relegated to Appendix.

In Chapter 5, the convergence rate of the consensus algorithm is studied for the networks with delay. After the introduction of the delayed consensus algorithm, the convergence properties are derived not only for networks with fixed topology, but also for varying topology networks. The possible effect of delay on convergence speed in discussed and it is shown that delay does not have detrimental effect on the convergence rate in directed acyclic graphs.

Finally, the concluding remarks are given in Chapter 6.

2. THE AVERAGING BASED DISTRIBUTED CONSENSUS ALGORITHM

In this chapter, we introduce a well known averaging based distributed consensus algorithm and review some graph theory concepts that are utilized to study its convergence properties. We further give the definition of the terms related to the convergence speed of the consensus algorithm that are studied in the rest of the thesis.

2.1. The Consensus Algorithm

In [35], Vicsek *et al.* proposed a discrete-time algorithm for self-driven particle systems where all particles in the plane have the same speed but different headings. A particle updates its heading by using a local rule based on its own heading and the headings of its neighbors so that all particles move towards the same direction as a group. While the original algorithm is not linear, most of the consensus algorithms designed so far are linearized versions of Vicsek's algorithm, one of which is given by

$$\theta_i(t+1) = \frac{1}{1+|N_i(t)|}\left(\theta_i(t) + \sum_{j\in N_i(t)} \theta_j(t)\right). \tag{2.1}$$

where $\theta_i(t)$ is the heading angle of particle i, $N_i(t)$ is the set of neighbors of particle i and $|N_i(t)|$ is the number of neighbors of particle i at time t. According to Vicsek's definition, particle j is a neighbor of particle i if particle j is within a circle of radius r_c surrounding particle j.

Based on Vicsek's algorithm, many authors studied a discrete–time, distributed consensus algorithm which is expressed as

$$x_i(t+1) = \sum_{j=1}^{n} w_{ij}(t)x_j(t) \tag{2.2}$$

where $x_i(t)$ is the local value of node i and $w_{ij}(t)$ is the non–negative weighting coefficient for the information coming from node j to node i. Note that each node uses the above equation to update its value in every step.

For an n–node network, (2.2) can be represented in matrix form as

$$x(t+1) = W(t)x(t) \tag{2.3}$$

where $x(t) = [x_1(t),\ \ldots,\ x_n(t)]^\top$ is the vector containing values of nodes and $W(t)$ is the system matrix which consists of the weighting coefficients $w_{ij}(t)$. Throughout the thesis, it is assumed that the system matrices satisfy the following assumption.

Assumption 2.1. (i) $w_{ij}(t) \geq 0$, if node i receives local information of node j at time t.

(ii) $w_{ij}(t) = 0$, if node i does not receive local information of node j at time t.

(iii) $\sum_{j=1}^{n} w_{ij}(t) = 1$, for all i and t.

Article (i) in Assumption 2.1 is related to network connectivity and ensures that data received from neighboring nodes are used with non-negative weights. Article (ii) in Assumption 2.1 requires that the weighting coefficient is zero when there is no information flow from node j to i and the averaging coefficients sum up to one for each node as stated in Article (iii) in Assumption 2.1.

The objective of the consensus algorithm is to achieve a common value asymptotically, i.e.,

$$\lim_{t\to\infty} x(t) = [c,\ \ldots,\ c]^T \tag{2.4}$$

for all initial conditions $x(0) = x_0$ where c is the common value of nodes.

Definition 2.2. We say that a non–negative matrix $W \in R^{n\times n}$ is *row–stochastic* if $\sum_{j=1}^{n} w_{ij} = 1$, for all $1 \leq i \leq n$.

For row–stochastic matrices, we have the following well–known result due to Markov.

Lemma 2.3. *[36] Let $W = [w_{ij}]$ be a row–stochastic matrix. Let x be a non–negative vector and $y = Wx$. Then we have*

$$\max_{i=1,2,\ldots,n} y_i - \min_{i=1,2,\ldots,n} y_i \leq \tau(W) \left(\max_{i=1,2,\ldots,n} x_i - \min_{i=1,2,\ldots,n} x_i \right) \tag{2.5}$$

where

$$\tau(W) = \frac{1}{2} \max_{i,j} \sum_{k} |w_{ik} - w_{jk}| \tag{2.6}$$

is defined to be the coefficient of ergodicity of W.

Definition 2.4. Let $W \in R^{n \times n}$ be a row–stochastic matrix. (i) We say that W is *scrambling* if $\tau(W) < 1$.
(ii) We say that W is *ergodic* if $\lim_{t \to \infty} W^t = \mathbf{1}c^\top$ for some $c \in \Re^n$.

For a row–stochastic matrix W, its coefficient of ergodicity satisfies $0 \leq \tau(W) \leq 1$. Although scrambling matrices are ergodic, not every ergodic system matrix is scrambling [2].

2.2. Convergence Analysis of the Consensus Algorithm

The convergence of the averaging based distributed consensus algorithm has been studied by many authors and different sufficient conditions have been determined [5, 10, 13, 14, 37–39].

In 1986, Tsitsiklis *et al.* studied the consensus algorithm in the context of asynchronous update of distributed processors and proved that convergence is achieved provided that $w_{ii} \geq \delta$ for some $\delta > 0$, a directed path in the network exists from every node to all others and communication delays are bounded [37].

In 2003, a sufficient condition for asymptotical convergence of the consensus algorithm is given by Jadbabaie *et al.* for undirected graphs [10]. The proof is based on Wolfowitz's theorem on convergence of stochastic matrix products. It is stated that convergence is achieved if there exists an infinite sequence of contiguous nonempty bounded time intervals across which all nodes are linked together.

Li and Wang suggested a weaker condition for the convergence of the system (2.1). They require the network graph to be connected in a limited number of steps if two vertices were connected to each other [38].

Olfati–Saber and Murray discussed consensus problems with fixed and switching topologies and time delays [13]. The convergence is shown using Lyapunov theory provided that the balance condition of the graph is satisfied.

In 2005, both Moreau and Ren and Beard stated necessary and sufficient conditions on the convergence of such multi–agent systems [14, 39]. Moreau showed that in order to achieve consensus either the associated graph of the infinite product of weighting matrices must be connected or a node must exist in the network from which the rest of the nodes are accessible across all time steps of length T where T is a positive integer [39]. In [14], this requirement is stated as the union of the collection of interaction graphs across some time intervals having a spanning tree frequently enough.

Blondel *et al.* also studied the convergence of such systems, for networks with time–varying coefficients and time delays. They showed that convergence is reached provided that the network graph is connected for all t and both intercommunication intervals and delays are bounded [5].

In 2011, Akar and Shorten used the properties of scrambling matrices and related the consensus conditions to the existence of a common norm for a set of row–stochastic matrices associated with the original set of system matrices [2].

The convergence of the algorithm can be proved by Perron–Frobenius Theorem

Figure 2.1. A simple network of two nodes.

[40]. Article (iii) in Assumption 2.1 ensures that averaging coefficients sum up to one and consequently 1 is an eigenvalue of the system matrix W with the corresponding eigenvector $\mathbf{1} = [1, \ldots, 1]^\top$. In order for the system (2.3) with a time–independent system matrix W to achieve consensus, the multiplicity of the eigenvalue 1 must be equal to one. If there does not exist a node from which all others are accessible in the network, then the multiplicity of the eigenvalue 1 is greater than one. This condition is necessary for the system matrix W to provide convergence, however it is not sufficient as illustrated in the following example.

Example 2.5. Consider the two node network depicted in Figure 2.1 and let the system matrix W be given as

$$W = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{2.7}$$

In this scenario, we have $x_1(t+1) = x_2(t)$ and $x_2(t+1) = x_1(t)$ so that the values of nodes switch in each time step. Therefore convergence cannot be achieved.

The above example shows that Assumption 2.1 combined with the node accessibility condition is not sufficient for a convergent system. If we further require each node to use its own value, i.e., $w_{ii} > 0$ for all i, asymptotical convergence of (2.3) with time–independent W is guaranteed. Yet, this condition is not necessary in general and it is not assumed to hold in Chapters 3 and 4.

2.3. Graph Representation of Networks

Data flow in communication networks is generally expressed in terms of its graph representation. In the following definitions [41], let $G = (\mathcal{V}, \mathcal{E})$ be an undirected graph without multiple edges where $\mathcal{V} = \{1, \ldots, n\}$ is the set of vertices and $\mathcal{E} \in \mathcal{V} \times \mathcal{V}$ is

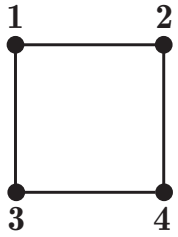


Figure 2.2. An undirected graph with four nodes.

the set of edges.

Definition 2.6. (Adjacency Matrix): The adjacency matrix of G is the $n \times n$ matrix $A = [A_{ij}]$ where $A_{ij} = 1$ when there is an edge from j to i in G, and $A_{ij} = 0$ otherwise. Note that $A_{ii} = 0$ for $i = 1, \ldots, n$ by definition.

Example 2.7. Consider the undirected graph depicted in Figure 2.2. The adjacency matrix corresponding to the graph is given as

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}. \tag{2.8}$$

Definition 2.8. (Spectrum of a graph): The spectrum of a graph G is the eigenvalue set of its adjacency matrix A with multiplicities.

Definition 2.9. (Neighbor): A vertex j is said to be a neighbor of vertex j if there exists a directed edge from vertex j to vertex i.

2.3.1. Graph Automorphism and Classes

The graphs G and H with the vertex set $\mathcal{V} = \{1, \ldots, n\}$ are called *isomorphic* if there exists a permutation p of $\mathcal{V}$ so that (i, j) is in the edge set of G if and only if $(p(i), p(j))$ is in the edge set of H. A graph automorphism of G is a graph isomorphism from G to itself, i.e., the graph is mapped onto itself while its edge–vertex connectivity is being preserved [28].

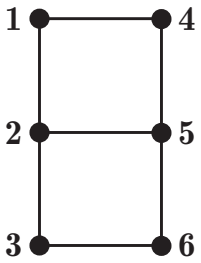


Figure 2.3. An undirected graph of six nodes with vertices labeled.

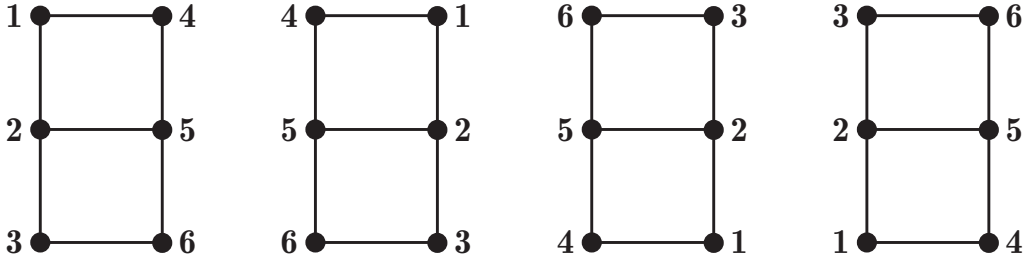
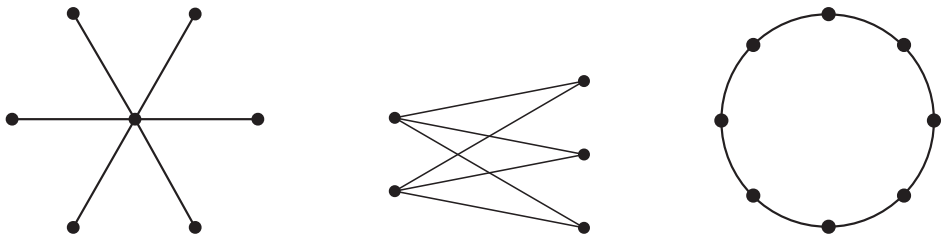


Figure 2.4. Four automorphisms of the graph illustrated in Figure 2.3.

Example 2.10. Consider the graph given in Figure 2.3. The automorphisms of the graph are depicted in Figure 2.4.

The set of all possible permutations is called the automorphism group of G and denoted by $\mathrm{Aut}(G)$. Following the definition of automorphism, we define edge–transitive, vertex–transitive, distance–transitive graphs which are of interest in this thesis.

2.3.1.1. Edge–Transitive Graphs. A graph G is called edge–transitive if every edge pair in G is equivalent under some element of its automorphism group. In other words, every edge in an edge–transitive graph has the same local environment, i.e., edges are indistinguishable from another based on surrounding edges and vertices. Some edge–transitive graphs are depicted in Figure 2.5. The n–node star graph given in Figure 2.5a consists of a central node that is connected to $n-1$ nodes. A complete bipartite graph consists of two groups where nodes of a group are connected to all nodes in the other group as illustrated in Figure 2.5b. Each vertex in a cycle is connected to two neighboring vertices to form a closed circle as depicted in Figure 2.5c.



(a) (b) (c)

Figure 2.5. Some edge–transitive graphs: (a) the star graph with $n = 7$ nodes, (b) the complete bipartite graph with $n_1 = 2$ and $n_2 = 3$, (c) the cycle graph with $n = 8$ nodes.

2.3.1.2. Vertex–Transitive Graphs. A graph G is called vertex–transitive if every vertex pair in G is equivalent under some element of its automorphism group. In other words, every vertex in a vertex–transitive graph has the same local environment, i.e., vertices are indistinguishable from another based on surrounding edges and vertices.

2.3.1.3. Distance–Transitive Graphs. In order to define distance–transitive graphs, the definition of *distance* is given below.

Definition 2.11. (Distance): The distance between the vertices i and j, denoted by $d(i, j)$, is the length of the shortest path from i to j.

A graph G is called distance–transitive if for any two vertex pairs $\{i, j\}$ and $\{i', j'\}$ with the same distance (i.e., $d(i, j) = d(i', j')$), there exists an automorphism that takes i to i' and j to j'. Note that distance–transitivity implies both edge and vertex transitivity.

Some distance–transitive graphs are depicted in Figure 2.6. In the complete graph, the distance between any two vertices is equal to one as shown in Figure 2.6a [28]. The Petersen graph, illustrated in Figure 2.6b, consists of 10 vertices and 15 edges whose vertices can be colored with three colors so that an edge does not connect vertices of the same color. The Heawood graph is an undirected graph with 14 vertices and 21 edges as depicted in Figure 2.6c. Hamming graphs are the Cartesian graph products of k copies of the complete graph (Figure 2.6d). A Johnson graph, denoted

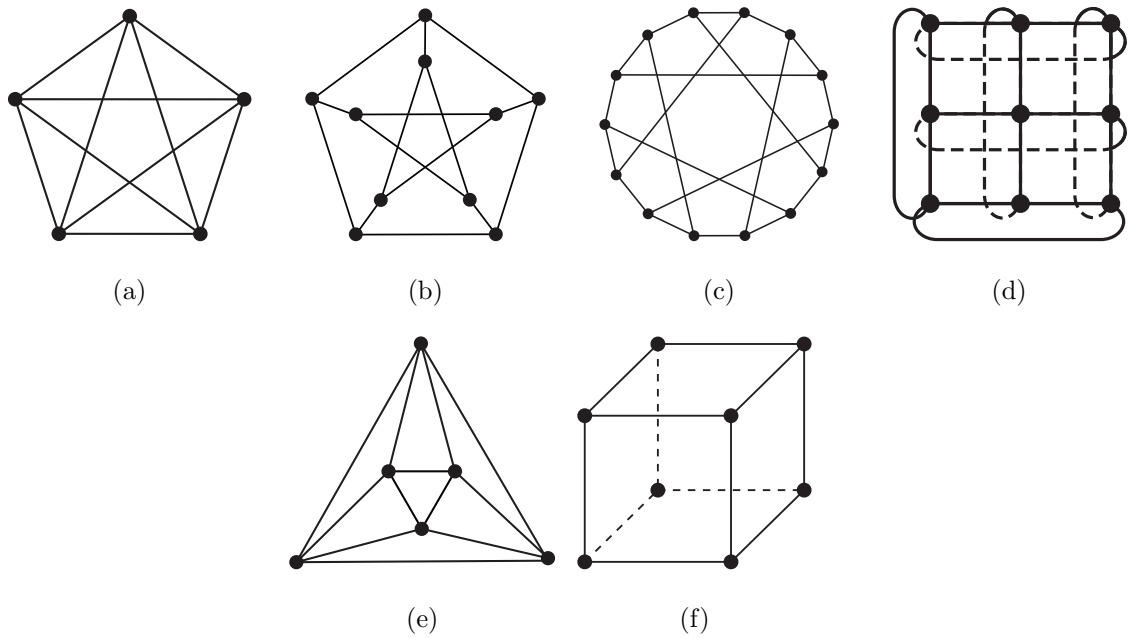


Figure 2.6. Some distance–transitive graphs: (a) a complete graph, (b) the Petersen graph, (c) the Heawood graph, (d) a Hamming graph, (e) a Johnson graph, (f) a hypercube.

by $J(n,k)$ and illustrated in Figure 2.6e, is an undirected graph whose vertices are the k–subsets of $\{1,\ldots,n\}$ where two vertices are connected if and only if their intersection has size $k-1$. The d–dimensional hypercube consists of 2^d vertices that can be labeled with binary words of length d where there exists an edge between two vertices if and only if their words differ only one component. The 3–d hypercube is given in Figure 2.6f.

2.4. Terms Related to Convergence Speed of Consensus Algorithms

Given an initial state $x(0)$ and

$$d_0 = \max_{i=1,2,\ldots,n} x_i(0) - \min_{i=1,2,\ldots,n} x_i(0), \tag{2.9}$$

we are interested in determining the number of steps, k, required to achieve a chosen consensus accuracy level, ϵ. In this section, we will discuss two terms related to the study of convergence speed of consensus algorithms: the coefficient of ergodicity and the second largest eigenvalue modulus (SLEM) of the system matrix.

2.4.1. The Coefficient of Ergodicity

Recall that the coefficient of ergodicity, denoted by $\tau(W)$, for a row–stochastic matrix W is computed from (2.5). One important property of the system matrix is that the coefficient of ergodicity satisfies $\tau(W) \leq 1$ for a row–stochastic matrix W. Furthermore, from Lemma 2.3, we know that the maximum state difference is contracting when the matrix is scrambling.

In order to derive an estimate on the convergence speed of a system with a scrambling matrix W, let d_k be the difference between the maximum and minimum components of the vector $x(k)$ at time k, $k \geq 0$. From (2.5), we have $d_{k+1} \leq \tau(W)d_k$, $k \geq 0$, which can be iterated to obtain $d_k \leq \tau(W)^k d_0$. Since it is desired that $d_k \leq \epsilon$, one can impose $d_k \leq \tau(W)^k d_0 \leq \epsilon$, i.e., $k \log \tau(W) \leq \log(\epsilon/d_0)$, which leads to $k \geq \log_{\tau(W)}(\epsilon/d_0)$ (since $\tau(W) < 1$ for a scrambling matrix). Therefore, in at most

$$k_{max} = \lceil \log_{\tau(A)}(\epsilon/d_0) \rceil \tag{2.10}$$

number of iterations, the specified accuracy level ϵ is achieved for the given initial difference d_0. It can be deduced from (2.10) that k_{max} is smaller for a smaller coefficient of ergodicity $\tau(W) < 1$; however, this does not necessarily ensure faster convergence as illustrated in the following example.

Example 2.12. Consider two systems with system matrices W_1 and W_2 given as

$$W_1 = \begin{bmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 3/4 & 1/4 \end{bmatrix} \quad \text{and} \quad W_2 = \begin{bmatrix} 3/5 & 2/5 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \end{bmatrix}. \tag{2.11}$$

The state evolution of these two systems are depicted in Figure 2.7 for the initial state $x(0) = [10,\ 2,\ 0]^\top$, whereas the maximum difference of the components of the state vector is shown in Figure 2.8 for ease of comparison. We note from Figures 2.7 and 2.8 that the first system achieves consensus in less steps than the second one.

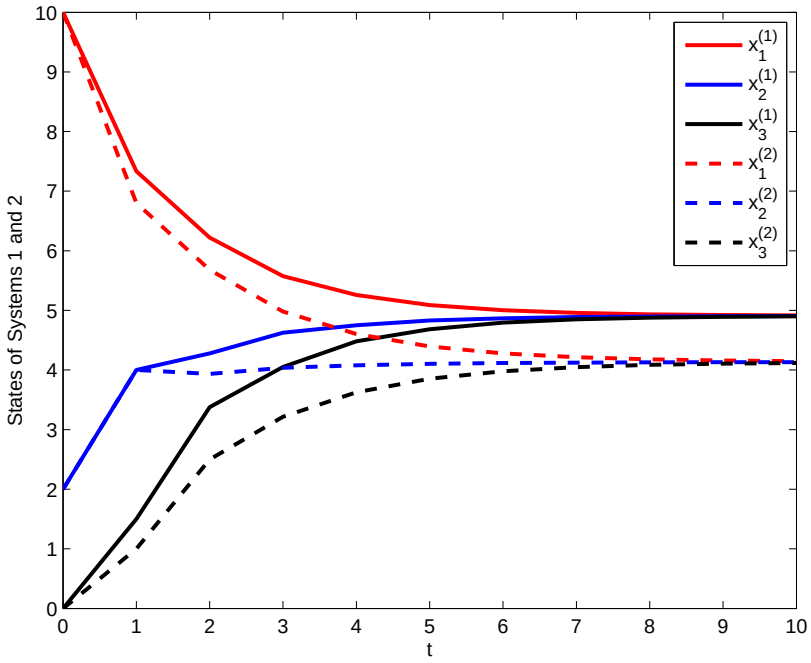


Figure 2.7. State values for System 1 and System 2 ($x_i^{(j)}$ denotes the i–th state of the j–th system).

In Figures 2.7 and 2.8 and the rest of the simulation results presented in this chapter, the selected numbers do not immediately correspond to actual physical values, but rather scaled versions of physical quantities, such as heading angle that is encountered in formation control, or clock frequencies of the nodes in e.g. sensor networks.

Table 2.1 summarizes the theoretical maximum and actual number of steps required for a range of ϵ/d_0. Although the theoretical k_{max} value computed via (2.10) for System 1 is greater than that of System 2 for all ϵ/d_0, the actual required number of steps is always less or equal. Moreover, it is also noted from the same table that k_{actual} increases as ϵ/d_0 decreases, which is expected since higher accuracy requires a larger number of iterations.

Since the coefficient of ergodicity is not a good means to compare convergence speeds as illustrated by the above example, we have to rely on alternative properties of the system matrices to make better comparisons.

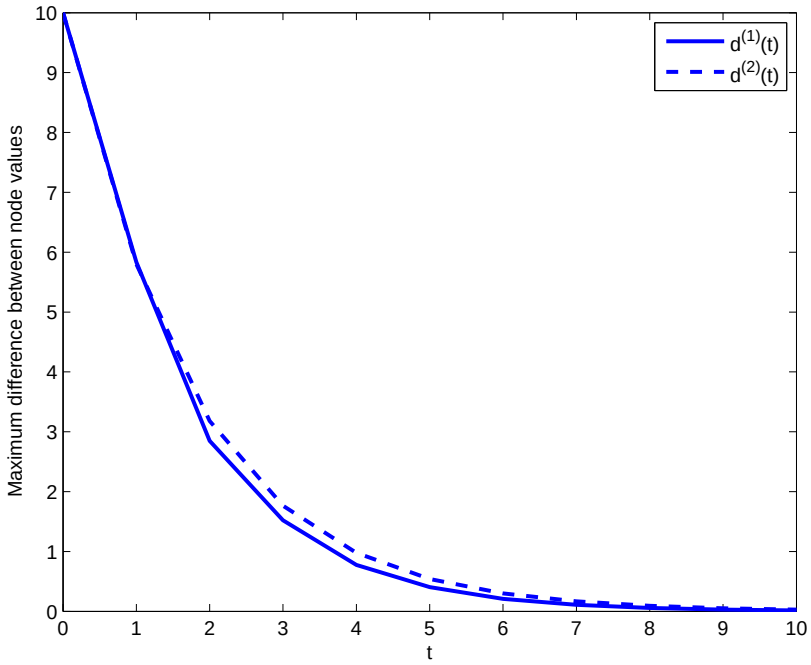


Figure 2.8. Maximum differences, $d^{(1)}$ and $d^{(2)}$, for System 1 and System 2.

Table 2.1. Theoretical maximum and actual number of steps required for varying ϵ/d_0.

		$\epsilon/d_0 = 10^{-2}$		$\epsilon/d_0 = 10^{-3}$		$\epsilon/d_0 = 10^{-4}$		$\epsilon/d_0 = 10^{-5}$	
System	τ	k_{max}	k_{actual}	k_{max}	k_{actual}	k_{max}	k_{actual}	k_{max}	k_{actual}
1	0.667	12	8	18	11	23	15	29	18
2	0.6	10	8	14	12	19	16	23	20

2.4.2. The Second Largest Eigenvalue (SLEM)

The largest eigenvalue of a row–stochastic matrix is equal to one. Moreover, all other eigenvalues are strictly less than one (in magnitude) when consensus is achieved. Let the eigenvalues of W be real and denoted in non–increasing order:

$$1 = \lambda_1(W) \geq \lambda_2(W) \geq \ldots \geq \lambda_n(W) \geq -1.$$

Then the second largest eigenvalue modulus, denoted by μ can be given by

$$\mu(W) = \max_{i=2,\ldots,n} |\lambda_i(W)| = \max\{\lambda_2(W), -\lambda_n(W)\}. \tag{2.12}$$

Suppose that the system matrix W is related to its Jordan form as $W = TJT^{-1}$. Then, the k–th power of the system matrix can be computed as

$$W^k = TJ^kT^{-1} = Q_{11}1^k + \sum_{i=2}^{q}\sum_{l=0}^{m_i-1} Q_{il}k^l\lambda_i^k \tag{2.13}$$

where q is the number of distinct eigenvalues of W; m_i is the multiplicity of i–th eigenvalue λ_i; and Q_{ij} are appropriate scaling matrices. Then the state vector at the k–th step is equal to

$$x(k) = W^kx_0 = Q_{11}x_0 + \sum_{i=2}^{q}\sum_{l=0}^{m_i-1} (Q_{il}x_0)k^l\lambda_i^k. \tag{2.14}$$

Note that the sum term in the above expression vanishes faster when the second largest eigenvalue modulus is smaller. Therefore, the convergence speed is directly related to μ, i.e., a smaller μ results in faster convergence.

Example 2.13. Consider the system matrices given in Example 2.12. For W_1 and W_2 we have $\mu(W_1) = 0.5181$ and $\mu(W_2) = 0.5537$, respectively. Since $\mu(W_1) < \mu(W_2)$, the first system converges faster.

2.5. Weighting Matrix Selection

For a given graph, the problem of finding the weighting matrix which yields the fastest convergence is not easy in general [26]. There are some heuristic methods in the literature to construct weighting matrices that provide a good (probably not the best) convergence speed. In this section, we consider three well known heuristic methods: the maximum–degree weights, the local–degree weights and the Metropolis–Hastings weights [26, 42].

2.5.1. The Maximum–Degree Weights

Let N_i and N_{max} be the degree (valency) of vertex i (number of neighbors of node i) and the maximum degree of an undirected graph, respectively. The maximum–degree

chain method assigns $1/N_{max}$ to all transition probabilities except holding probabilities. The holding probabilities are assigned properly so that each row adds up to one.

The weighting matrix denoted by $P^{md} = [p_{ij}^{md}]$ is given in [26] by

$$p_{ij}^{md} = \begin{cases} 1/N_{max}, & \text{if } i \neq j \text{ and } (i,j) \in \mathcal{E} \\ 1 - N_i/N_{max}, & \text{if } i = j \\ 0, & \text{if } (i,j) \notin \mathcal{E}. \end{cases} \tag{2.15}$$

2.5.2. The Local–Degree Weights

The local–degree weights method uses the out–degree information of vertices. The transition probability of an edge is inversely proportional to the maximum of its incident vertices [42], that is

$$p_{ij}^{ld} = \begin{cases} 1/\max\{N_i^{out}, N_j^{out}\}, & \text{if } i \neq j \text{ and } (i,j) \in \mathcal{E} \\ 0, & \text{if } (i,j) \notin \mathcal{E} \\ 1 - \sum_{j \neq i} p_{ij}^{ld}, & \text{if } i = j. \end{cases} \tag{2.16}$$

where N_i^{out} is the number of directed edges where i is the head of the edge. Note that in order to implement the local–degree weights method, each vertex must know the out–degrees of its neighbors.

2.5.3. The Metropolis–Hastings Weights

For the random walk on a graph, the transition probability matrix $P^{rw} = [p_{ij}^{rw}]$ is given in [26] by

$$p_{ij}^{rw} = \begin{cases} 1/N_i, & \text{if } i \neq j \text{ and } (i,j) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \tag{2.17}$$

Note that the holding probabilities of vertices are zero and the transition probabilities from a given vertex to its neighbors are equal. The matrix P^{rw} is not symmetric and the stationary distribution of the random walk is proportional to vertex degrees. For a given stationary distribution $\pi = [\pi_1, \ldots, \pi_n]$, applying the Metropolis–Hastings algorithm to the random walk results in

$$p_{ij}^{mh} = \begin{cases} p_{ij}^{rw} \min\{1, r_{ij}\}, & \text{if } i \neq j \text{ and } (i,j) \in \mathcal{E} \\ p_{ii}^{rw} + \sum\limits_{(i,k)\in\mathcal{E}} p_{ik}^{rw}(1 - \min\{1, r_{ik}\}), & \text{if } i = j. \end{cases} \tag{2.18}$$

where $r_{ij} = (\pi_j p_{ji}^{rw})/(\pi_i p_{ij}^{rw})$. When the stationary distribution is the uniform distribution, i.e., $\pi = [1/n, \ldots, 1/n]$, the transition probability matrix is symmetric and computed as

$$p_{ij}^{mh} = \begin{cases} \min\{1/N_i, 1/N_j\}, & \text{if } i \neq j \text{ and } (i,j) \in \mathcal{E} \\ \sum\limits_{(i,k)\in\mathcal{E}} \max\{0, 1/N_i - 1/N_k\}, & \text{if } i = j \\ 0, & \text{if } (i,j) \notin \mathcal{E}. \end{cases} \tag{2.19}$$

2.6. Chapter Summary

In this chapter, we have introduced the averaging based distributed consensus algorithm and discussed its convergence properties that is widely studied in the literature. The relevant graph theory has been reviewed and some graph classes (edge–transitive, vertex–transitive and distance–transitive graphs) have been introduced. Since we study the convergence rate of the consensus algorithms in the thesis, two terms that are related to the convergence speed have been explained and some heuristic weighting matrix selection algorithms to achieve a good (probably not the best) convergence rate have been presented.

3. THE FASTEST MIXING MARKOV CHAINS ON UNDIRECTED GRAPHS

The problem of achieving the fastest convergence in consensus algorithms can be investigated using the theory of probabilistic matrices, namely Markov chains. Since the system matrix of the consensus algorithm given in (2.3) is a Markov chain, the mixing rate of the Markov chain is equivalent to the convergence speed of the consensus algorithm. In this chapter, we study two alternative methods of assigning transition probabilities to a Markov chain in order to optimize its mixing rate. In the first SDP formulation, there is a single transition probability parameter to be optimized (the holding probability of vertices) which yields to faster computation as opposed to the more general reversible Markov chain formulation corresponding to a stationary distribution that is proportional to the degree of vertices. Both the single parameter and the degree proportional reversible FMMC formulations tend to yield better results than the symmetric SDP formulation for some well–known graphs. This claim is verified by deriving exact analytical results for the graphs under consideration.

3.1. Introduction to Markov Chains

Let $G = (\mathcal{V}, \mathcal{E})$ be an undirected graph without multiple edges where $\mathcal{V} = \{1, \ldots, n\}$ is the set of vertices and $\mathcal{E} \in \mathcal{V} \times \mathcal{V}$ is the set of edges. Each vertex is assumed to have a self–loop, i.e., $(i, i) \in \mathcal{E}$ for $i = 1, \ldots, n$.

Let the state at time step k be denoted by $X(k)$. The probability of transition from state i to j can be defined as

$$P_{ij} = \mathbf{Prob}\left(X(k+1) = j \mid X(k) = i\right), \quad i, j = 1, \ \ldots, \ n.$$

Throughout this chapter, the Markov chain is defined by its transition probability matrix, $P = [P_{ij}]$. The transition probability matrix P satisfies $P \geq 0$ and $P\mathbf{1} = \mathbf{1}$

which ensure that P_{ij} are non–negative and they sum up to one for each vertex and $\mathbf{1}$ is the vector of ones. Let the probability distribution at time k be denoted by the vector $\pi(k) \in \Re^n$ where $\pi_i(k) = \mathbf{Prob}(X(k) = i)$. The relationship between the state distributions at time k and $k+1$ is given as

$$\pi(k+1)^\top = \pi(k)^\top P$$

which can be iterated to obtain

$$\pi(k)^\top = \pi(0)^\top P^k.$$

3.2. The Fastest Mixing Symmetric Markov Chain

The problem of determining the fastest mixing symmetric Markov chain is considered in [26–32] and shown to be a convex optimization problem which is solvable by Semi Definite Programming (SDP) techniques. In [26], the FMMC problem for symmetric Markov chains is formulated as

$$\begin{array}{ll} \text{minimize} & \mu(P) \\ \text{subject to} & P \geq 0, \quad P\mathbf{1} = \mathbf{1}, P = P^\top, \\ & P_{ij} = 0, (i,j) \notin \mathcal{E}. \end{array} \tag{3.1}$$

Let the optimal second largest eigenvalue modulus (SLEM) be denoted by μ^*:

$$\mu^* = \inf \left\{ \mu(P) | P \geq 0, P\mathbf{1} = \mathbf{1}, P = P^\top, P_{ij} = 0, (i,j) \notin \mathcal{E} \right\}$$

3.3. The Fastest Mixing Reversible Markov Chain

The FMMC problem can be generalized to reversible Markov chains on undirected graphs [26]. In the fastest mixing reversible Markov chain problem, the transition

probability matrix P need not be symmetric, however, the balance condition given by

$$\pi_i P_{ij} = \pi_j P_{ji}, \quad i, j = 1, \ldots, n \tag{3.2}$$

must be satisfied where π_i is the stationary distribution of vertex i. Let the vector $\pi = [\pi_1, \ldots, \pi_n]$ denote the equilibrium distribution of the Markov chain. For fixed π, it is desired to determine the transition probability matrix P with the smallest SLEM. The balance condition (3.2) can be re–written as $T^{-2}P = P^\top T^{-2}$ where $T = diag\left\{\frac{1}{\sqrt{\pi_1}}, \ldots, \frac{1}{\sqrt{\pi_n}}\right\}$ is a similarity transformation matrix such that $\hat{P} = T^{-1}PT$ is symmetric. Since similar matrices have the same spectra, we have $\mu(P) = \mu(\hat{P})$. The unit eigenvector corresponding to the largest eigenvalue of $\hat{P}$, $\lambda_1 = 1$, is calculated as

$$v_T = \frac{T^{-1}\mathbf{1}}{||T^{-1}\mathbf{1}||} \tag{3.3}$$

where $||.||$ denotes the Euclidean norm.

The convexity of $\mu(\hat{P})$ can be proved by using spectral norm and the properties of $\hat{P}$. The second maximum singular value of $\hat{P}$ can be expressed as the spectral norm of the orthogonal projection of $\hat{P}$ on $v_T^\perp$, i.e.,

$$\mu(\hat{P}) = ||(I_n - v_T v_T^\top)\hat{P}(I_n - v_T v_T^\top))||_2 = ||\hat{P} - v_T v_T^\top||_2 \tag{3.4}$$

where I_n is the identity matrix of size n. Moreover, $\hat{P}$ is symmetric which ensures that its singular values and eigenvalues are the same. Therefore, magnitude of the second largest eigenvalue, expressed as the norm of an affine function in (3.4), is convex. Consequently, the fastest mixing reversible Markov chain problem can be formulated as

$$\begin{array}{ll} \text{minimize} & r \\ \text{subject to} & -rI_n \preceq T^{-1}PT - v_T v_T^\top \preceq rI_n, \\ & P \geq 0, \;\; P\mathbf{1} = \mathbf{1}, \;\; T^{-2}P = P^\top T^{-2} \\ & P_{ij} = 0, \quad (i,j) \notin \mathcal{E} \end{array} \tag{3.5}$$

The symbol $\leq$ represents elementwise inequality while the symbol $\preceq$ denotes matrix inequality (i.e., $A \preceq B$ corresponds to positive semi–definiteness of $B - A$). The variables of the formulation are r and P_{ij}.

Remark 3.1. In the fastest mixing symmetric Markov chain formulation of [26], we require $P_{ij} = P_{ji}$ for $1 \leq i, j \leq n$, that implies $\pi = [1/n, \ldots, 1/n]$. Therefore, the symmetric Markov chain of [26] is a reversible Markov chain with the stationary distribution $\pi = [1/n, \ldots, 1/n]$.

3.4. FMMC with Identical Holding Probabilities

In this chapter, we also consider the FMMC with identical holding and equal neighbor transition probabilities. More specifically, the following construction is assumed.

Assumption 3.2. Let the number of neighbors (degree) of vertex i be denoted by N_i. For some $a \in [0, 1]$, suppose that the transition probabilities are of the following form:

(i) $P_{ii} = a$ for $i = 1, \ldots, n$
(ii) $P_{ij} = \frac{1-a}{N_i}$ if $i \neq j$ and $(i, j) \in \mathcal{E}$ for $i, j = 1, \ldots, n$.

Article (i) in Assumption 3.2 ensures that all vertices have identical holding probabilities while Article (ii) induces that all neighbors of a given vertex have equal transition probabilities. Since P is not required to be symmetric, we use the following result in demonstrating the convexity of our SDP formulation.

Theorem 3.3. *Let P be a Markov chain on an undirected graph and suppose that Assumption 3.2 is satisfied. Then the spectrum of P is real.*

Proof. Let $A = A^\top$ be the adjacency matrix of the undirected graph. Then the Markov chain can be represented as

$$P = (1 - a)\ \mathrm{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \ldots, \frac{1}{N_n}\right) A + aI_n. \tag{3.6}$$

The matrix $T = \text{diag}\left(\frac{1}{\sqrt{N_1}}, \frac{1}{\sqrt{N_2}}, \ldots, \frac{1}{\sqrt{N_n}}\right)$ is a similarity transformation matrix such that $\hat{P} = T^{-1}PT$ is computed as

$$\hat{P} = [\hat{P}_{ij}] = \begin{cases} a, & \text{if } i = j \\ \frac{1-a}{\sqrt{N_i N_j}}, & \text{if } i \neq j \text{ and } (i,j) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \tag{3.7}$$

which is symmetric. Hence the spectrum of P is real. □

The unit eigenvector corresponding to the largest eigenvalue of $\hat{P}$, $\lambda_1 = 1$, is calculated as

$$v_T = \frac{T^{-1}\mathbf{1}}{||T^{-1}\mathbf{1}||} = \left(\sum_{i=1}^{n} N_i\right)^{-1/2} \left[\sqrt{N_1}, \ldots, \sqrt{N_n}\right]^\top. \tag{3.8}$$

In this chapter, the fastest mixing Markov chain problem on an undirected graph with identical holding probabilities is formulated as

$$\begin{aligned} \underset{a}{\text{minimize}} \quad & \mu(P) \\ \text{subject to} \quad & P_{ij} = \begin{cases} a, & \text{if } i = j \\ \frac{1-a}{N_i}, & \text{if } i \neq j \text{ and } (i,j) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \\ & 0 \leq a \leq 1 \end{aligned} \tag{3.9}$$

From Theorem 3.3, there exists a symmetric matrix $\hat{P}$ which is similar to P. Hence the FMMC problem can be reformulated as

$$\begin{aligned} \underset{a}{\text{minimize}} \quad & \mu(\hat{P}) \\ \text{subject to} \quad & \hat{P}_{ij} = \begin{cases} a, & \text{if } i = j \\ \frac{1-a}{\sqrt{N_i N_j}}, & \text{if } i \neq j \text{ and } (i,j) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \\ & 0 \leq a \leq 1 \end{aligned} \tag{3.10}$$

The formulation (3.10) is expressed as an SDP as follows:

$$\begin{aligned} \underset{a}{\text{minimize}} \quad & r \\ \text{subject to} \quad & -rI_n \preceq \hat{P} - v_T v_T^\top \preceq rI_n, \\ & \hat{P}_{ij} = \begin{cases} a, & \text{if } i = j \\ \frac{1-a}{\sqrt{N_i N_j}}, & \text{if } i \neq j \text{ and } (i,j) \in \mathcal{E}, \\ 0, & \text{otherwise} \end{cases} \\ & 0 \leq a \leq 1 \end{aligned} \tag{3.11}$$

The variables of the formulation are r and a. The inequalities in (3.11) can also be represented as a one–line linear matrix inequality (LMI):

$$\text{diag}(\hat{P} - v_T v_T^\top + rI_n,\ rI_n - \hat{P} + v_T v_T^\top,\ a,\ 1-a) \succeq 0. \tag{3.12}$$

Remark 3.4. Recall from Remark 3.1 that the symmetric Markov chain of [26] is reversible with the stationary distribution $\pi = [1/n, \ldots, 1/n]$. Similarly, the FMMC with identical holding probabilities given in (3.11) can be related to the fastest mixing reversible Markov chain in [26]. To this end, let the stationary distribution be

$$\pi = \frac{1}{\sum_{i=1}^{n} N_i}[N_1, N_2, \ldots, N_n]. \tag{3.13}$$

Although the two problem formulations (3.5) and (3.11) correspond to the same equilibrium distribution (3.13), note that their search spaces are different in general and therefore may not always lead to the same result. In order for the formulation (3.5) to be equivalent to the formulation (3.11), the following restriction must be added to (3.5):

$$P_{ij} = P_{ik}, \forall\ (i,j) \in \mathcal{E} \text{ and } (i,k) \in \mathcal{E},\ i \neq j,\ i \neq k.$$

Note that the above restriction requires all non–zero and non–diagonal elements of a row to be equal. Together with the balance condition (3.2), this directly implies the

conditions in Assumption 3.2.

In the rest of this chapter, we study edge–transitive graphs, distance–transitive graphs and orbit graphs in detail. We derive the analytical expressions of the optimal SLEM using formulations of (3.5) and (3.11), denoted by μ_r and μ_i, respectively. Subsequently, the results are compared with the symmetric solution of (3.1) that is denoted by μ_s.

3.5. FMMC on Edge–Transitive Graphs

A graph $\mathcal{G}$ is called edge–transitive if all edge pairs, (i, j), are equivalent under some element of its automorphism group. In this section, we derive the analytical expressions of the best mixing rate for 3 representative edge–transitive graphs: A star, a complete bipartite graph and a cycle. Since (3.11) and (3.5) with π in (3.13) lead to the same solution for the graphs under consideration, the analytical derivations are shown only for (3.11) for ease of exposition.

3.5.1. Star

The star graph is illustrated in Figure 2.5a. Without loss of generality, let the node at the center of the star be labeled as 1. Then the Markov chain on an n–node ($n \geq 3$) undirected star when vertices have identical holding probabilities is given by

$$S_n = \begin{bmatrix} a & \frac{1-a}{n-1} & \cdots & \frac{1-a}{n-1} & \frac{1-a}{n-1} \\ 1-a & a & 0 & \cdots & 0 \\ 1-a & 0 & a & \cdots & 0 \\ \vdots & & & \ddots & \\ 1-a & 0 & \cdots & 0 & a \end{bmatrix}. \tag{3.14}$$

The spectrum of S_n can be computed as follows. Consider the matrix $aI_n - S_n$. Since $rank(aI_n - S_n) = 2$, the matrix $aI_n - S_n$ has $n-2$ eigenvalues at zero; therefore a is an eigenvalue of S_n with multiplicity $n-2$. Since S_n is row–stochastic, we have

$\lambda_1(S_n) = 1$. The remaining eigenvalue can be computed as

$$trace(S_n) - \sum_{i=1}^{n-1} \lambda_i = 2a - 1. \tag{3.15}$$

where $trace(S_n)$ denotes the sum of diagonal elements of S_n. Hence the SLEM of S_n is

$$\mu_i(S_n) = \max\{a, |2a-1|\} \tag{3.16}$$

and the fastest mixing, $\mu_i^*(S_n) = 1/3$, is achieved for $a^* = 1/3$.

Remark 3.5. The mixing rate is independent of network size. Moreover, the optimal SLEM of the Markov chain on a star under Assumption 3.2 (that is equal to the solution of (3.5) with π in (3.13)) is always less than the SLEM of the optimal symmetric Markov chain, which is given in [28] as

$$\mu_s^*(S_n) = \frac{n-2}{n-1}. \tag{3.17}$$

Remark 3.6. In an edge–transitive graph, the fastest mixing symmetric Markov chain has a single transition probability that is equal for all edges, i.e., $P_{ij} = p$ for $i \neq j$ and $(i,j) \in \mathcal{E}$ [28]. Furthermore, the non–negativity of the Markov chain requires $0 < p \leq \frac{1}{\max\limits_j N_j}$. Consequently, the diagonal elements of the symmetric Markov chain (that correspond to the holding probability of vertices) are lower bounded by $P_{ii} \geq 1 - \frac{N_i}{\max\limits_j N_j}$. For the specific case of the symmetric Markov chain on a star (which is also edge–transitive), note that $n-2$ eigenvalues are among the diagonal elements that correspond to the vertices of degree 1. This in turn implies that the symmetric formulation of [28] cannot yield a result any better than $\mu_s(P) \geq P_{ii} \geq 1 - \frac{1}{n-1}$ for $i = 2, \ldots, n$.

On the other hand, the diagonal element of the Markov chain (that is among the eigenvalues) in our formulation is an optimization parameter. Hence, we do not have the restriction that the fastest mixing symmetric Markov chain formulation of [28] has. Furthermore, note from (3.16) that the SLEM is $\max\{a, |2a-1|\}$. This might provide

some insight into why our formulation provides better results in edge–transitive graphs, whose eigenvalues consist of the diagonal elements.

3.5.2. Complete Bipartite Graph

The complete bipartite graph is illustrated in Figure 2.5b. Let the number of vertices in two parts of the network be n_1 and n_2, respectively, and without loss of generality let $n_2 \geq n_1$. When all vertices have identical holding probabilities, the Markov chain has the form

$$B_{n_1,n_2} = \begin{bmatrix} aI_{n_1} & \frac{1-a}{n_2}\mathbf{1}_{n_1}\mathbf{1}_{n_2}^\top \\ \frac{1-a}{n_1}\mathbf{1}_{n_2}\mathbf{1}_{n_1}^\top & aI_{n_2} \end{bmatrix}. \tag{3.18}$$

Since $\text{rank}(aI_{n_1+n_2} - B_{n_1,n_2}) = 2$, a is an eigenvalue of B_{n_1,n_2} with multiplicity n_1+n_2-2. Moreover, B_{n_1,n_2} is row–stochastic so that we have $\lambda_1 = 1$. The remaining eigenvalue can be computed as

$$\text{trace}(B_{n_1,n_2}) - (n_1 + n_2 - 2)a - 1 = 2a - 1. \tag{3.19}$$

Therefore, the SLEM can be analytically formulated as

$$\mu_i(B_{n_1,n_2}) = \max\{a, |2a-1|\} \tag{3.20}$$

and the fastest mixing, $\mu_i^*(B_{n_1,n_2}) = 1/3$, is achieved for $a^* = 1/3$.

Remark 3.7. The optimal SLEM of a Markov chain on a complete bipartite graph under Assumption 3.2 is independent of network size. Moreover, it is always less than or equal to the SLEM of the optimal symmetric Markov chain in [28] given by

$$\mu_s^* = \max\left\{\frac{n_2 - n_1}{n_2}, \quad \frac{n_2}{n_2 + 2n_1}\right\}. \tag{3.21}$$

Remark 3.8. Since complete bipartite graphs are edge–transitive, the holding proba-

bilities of a symmetric Markov chain are lower bounded by $P_{ii} \geq 1 - \frac{N_i}{\max\limits_j N_j}$. For the symmetric Markov chain on a bipartite graph, $n_1 + n_2 - 2$ eigenvalues are located in the diagonals, $n_2 - 1$ of which are lower bounded by $1 - \frac{n_1}{n_2}$. Consequently the symmetric formulation of [28] yields $\mu_s \geq 1 - \frac{n_1}{n_2}$. One underlying reason for the better performance of our formulation in this case is the flexibility to optimize the mixing rate given by (3.20) over the optimization parameter, which is the diagonal element of the Markov chain.

3.5.3. Cycle

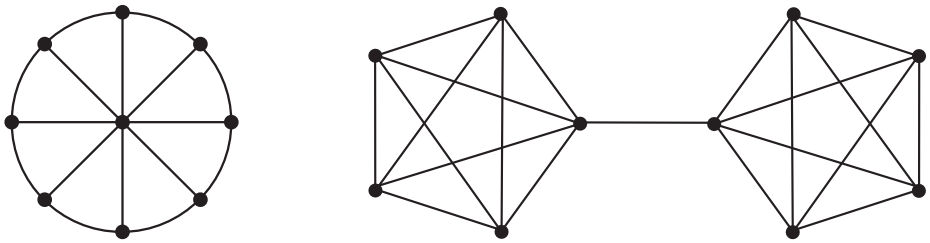
The Markov chain on an n–node undirected cycle depicted in Figure 2.5c with identical holding probabilities is given as

$$C_n = \begin{bmatrix} a & \frac{1-a}{2} & 0 & \cdots & \frac{1-a}{2} \\ \frac{1-a}{2} & a & \frac{1-a}{2} & 0 & 0 \\ 0 & \frac{1-a}{2} & a & \frac{1-a}{2} & 0 \\ \vdots & & & \ddots & \\ \frac{1-a}{2} & 0 & \cdots & \frac{1-a}{2} & a \end{bmatrix}. \tag{3.22}$$

Since all edges have the same transition probability, $P_{ij} = \frac{1-a}{2}$, $(i,j) \in \mathcal{E}$, this problem is identical not only to (3.5) with π in (3.13) but also to the FMMC problem for symmetric Markov chains discussed in [28] where the solution is given by

$$a^* = \frac{\cos\frac{2\pi}{n} + \cos\frac{2\lfloor n/2 \rfloor \pi}{n}}{\cos\frac{2\pi}{n} + \cos\frac{2\lfloor n/2 \rfloor \pi}{n} - 2} \tag{3.23}$$

$$\mu_i^*(C_n) = \frac{\cos\frac{2\pi}{n} - \cos\frac{2\lfloor n/2 \rfloor \pi}{n}}{2 - \cos\frac{2\pi}{n} - \cos\frac{2\lfloor n/2 \rfloor \pi}{n}} \tag{3.24}$$



(a) (b)

Figure 3.1. Some orbit graphs under consideration: (a) the wheel graph with 9 nodes ($n = 8$), (b) the graph $K_n - K_n$ for $n = 5$.

3.6. FMMC on Distance–Transitive Graphs

Distance transitive graphs are both edge–transitive and vertex–transitive. All vertices of a distance–transitive graph have the same degree, i.e., $N_i = N_j$ for all $i, j = 1, \ldots n$. Therefore, the Markov chain P is symmetric and the analytical results for symmetric FMMC problem are also valid for the non–symmetric FMMC problem under Assumption 3.2 and for (3.5) with π in (3.13). The complete graph, the Petersen graph, the Heawood graph, Hamming graphs, Johnson graphs, hypercubes and cycles are some examples of distance–transitive graphs which are illustrated in Figure 2.6 [28].

3.7. FMMC on Orbit Graphs

In this section, we consider graphs that have large automorphism groups so that the eigenvalue multiplicities of the transition probability matrix are high in general [28]. In [43], the authors utilized orbit theory to build a reduced order matrix which consists of all distinct eigenvalues of the transition probability matrix. This provides a reduced the number of optimization variables which results in faster solution of the SDP problem. In the following sections, we present the analytical results for some well–known orbit graphs including the wheel graph and the graph $K_n - K_n$.

3.7.1. The Wheel Graph

The wheel graph, depicted in Figure 3.1a, is composed of a central hub and a cycle of n nodes each of which is connected to the hub. The total number of nodes in a wheel graph is $n+1$. Let the central hub be labeled by 1. Then the optimal reversible Markov chain on a wheel graph for π in (3.13) has the form

$$U_n = \begin{bmatrix} a_1 & \frac{1-a_1}{n} & \frac{1-a_1}{n} & \cdots & \cdots & \frac{1-a_1}{n} \\ \frac{1-a_1}{3} & a_2 & b & 0 & \cdots & b \\ \frac{1-a_1}{3} & b & a_2 & b & 0 & 0 \\ \frac{1-a_1}{3} & 0 & b & a_2 & b & 0 \\ \frac{1-a_1}{3} & \vdots & & & \ddots & \\ \frac{1-a_1}{3} & b & 0 & \cdots & b & a_2 \end{bmatrix}, \tag{3.25}$$

where $b = \frac{1-a_2-\frac{1-a_1}{3}}{2}$. Let T_w be the transformation matrix given by $T_w = \mathbf{Diag}(1, \mathcal{F})$ where $\mathbf{Diag}(1, \mathcal{F})$ is a block diagonal matrix with scalar 1 as its first block and

$$\mathcal{F} = [\mathcal{F}_{jk}] = \left[\frac{1}{\sqrt{n}} e^{\frac{2\pi i(j-1)(k-1)}{n}}\right]$$

is the $n \times n$ unitary Fourier matrix, as the second. Then $T_w^{-1} U T_w$ is a block diagonal matrix where the first block is a 2×2 matrix given by

$$\begin{bmatrix} a_1 & \frac{1-a_1}{\sqrt{n}} \\ \frac{(1-a_1)\sqrt{n}}{3} & \frac{a_1+2}{3} \end{bmatrix} \tag{3.26}$$

and the rest $n-1$ blocks are scalars that satisfy

$$a_2 + \frac{(1-a_2-\frac{1-a_1}{3})}{n}\left(\frac{n-1}{2}(\omega_n^k + \omega_n^{-k}) + \frac{1}{2}(\omega_n^{n-k} + \omega_n^{-(n-k)})\right) \tag{3.27}$$

where $\omega_n = e^{\frac{2\pi i}{n}}$, for $k = 1, \ldots n-1$ or equivalently

$$a_2 + (1 - a_2 - \frac{1-a_1}{3}) \cos\frac{2\pi k}{n}. \tag{3.28}$$

The optimal SLEM μ^* and the optimal holding probabilities a_1^* and a_2^* can be computed from the above decomposition as

$$a_1^* = \begin{cases} \dfrac{2 - 3\cos\frac{2\pi}{n} + \cos\frac{2\lfloor n/2\rfloor\pi}{n}}{8 - 3\cos\frac{2\pi}{n} - 5\cos\frac{2\lfloor n/2\rfloor\pi}{n}}, & \text{if } n \leq 5 \\ 0, & \text{if } n > 5 \end{cases} \tag{3.29}$$

$$a_2^* = \begin{cases} \dfrac{-3\left(\cos\frac{2\pi}{n} + \cos\frac{2\lfloor n/2\rfloor\pi}{n}\right)}{8 - 3\cos\frac{2\pi}{n} - 5\cos\frac{2\lfloor n/2\rfloor\pi}{n}}, & \text{if } n \leq 5 \\ \dfrac{-2\left(\cos\frac{2\pi}{n} + \cos\frac{2\lfloor n/2\rfloor\pi}{n}\right)}{3\left(2 - \cos\frac{2\pi}{n} - \cos\frac{2\lfloor n/2\rfloor\pi}{n}\right)}, & \text{if } n > 5 \end{cases} \tag{3.30}$$

$$\mu_r^*(U) = \begin{cases} \dfrac{3\left(\cos\frac{2\pi}{n} - \cos\frac{2\lfloor n/2\rfloor\pi}{n}\right)}{8 - 3\cos\frac{2\pi}{n} - 5\cos\frac{2\lfloor n/2\rfloor\pi}{n}}, & \text{if } n \leq 5 \\ \dfrac{2\left(\cos\frac{2\pi}{n} - \cos\frac{2\lfloor n/2\rfloor\pi}{n}\right)}{3\left(2 - \cos\frac{2\pi}{n} - \cos\frac{2\lfloor n/2\rfloor\pi}{n}\right)}, & \text{if } n > 5 \end{cases} \tag{3.31}$$

When all holding probabilities are identical, we have $a_2^* = a_1^*$. In this case, the optimal holding probability and the optimal SLEM is given as

$$\begin{aligned} a_2^* = a_1^* &= -\frac{\cos\frac{2\pi}{n} + \cos\frac{2\lfloor n/2\rfloor\pi}{n}}{3 - \cos\frac{2\pi}{n} - \cos\frac{2\lfloor n/2\rfloor\pi}{n}} \\ \mu_i^*(U) &= \frac{\cos\frac{2\pi}{n} - \cos\frac{2\lfloor n/2\rfloor\pi}{n}}{3 - \cos\frac{2\pi}{n} - \cos\frac{2\lfloor n/2\rfloor\pi}{n}}. \end{aligned} \tag{3.32}$$

Remark 3.9. The fastest mixing reversible Markov chain on a wheel graph mixes faster than the FMMC with identical holding probabilities: $\mu_i^*(U) > \mu_r^*(U)$. Furthermore, for $n \geq 5$, the mixing performance of the FMMC with identical holding probabilities

outperforms that of [28] that is given by

$$\mu_s^*(U) = \left(1 - \frac{1}{n}\right) \frac{\cos\frac{2\pi}{n} - \cos\frac{2\lfloor n/2\rfloor\pi}{n}}{2 - \cos\frac{2\pi}{n} - \cos\frac{2\lfloor n/2\rfloor\pi}{n}}. \tag{3.33}$$

For $n \leq 4$, the two methods yield exactly the same result. Note also that for both formulations considered in this chapter, the optimal SLEM, $\mu_i^*(U)$ and $\mu_r^*(U)$, converge to $2/3$ whereas the result of [28] converges to 1 as $n \to \infty$.

3.7.2. The Graph $K_n - K_n$

The graph $K_n - K_n$ is the union of two complete graphs connected by a bridge as illustrated in Figure 3.1b. Without loss of generality, let the nodes connecting these two complete graphs be labeled by n and $n+1$. Then the fastest mixing reversible Markov chain $K_{n,n} = [k_{ij}]$ on a $K_n - K_n$ graph for the stationary distribution in (3.13) has the following components

$$k_{ij} = \begin{cases} a_1, \text{if } i = j \text{ and } i, j \notin \{n-1, n\} \\ a_2, \text{if } i = j \text{ and } i, j \in \{n-1, n\} \\ b_1, \text{if } i \neq j,\ i, j \notin \{n, n+1\} \text{ and } (i,j) \in \mathcal{E} \\ b_2, \text{if } i \neq j,\ j \in \{n, n+1\} \text{ and } (i,j) \in \mathcal{E} \\ b_3, \text{if } i \neq j,\ i \in \{n, n+1\} \text{ and } (i,j) \in \mathcal{E} \\ b_4, \text{if } (i,j) \in \{(n, n+1), (n+1, n)\} \\ 0, \text{ if } (i,j) \notin \mathcal{E} \end{cases} \tag{3.34}$$

where $b_4 = 1 - a_2 - (n-1)b_3$. Note that the balance condition in (3.2) requires $b_3 = \frac{n}{n-1} b_2$. In order to determine the eigenvalues of $K_{n,n}$, let

$$T_1 = \begin{bmatrix} V & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & V \end{bmatrix}$$

where V is the $(n-1)\times(n-1)$ eigenvector matrix of $b_1\mathbf{1}\mathbf{1}^\top+(a_1-b_1)I_{n-1}$. The $2n-4$ eigenvalues of $T_1^{-1}K_{n,n}T_1$ are a_1-b_1, whereas the rest are obtained from those of the 4×4 matrix

$$K_r=\begin{bmatrix} 1-b_2 & \sqrt{n-1}b_2 & 0 & 0 \\ \frac{(n-1)^{1.5}}{n}b_2 & a_2 & b_4 & 0 \\ 0 & b_4 & a_2 & \frac{(n-1)^{1.5}}{n}b_2 \\ 0 & 0 & \sqrt{n-1}b_2 & 1-b_2 \end{bmatrix}.$$

With

$$T_2=\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix},$$

it is easy to note that the eigenvalues of $T_2^{-1}K_rT_2$ are obtained from those of

$$\begin{bmatrix} 1-b_2 & \sqrt{n-1}b_2 \\ \frac{(n-1)^{1.5}}{n}b_2 & 1-\frac{(n-1)^2}{n}b_2 \end{bmatrix}, \begin{bmatrix} 2a_2-1+\frac{(n-1)^2}{b_2} & \frac{(n-1)^{1.5}}{n}b_2 \\ \sqrt{n-1}b_2 & 1-b_2 \end{bmatrix}.$$

The matrix on the left has eigenvalues at 1 and $(1-b_2)\left(1+\frac{(n-1)^2}{n}\right)-\frac{(n-1)^2}{n}$. The spectrum of the second matrix consists of the second largest eigenvalue (in modulus). Hence, the optimal SLEM and transition/holding probabilities are calculated as

$$\mu_r^*(K_{n,n})=\frac{n^2+(2\sqrt{2}-5)n+1}{n^2+(1-2\sqrt{2})n+1} \tag{3.35}$$

$$a_2^*=0,\quad b_2^*=\frac{(2-\sqrt{2})n}{n^2+(1-2\sqrt{2})n+1} \tag{3.36}$$

The SLEM of $K_{n,n}$ depends only on the choice of a_2 and b_2. There is no restriction on a_1 and b_1 so long as they are non–negative and each row of $K_{n,n}$ sums up to one.

This implies that only the transition probability to the bridge node is important in the solution of the SLEM. However, the choice $b_1 = a_1 = \frac{1-b_2}{n-1}$ moves the $2n-4$ eigenvalues at $a_1 - b_1$ to zero.

3.7.2.1. The FMMC with Identical Holding Probabilities. The Markov chain under Assumption 3.2 requires $a_1 = a_2 = a$, $b_1 = b_2 = \frac{1-a}{n-1}$ and $b_3 = \frac{1-a}{n}$, In this case, the optimal mixing rate

$$\mu_i^*(K_{n,n}) = \frac{n^2 - 3n + 1 + \sqrt{n^4 + 2n^3 - 9n^2 + 6n + 1}}{2n(n-1)} \tag{3.37}$$

is obtained for the optimal choice of the identical holding probability of $a^* = 0$.

Remark 3.10. The optimal SLEM of the symmetric Markov chain on a $K_n - K_n$ graph in [28] is given by

$$\mu_s^*(K_{n,n}) = \frac{n - 4 + 2\sqrt{2}}{n + 2 - 2\sqrt{2}}. \tag{3.38}$$

Despite using a single parameter (a), we have $\mu_i^*(K_{n,n}) < \mu_s^*(K_{n,n})$ for $2 \leq n \leq 4$ while the symmetric SDP solution of [28] employing three parameters is superior for $n \geq 5$. However note also that $\mu_r^*(K_{n,n}) < \mu_s^*(K_{n,n})$ for all $n \geq 2$, i.e., the optimal reversible Markov chain mixes faster than its symmetric counterpart.

3.7.3. Alternative Transition Probability Dormulation for the Graph $K_n - K_n$:

In order to achieve faster mixing for the graph $K_n - K_n$, we modify our single parameter SDP formulation by revising the construction in Assumption 3.2 as follows.

Assumption 3.11. Let P be the transition probability matrix. For $n \geq 3$ and some $a \in [0, 1]$, suppose that the transition probabilities are of the following form:

(i) $P_{ii} = a$ for $i = 1, \ldots, n$

$$\text{(ii)}\ P_{ij} = \begin{cases} \frac{1-a}{2}, & \text{if } i \neq j,\ j = n, n+1 \text{ and } (i,j) \in \mathcal{E} \\ \frac{1-a}{2(N_i-1)}, & \text{if } i \neq j,\ j \neq n, n+1 \text{ and } (i,j) \in \mathcal{E} \\ 0 & \text{if } (i,j) \notin \mathcal{E} \end{cases}$$

Article (i) in Assumption 3.11 ensures that all vertices have identical holding probabilities, and Article (ii) guarantees that transition probability from a vertex to a vertex on the bridge is equal to the sum of transition probabilities from a vertex to any vertex that is not on the bridge. This set up corresponds to the stationary distribution

$$\pi' = \frac{1}{4n-4}[1, \ldots, 1, n-1, n-1, 1, \ldots, 1]. \tag{3.39}$$

The fastest mixing reversible Markov chain on a $K_n - K_n$ graph corresponding to the given stationary distribution is of the form (3.34) where $b_3 = \frac{b_2}{n-1}$. Following the same procedure as in the spectrum analysis of the original formulation, we compute the optimal SLEM and the transition/holding probabilities as

$$\mu_r'^*(K_{n,n}) = \sqrt{2}/2, \quad a_2^* = 0, \quad b_2^* = 0.5 \tag{3.40}$$

Once again, the SLEM of $K_{n,n}$ depends only on a_2 and b_2. There is no restriction on a_1 and b_1 so long as they are non–negative and each row of $K_{n,n}$ adds up to one. This implies that only the transition probability to the bridge node is important in the solution of the SLEM. However, the choice $b_1 = a_1 = \frac{1/2}{n-1}$ moves the $2n-4$ eigenvalues at $a_1 - b_1$ to zero.

Remark 3.12. In the modified case, the SDP formulation of identical holding probabilities leads to the same optimal mixing rate given in (3.40) for $a_1^* = a_2^* = a^* = 0$. Despite using a single parameter (a), the result, $\mu^* = \sqrt{2}/2$, based on our extended SDP formulation, is independent of the number of the vertices of the graph, whereas the symmetric SDP formulation of [28] employing three parameters tends to 1 as $n \to \infty$.

3.8. Chapter Summary & Discussion

In this chapter, we study two alternative methods of assigning transition probabilities to a Markov chain in order to optimize its mixing rate. In the first SDP formulation, there is a single transition probability parameter to be optimized (the holding probability of vertices) which yields to faster computation as opposed to the more general reversible Markov chain formulation corresponding to a stationary distribution that is proportional to the degree of vertices.

Exact analytical results verify that the two proposed formulations lead to better mixing for a star, a complete bipartite graph and a wheel graph than the optimal symmetric Markov chain. For distance transitive graphs, the symmetric SDP formulation and the ones considered herein lead to the same FMMC solution. The straightforward application of the identical holding probability SDP formulation results in faster mixing than the optimal symmetric $K_n - K_n$ graph, $2 \leq n \leq 4$, despite using a single parameter, while it is slower for $n \geq 5$. Even in this case, it has been analytically demonstrated that the reversible Markov chain formulation corresponding to the stationary distribution that is proportional to the degree of vertices outperforms that of the best symmetric Markov chain.

Finally, the single parameter SDP formulation has been extended for the $K_n - K_n$ graph resulting in a Markov chain whose SLEM is equal to $1/\sqrt{2}$ and is independent of the graph size. We also note that the single parameter solution is the same as the more general reversible formulation corresponding to the equilibrium distribution in (3.39).

From the above discussion, it is noted that we obtain better or equal mixing for the graphs under consideration with the reversible Markov chain formulation corresponding to a stationary distribution that is proportional to the degree of vertices. Although this seems to be valid for many other connected graphs including all with number of vertices less than six, it can not be generalized. To this end, consider the graph depicted in Figure 3.8. The fastest mixing reversible and the fastest mixing symmetric

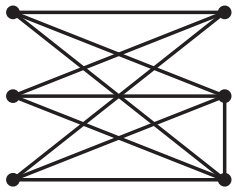


Figure 3.2. A 6–node undirected graph.

Markov chains are respectively given by

$$\begin{bmatrix} \frac{19}{68} & 0 & 0 & \frac{1}{4} & \frac{4}{17} & \frac{4}{17} \\ 0 & \frac{19}{68} & 0 & \frac{1}{4} & \frac{4}{17} & \frac{4}{17} \\ 0 & 0 & \frac{19}{68} & \frac{1}{4} & \frac{4}{17} & \frac{4}{17} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{3}{17} & \frac{3}{17} & \frac{3}{17} & 0 & \frac{4}{17} & \frac{4}{17} \\ \frac{3}{17} & \frac{3}{17} & \frac{3}{17} & 0 & \frac{4}{17} & \frac{4}{17} \end{bmatrix}, \begin{bmatrix} \frac{1}{3} & 0 & 0 & \frac{2}{9} & \frac{2}{9} & \frac{2}{9} \\ 0 & \frac{1}{3} & 0 & \frac{2}{9} & \frac{2}{9} & \frac{2}{9} \\ 0 & 0 & \frac{1}{3} & \frac{2}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{2}{9} & \frac{2}{9} & \frac{1}{3} & 0 & 0 \\ \frac{2}{9} & \frac{2}{9} & \frac{2}{9} & 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{2}{9} & \frac{2}{9} & \frac{2}{9} & 0 & \frac{1}{6} & \frac{1}{6} \end{bmatrix}$$

that lead to the SLEMs $\sqrt{2/17}$ and $1/3$. On the other hand, the fastest mixing Markov chain with identical holding probabilities is obtained for $a^* = 3/11$ and has a slightly greater SLEM, that is equal to $\sqrt{129}/33$.

Although the reversible formulation corresponding to a stationary distribution that is proportional to the degree of vertices yields better mixing for many graphs under consideration, the above example demonstrates that this claim is not always true. Future research should concentrate on gaining further insight into why the two proposed formulations tend to yield better analytical results than the symmetric formulation for certain classes of graphs. We hope that such research will also pave the way in developing alternative methods for the solution of the FMMC problem.

4. THE FASTEST MIXING MARKOV CHAIN ON A PATH

The fastest mixing Markov chain on undirected graphs is discussed in Chapter 3 and some novel results are derived for achieving the fastest mixing with two different formulations (FMMC with identical holding probabilities and fastest mixing reversible Markov chain with degree proportional stationary distributions). In this chapter, we discuss the FMMC problem on an important network topology, a path.

4.1. The Fastest Mixing Symmetric Markov Chain Formulation for a Path

The SDP formulation of the fastest mixing symmetric Markov chain problem on a path is as follows:

$$\begin{array}{ll} \text{minimize} & \mu(P) \\ \text{subject to} & P \geq 0, \quad P\mathbf{1} = \mathbf{1}, P = P^\top, \\ & P_{ij} = 0, \text{ if } |i-j| > 1. \end{array} \tag{4.1}$$

In [27], the authors derived the fastest mixing symmetric Markov chain on a path which is given by

$$S_n^* = \begin{bmatrix} 1/2 & 1/2 & & & \\ 1/2 & 0 & 1/2 & & \\ & \ddots & \ddots & \ddots & \\ & & 1/2 & 0 & 1/2 \\ & & & 1/2 & 1/2 \end{bmatrix} \tag{4.2}$$

and the optimal SLEM is shown to be equal to

$$\mu(S_n^*) = \cos\frac{\pi}{n}. \tag{4.3}$$

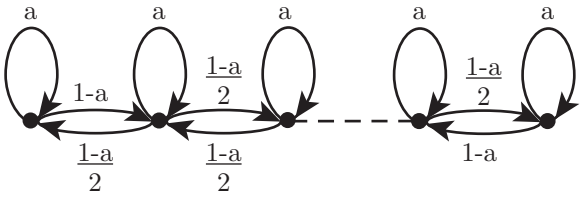


Figure 4.1. A directed path with identical holding probabilities.

In the above work, the authors deal with symmetric Markov chains. In Chapter 3, we have proposed two FMMC formulations on undirected graphs and derive the analytical expression of their SLEM. In this chapter, we consider the identical holding probability formulation for a path graph and obtain its SLEM which provides faster mixing than that of [27].

4.2. The FMMC with Identical Holding Probabilities Formulation for a Path

In our formulation, we let all vertices have the same holding probability, i.e., $P_{ii} = a$, and transition probabilities from a vertex to its neighbors are equal, i.e., $P_{ij} = \frac{1-a}{N_i}$ for $|i-j| = 1$ where N_i be the number of neighbors of vertex i. Note that the Markov chain is not symmetric in this formulation. The transition and holding probabilities are labeled for our setup in Figure 4.1. Let $P_{n,n}(a)$ be the Markov chain on a path with n vertices for our setup, i.e.,

$$P_{n,n}(a) = \begin{bmatrix} a & 1-a & & & & \\ \frac{1-a}{2} & a & \frac{1-a}{2} & & & \\ & & \ddots & & & \\ & & \frac{1-a}{2} & a & \frac{1-a}{2} & \\ & & & 1-a & a & \end{bmatrix}. \tag{4.4}$$

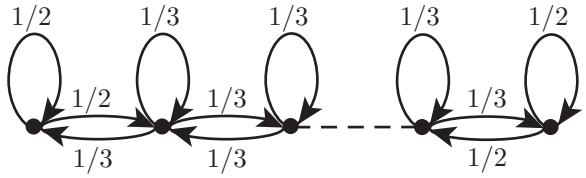


Figure 4.2. A directed path with equal transition probabilities.

The SDP formulation of the FMMC with identical holding probabilities for a path is as follows:

$$\begin{aligned} \underset{a}{\text{minimize}} \quad & \mu(\hat{P}) \\ \text{subject to} \quad & \hat{P}_{ij} = \begin{cases} a, & \text{if } i = j \\ \frac{1-a}{\sqrt{N_i N_j}}, & \text{if } |i-j| = 1 \\ 0, & \text{otherwise} \end{cases} \\ & 0 \leq a \leq 1 \end{aligned} \tag{4.5}$$

One of the objectives in this thesis is to derive the mathematical expression of the optimal holding probability, a^*, and the optimal SLEM, $\mu^*(P_{n,n}(a))$ as functions of n and to show that it is less than the SLEM of the fastest mixing symmetric Markov chain given by (4.3). We also investigate the Markov chains on a path with equal transition probabilities as shown in Figure 4.2, i.e, $P_{ij} = \frac{1}{N_i+1}$ for $|i-j| = 1$. This probability assignment is quite intuitive and its performance needs to be investigated as well. In this chapter, we not only derive bounds on the SLEM of this Markov chain, but also prove that its SLEM is larger than the optimal solution proposed as per (4.4).

4.3. Spectra of Markov Chain on a Path with Identical Holding Probabilities

In this section, we study the spectrum of the Markov chain on a path with identical holding probabilities and derive the optimal holding probability that yields the fastest mixing rate.

Theorem 4.1. *The spectrum of $P_{n,n}(a)$ consists of values $a + (1-a)\cos\frac{\pi k}{n-1}$, $0 \leq$*

$k \leq n-1$. *Furthermore,* $\mu_i(P_{n,n}(a)) = \max\left\{\left|a + (1-a)\cos\frac{\pi k}{n-1}\right|\right\}$, $1 \leq k \leq n-1$, *is minimum for*

$$a = a^* = \frac{1 - \cos\frac{\pi}{n-1}}{3 - \cos\frac{\pi}{n-1}} \tag{4.6}$$

and the optimal SLEM is given by

$$\mu_i^* = \frac{1 + \cos\frac{\pi}{n-1}}{3 - \cos\frac{\pi}{n-1}}. \tag{4.7}$$

Proof. Let s_k, $k = 0, \ldots, n-1$, be the inverse of the unordered eigenvalues of $P_{n,n}(a)$, without any specific order (i.e., λ_k and s_k^{-1} need not be equal). Let $x = [x_1,\ \ldots,\ x_n]^\top$ be an eigenvector of $P_{n,n}(a)$. With $s_k = s$, from the eigenvalue–eigenvector equation we get

$$x_1 = s(x_1 + (1-a)x_2) \tag{4.8}$$

$$x_j = s\left(\frac{1-a}{2}x_{j-1} + ax_j + \frac{1-a}{2}x_{j+1}\right), j = 2, \ldots, n-1 \tag{4.9}$$

$$x_n = s((1-a)x_{n-1} + ax_n) \tag{4.10}$$

In order to solve the above equations for s, the method of particular solutions will be applied. For $x_j = r^j$, (4.9) becomes $s(\frac{1-a}{2})r^2 + (sa-1)r + s(\frac{1-a}{2}) = 0$ and the roots of this quadratic equation are

$$r_{1,2}(s) = \frac{1 - sa \pm \sqrt{s^2(2a-1) - 2sa + 1}}{s(1-a)}. \tag{4.11}$$

The general solution of x_j is of the form

$$x_j = A(s)r_1^j(s) + B(s)r_2^j(s) \tag{4.12}$$

where $A(s)$ and $B(s)$ are functions of s. In order for (4.9) to be valid for all j, the boundary conditions $x_0 = x_2$ and $x_{n+1} = x_{n-1}$ must hold, which lead to the following

equations:

$$\begin{aligned} A(s)(r_1^2(s)-1)+B(s)(r_2^2(s)-1) &= 0 \\ A(s)r_1^{n-1}(s)(r_1^2(s)-1)+B(s)r_2^{n-1}(s)(r_2^2(s)-1) &= 0 \end{aligned} \tag{4.13}$$

For some values of s, if the above equations are satisfied, then the solution of (4.9) is as given in (4.11). If $r_1(s) = r_2(s)$, then the solution of the linear system (4.9) is identically zero. Thus, we seek the values of s that satisfy

$$r_1^{n-1}(s) = r_2^{n-1}(s) \quad \text{but} \quad r_1(s) \neq r_2(s). \tag{4.14}$$

Since $r_1(s)r_2(s) = 1$, we compute r_1 as a $2(n-1)$–th root of unity, i.e.,

$$r_1(s) = e^{\frac{j\pi k}{n-1}},\ 0 \leq k \leq n-1. \tag{4.15}$$

It follows easily from (4.11) that (4.15) holds when $s = s_k$ where

$$s_k^{-1} = a + (1-a)\cos\frac{\pi k}{n-1},\ 0 \leq k \leq n-1. \tag{4.16}$$

Consequently, the SLEM of $P_{n,n}(a)$ can be expressed as

$$\mu_i(P_{n,n}(a)) = \max\left\{\left|a + (1-a)\cos\frac{\pi k}{n-1}\right|\right\} \tag{4.17}$$

$$= \max\left\{\left|a + (1-a)\cos\frac{\pi}{n-1}\right|, |2a-1|\right\} \tag{4.18}$$

for $1 \leq k \leq n-1$. Note that $\mu_i(P_{n,n}(a))$ is minimized for $a = a^*$ in (4.6) and the optimal SLEM is μ_i^* as given by (4.7). □

The SLEM of the fastest mixing Markov chain derived in Theorem 4.1 and result of [27] in (4.3) are compared in the following theorem.

Theorem 4.2.

$$\mu_s^*(S_n) > \mu_i^*(P_{n,n}(a)) \quad \textit{for } n \geq 3. \tag{4.19}$$

Proof. For $n = 3$, $\mu^*(S_3) = \frac{1}{2} > \frac{1}{3} = \mu_i^*(P_{3,3}(a))$. We will show that (4.19) is satisfied for also $n \geq 4$. Let $f(n)$ be the function defined as follows

$$f(n) = \cos\frac{\pi}{n}\left(3 - \cos\frac{\pi}{n-1}\right) - \cos\frac{\pi}{n-1} - 1 \tag{4.20}$$

so that $f(n) > 0$ for $n \geq 4$ implies (4.19). From the Taylor series expansion of $\cos(x)$, we have

$$1 - \frac{1}{2}x^2 \leq \cos(x) \leq 1 - \frac{x^2}{2} + \frac{x^4}{4!}, \quad |x| \leq \pi \tag{4.21}$$

Using (4.21), we derive a lower bound for $f(n)$

$$f(n) \geq f_{a_1}(n) + f_{a_2}(n) + f_{a_3}(n) \tag{4.22}$$

where

$$f_{a_1}(n) = \frac{2\pi^2}{n^2(n-1)} \tag{4.23}$$

$$f_{a_2}(n) = \frac{\pi^2 - \pi^4/4}{n^2(n-1)^2} - \frac{1}{4!}\left(\frac{\pi}{n}\right)^4 - \frac{2}{4!}\left(\frac{\pi}{n-1}\right)^4 \tag{4.24}$$

$$f_{a_3}(n) = \frac{1}{4!}\left(\frac{\pi}{n}\right)^2\left(\frac{\pi}{n-1}\right)^2\left[\frac{1}{2}\left(\frac{\pi}{n}\right)^2 + \frac{1}{2}\left(\frac{\pi}{n-1}\right)^2 - \frac{1}{4!}\left(\frac{\pi}{n}\right)^2\left(\frac{\pi}{n-1}\right)^2\right]. \tag{4.25}$$

Note that $f_{a_3}(n) > 0$. Since $\frac{1}{n^2} < \frac{1}{(n-1)^2} \leq \frac{16}{9}\frac{1}{n^2}$ for $n \geq 4$, we obtain

$$f(n) > f_{a_1}(n) + \frac{\pi^2 - \pi^4/4}{n^2(n-1)^2} - \frac{3}{4!}\frac{16}{9}\frac{\pi^4}{n^2(n-1)^2}. \tag{4.26}$$

By multiplying both sides by $n^2(n-1)^2$:

$$n^2(n-1)^2 f(n) > (2-c)\pi^2(n-1) + \left[c\pi^2(n-1) + \pi^2 - \frac{17}{36}\pi^4\right]. \tag{4.27}$$

Note that the second term on the right hand side is positive for $c \geq 1.23$ which implies that $f(n) > \frac{0.77\pi^2}{n^2(n-1)} > 0$. □

4.4. The Fastest Mixing Reversible Markov Chains with Degree Proportional Stationary Distribution for a Path

Let R_n be the reversible Markov chain with degree proportional stationary distribution on a path. For $n = 4$, R_n is of the following form

$$R_4 = \begin{bmatrix} a_1 & 1-a_1 & 0 & 0 \\ \frac{1-a_1}{2} & a_2 & 1-a_2-\frac{1-a_1}{2} & 0 \\ 0 & 1-a_2-\frac{1-a_1}{2} & a_2 & \frac{1-a_1}{2} \\ 0 & 0 & 1-a_1 & a_1 \end{bmatrix}. \tag{4.28}$$

With

$$T = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \tag{4.29}$$

it is easy to note that the eigenvalues of $T^{-1}R_4T$ are obtained from those of

$$\begin{bmatrix} a_1 & 1-a_1 \\ \frac{1-a_1}{2} & \frac{1+a_1}{2} \end{bmatrix}, \begin{bmatrix} 2a_2 - \frac{a_1}{2} - \frac{1}{2} & \frac{1-a_1}{2} \\ 1-a_1 & a_1 \end{bmatrix}. \tag{4.30}$$

The first matrix has eigenvalues at 1 and $\frac{3a_1-1}{2}$. The second matrix consists of the

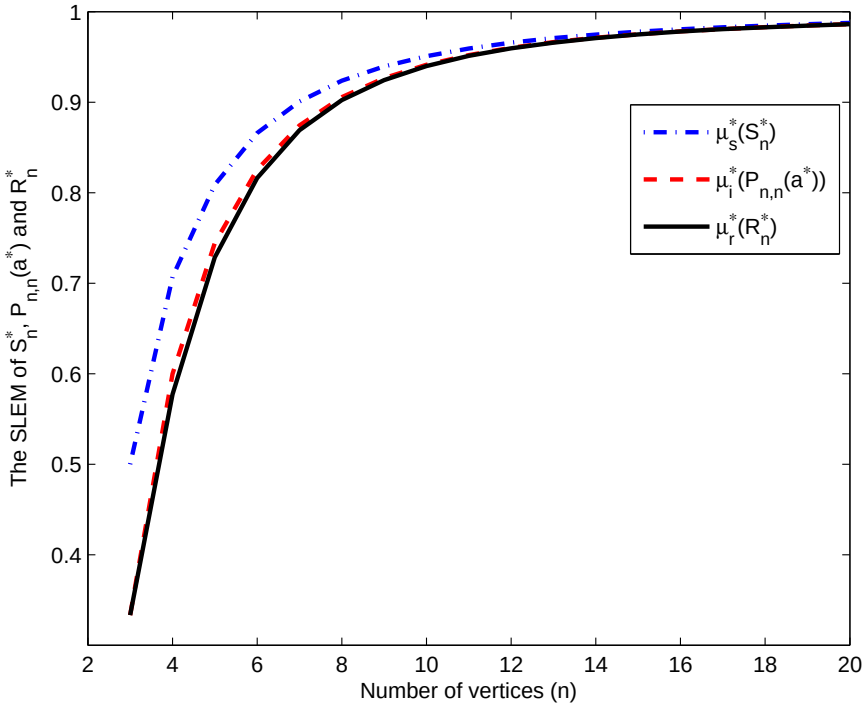


Figure 4.3. The SLEM of S_n^* and $P_{n,n}(a^*)$.

optimal SLEM. Hence the optimal SLEM and optimal parameters are calculated as

$$a_1 = 1/3, \quad a_2 = 1/6, \quad \mu^* = 1/\sqrt{3}. \tag{4.31}$$

For larger n, the numerical solution of the degree proportional reversible FMMC formulation can be obtained by SDP techniques of [33] as follows

$$R_n^* = \begin{bmatrix} 1/3 & 2/3 & & & & & \\ 1/3 & 1/6 & 1/2 & & & & \\ & 1/2 & 0 & 1/2 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1/2 & 0 & 1/2 & \\ & & & & 1/2 & 1/6 & 1/3 \\ & & & & & 2/3 & 1/3 \end{bmatrix}. \tag{4.32}$$

The SLEM of $S_n^*, P_{n,n}(a^*)$ and R_n^* are depicted in Figure 4.3 which verifies that the degree proportional reversible formulation outperforms that of [27] and (4.5).

4.5. Spectra of Markov Chain on a Path with Equal Weights

Let $L_{n,n}$ be the Markov chain on a path with with equal transition probabilities, i.e, $P_{ij} = \frac{1}{N_i+1}$ for $|i-j| = 1$:

$$L_{n,n} = \begin{bmatrix} 1/2 & 1/2 & & & \\ 1/3 & 1/3 & 1/3 & & \\ & & \ddots & & \\ & & 1/3 & 1/3 & 1/3 \\ & & & 1/2 & 1/2 \end{bmatrix} \tag{4.33}$$

Let L_{n,n_1}, $1 \leq n_1 < n$, be the $n_1 \times n_1$ sub–matrix of $L_{n,n}$ formed by deleting the last $n - n_1$ rows and columns, i.e.,

$$L_{n,n_1} = \begin{bmatrix} 1/2 & 1/2 & & & \\ 1/3 & 1/3 & 1/3 & & \\ & & \ddots & & \\ & & 1/3 & 1/3 & 1/3 \\ & & & 1/3 & 1/3 \end{bmatrix}. \tag{4.34}$$

Theorem 4.3. *The spectra of $L_{n,n}$, denoted by $\sigma(L_{n,n})$, is real. Furthermore, it can be computed as the union of the spectra of two reduced–order matrices as*

$$\sigma\left(L_{n,n}\right) = \begin{cases} \sigma\left(L_{n,\lfloor \frac{n}{2} \rfloor}\right) \bigcup \sigma\left(L_{n,\lceil \frac{n}{2} \rceil} + \frac{1}{3} q_{\lceil \frac{n}{2} \rceil} q_{\lfloor \frac{n}{2} \rfloor}^\top\right), & \text{if } n \text{ is odd} \\ \sigma\left(L_{n,\frac{n}{2}} - \frac{1}{3} q_{\frac{n}{2}} q_{\frac{n}{2}}^\top\right) \bigcup \sigma\left(L_{n,\frac{n}{2}} + \frac{1}{3} q_{\frac{n}{2}} q_{\frac{n}{2}}^\top\right), & \text{if } n \text{ is even} \end{cases} \tag{4.35}$$

where q_i is the basis vector that has 1 as its i–th component and zeros elsewhere; $\lfloor \frac{n}{2} \rfloor$ and $\lceil \frac{n}{2} \rceil$ denote the largest integer not greater than $\frac{n}{2}$ and smallest integer not less than $\frac{n}{2}$, respectively.

Proof. See Appendix B. □

Given that the Markov chain on a path with n vertices when all vertices have the same holding probability, a, is denoted by $P_{n,n}(a)$ in (4.4), let $P_{n,n_1}(a)$, $1 \leq n_1 < n$, be the $n_1 \times n_1$ sub–matrix of $P_{n,n}(a)$ formed by deleting the last $n - n_1$ rows and columns, i.e.,

$$P_{n,n_1}(a) = \begin{bmatrix} a & 1-a & & \\ \frac{1-a}{2} & a & \frac{1-a}{2} & \\ & & \ddots & \\ & & \frac{1-a}{2} & a \end{bmatrix}. \tag{4.36}$$

Note from Theorem 4.3 that the spectra of $P_{n,n}(a)$ for any $0 \leq a \leq 1$, denoted by $\sigma(P_{n,n}(a))$, is real. Furthermore as in Theorem 4.3, $\sigma(P_{n,n}(a))$ can be decomposed into two parts as follows.

Corollary 4.4. *$\sigma(P_{n,n}(a))$ can be computed as the union of the spectra of two reduced–order matrices as*

$$\sigma\left(P_{n,n}(a)\right) = \begin{cases} \sigma\left(P_{n,\lfloor \frac{n}{2} \rfloor}(a)\right) \bigcup \sigma\left(\bar{P}\right), & \text{if } n \text{ is odd} \\ \sigma\left(P_{n,\frac{n}{2}}(a) - \frac{1-a}{2} q_{\frac{n}{2}} q_{\frac{n}{2}}^{\top}\right) \bigcup \sigma\left(\bar{P}\right), & \text{if } n \text{ is even} \end{cases} \tag{4.37}$$

where $\bar{P} = P_{n,\lceil \frac{n}{2} \rceil}(a) + \frac{1-a}{2} q_{\lceil \frac{n}{2} \rceil} q_{\lfloor \frac{n}{2} \rfloor}^{\top}$.

The reduced–order matrix, whose spectrum consists of the SLEM of $L_{n,n}$, is given in Theorem 4.5.

Theorem 4.5.

$$\mu\left(L_{n,n}\right) = \begin{cases} \lambda_1\left(L_{n,\lfloor \frac{n}{2} \rfloor}\right), & \text{if } n \text{ is odd} \\ \lambda_1\left(L_{n,\frac{n}{2}} - \frac{1}{3} q_{\frac{n}{2}} q_{\frac{n}{2}}^{\top}\right), & \text{if } n \text{ is even} \end{cases} \tag{4.38}$$

Proof. See Appendix C. □

Corollary 4.6.

$$\mu\left(P_{n,n}\right) = \begin{cases} \lambda_1\left(P_{n,\lfloor\frac{n}{2}\rfloor}(a)\right), & \text{if } n \text{ is odd} \\ \lambda_1\left(P_{n,\frac{n}{2}}(a) - \frac{1-a}{2} q_{\frac{n}{2}} q_{\frac{n}{2}}^{\top}\right), & \text{if } n \text{ is even} \end{cases} \tag{4.39}$$

4.5.1. Effect of Number of Vertices on the Mixing Rate

It should not be surprising to determine that the mixing rate is strictly decreasing as the length of the path increases. Note that this fact can easily be verified for the solutions in (4.3) and (4.7). However, the situation is a bit more involved for the Markov chain with equal weights.

Theorem 4.7. *$\mu\left(L_{n,n}\right)$ is strictly increasing with increasing n.*

Proof. See Appendix D. □

4.6. Discussion on the Convergence Speed of the FMMC on a Path

In the previous section, we derived some results on the SLEM of $L_{n,n}$. These results and the results of the following theorem will be used in Theorem 4.9 to compare its SLEM with the optimal SLEM of $P_{n,n}(a)$ derived in (4.7).

Lemma 4.8. *(i) $\mu\left(P_{n,n}(\frac{1}{3})\right) > \mu_i^*\left(P_{n,n}(a)\right)$ for $n > 3$. (ii) The difference matrices $\tilde{P}_{n,\lfloor\frac{n}{2}\rfloor}(\frac{1}{3}) - \tilde{P}_{n,\lfloor\frac{n}{2}\rfloor}(a)$ and $\left(\tilde{P}_{n,\frac{n}{2}}(\frac{1}{3}) - \frac{1}{3} q_{\frac{n}{2}} q_{\frac{n}{2}}^{\top}\right) - \left(\tilde{P}_{n,\frac{n}{2}}(a) - \frac{1-a}{2} q_{\frac{n}{2}} q_{\frac{n}{2}}^{\top}\right)$ are positive definite for n odd and n even, respectively, where $\tilde{P}_{n,\frac{n}{2}}(a) = T^{-1} P_{n,\frac{n}{2}}(a) T$ is symmetric for $0 \leq a < \frac{1}{3}$ and $T = diag\{\sqrt{2}, 1, \ldots, 1\}$.*

Proof. It is straightforward to see that (4.18) is greater than (4.7) for $a = 1/3$ and $n > 3$. In order to prove (ii), we will consider two cases: n odd and n even. For n odd and $0 \leq a < 1/3$, the difference matrix $\tilde{P}_{n,\lfloor\frac{n}{2}\rfloor}(1/3) - \tilde{P}_{n,\lfloor\frac{n}{2}\rfloor}(a)$ is an M–matrix whose diagonal elements are $1/3 - a > 0$, non–diagonal elements are either $\frac{a-1/3}{2} < 0$

or $\frac{a-1/3}{\sqrt{2}} < 0$, and leading principal minors are $\frac{(1/3-a)^k}{2^{k-1}} > 0$, $k = 1, \ldots, \lfloor \frac{n}{2} \rfloor$. Hence $\tilde{P}_{n,\lfloor \frac{n}{2} \rfloor}(\frac{1}{3}) - \tilde{P}_{n,\lfloor \frac{n}{2} \rfloor}(a)$ is positive definite for $0 \leq a < 1/3$. For n even and $0 \leq a < 1/3$, the difference maxtrix is an M–matrix whose diagonal elements are either $1/3 - a > 0$ or $\frac{3}{2}(1/3 - a) > 0$, non–diagonal elements are either $\frac{a-1/3}{\sqrt{2}} < 0$ or $\frac{a-1/3}{2} < 0$, and the first $\frac{n}{2} - 1$ leading principal minors are $\frac{(1/3-a)^k}{2^{k-1}} > 0$, $k = 1, \ldots, \frac{n}{2} - 1$, and the determinant is $\frac{(1/3-a)^{\frac{n}{2}-1}}{2^{\frac{n}{2}-2}} > 0$. Hence $\left(\tilde{P}_{n,\frac{n}{2}}(\frac{1}{3}) - \frac{1}{3} q_{\frac{n}{2}} q_{\frac{n}{2}}^\top\right) - \left(\tilde{P}_{n,\frac{n}{2}}(a) - \frac{1-a}{2} q_{\frac{n}{2}} q_{\frac{n}{2}}^\top\right)$ is positive definite for $0 \leq a < 1/3$. □

The following theorem states that for $a = a^*$, (4.4) mixes faster than (4.33).

Theorem 4.9.

$$\mu(L_{n,n}) > \mu_i^*(P_{n,n}(a^*)). \tag{4.40}$$

Proof. For $n = 3$, $\mu(L_{3,3}) = \frac{1}{2} > \frac{1}{3} = \mu_i^*\left(P_{3,3}(a^*)\right)$. From Lemma 4.8, we have $\mu\left(P_{n,n}(\frac{1}{3})\right) > \mu^*\left(P_{n,n}(a)\right)$ for $n > 3$. In order to prove the theorem, we will show that $\mu(L_{n,n}) \geq \mu(P_{n,n}\left(\frac{1}{3}\right))$ for $n > 3$.

Let us consider two cases: (i) n is odd, (ii) n is even.

Case (i) n is odd: From Theorem 4.5, we have $\mu(L_{n,n}) = \lambda_1(L_{n,\lfloor \frac{n}{2} \rfloor})$ and from Corollary 4.6, we have $\mu(P_{n,n}\left(\frac{1}{3}\right)) = \lambda_1(P_{n,\lfloor \frac{n}{2} \rfloor})\left(\frac{1}{3}\right)$. Therefore, mixing rate comparison of $L_{n,n}$ and $P_{n,n}\left(\frac{1}{3}\right)$ can be done by comparing the largest eigenvalues of $L_{n,\lfloor \frac{n}{2} \rfloor}$ and $P_{n,\lfloor \frac{n}{2} \rfloor}\left(\frac{1}{3}\right)$. Let $T_2 = diag\{\sqrt{2}, 1, \ldots, 1\}$ so that $P_{n,\lfloor \frac{n}{2} \rfloor}\left(\frac{1}{3}\right)$ is similar to the symmetric matrix $\tilde{P}_{n,\lfloor \frac{n}{2} \rfloor}\left(\frac{1}{3}\right) = T_2^{-1}\left(P_{n,\lfloor \frac{n}{2} \rfloor}\left(\frac{1}{3}\right)\right) T_2$ given by

$$\tilde{P}_{n,\lfloor \frac{n}{2} \rfloor}\left(\frac{1}{3}\right) = \begin{bmatrix} 1/3 & \frac{\sqrt{2}}{3} & & \\ \frac{\sqrt{2}}{3} & 1/3 & 1/3 & \\ & & \ddots & \\ & & 1/3 & 1/3 \end{bmatrix}. \tag{4.41}$$

Let

$$x = \left[x_1,\ x_2,\ \ldots,\ x_{\lfloor \frac{n}{2} \rfloor}\right]^\top \tag{4.42}$$

be the unit length eigenvector of $\tilde{P}_{n,\lfloor \frac{n}{2} \rfloor}\left(\frac{1}{3}\right)$ corresponding to the largest eigenvalue, $\lambda_1^* = \lambda_1\left(\tilde{P}_{n,\lfloor \frac{n}{2} \rfloor}\left(\frac{1}{3}\right)\right)$.

From the eigenvalue–eigenvector equation, we have $x_1 = cx_2$ in (4.42) where

$$c = \frac{\sqrt{2}}{3\lambda_1^* - 1}. \tag{4.43}$$

Since $\tilde{P}_{n,\lfloor \frac{n}{2} \rfloor}\left(\frac{1}{3}\right)$ is a real symmetric matrix, its largest eigenvalue can be computed from the Rayleigh–Ritz Theorem ([40] Theorem 4.2.2 pp 176–177) as

$$\lambda_1^* = \max_{||z||=1} z^\top \left(\tilde{P}_{n,\lfloor \frac{n}{2} \rfloor}(1/3)\right) z \tag{4.44}$$

$$= x^\top \left(\tilde{P}_{n,\lfloor \frac{n}{2} \rfloor}(1/3)\right) x. \tag{4.45}$$

Let $y = \tilde{c}\left[\frac{2}{\sqrt{3}}cx_2,\ x_2,\ \ldots,\ x_{\lfloor \frac{n}{2} \rfloor}\right]^\top$ where $\tilde{c} = \dfrac{1}{\sqrt{1 + \frac{c^2}{3}x_2^2}}$ and x_i, $i = 2, \ldots, \lfloor \frac{n}{2} \rfloor$, are the components of x. Note that y has unit length. Using the Rayleigh–Ritz Theorem together with (4.45), we obtain

$$\begin{aligned}
\lambda_1(\tilde{L}_{n,\lfloor \frac{n}{2} \rfloor}) &= \max_{z \in R^{\lfloor \frac{n}{2} \rfloor}, ||z||=1} z^\top \tilde{L}_{n,\lfloor \frac{n}{2} \rfloor} z && (4.46)\\
&\geq y^\top \tilde{L}_{n,\lfloor \frac{n}{2} \rfloor} y && (4.47)\\
&= \tilde{c}^2\left(\frac{c^2 x_2^2}{3} + \lambda_1^*\right) && (4.48)\\
&= \lambda_1^* \frac{1 + \dfrac{c^2 x_2^2}{3\lambda_1^*}}{1 + \frac{c^2 x_2^2}{3}} && (4.49)
\end{aligned}$$

Since $\lambda_1^* < 1$, we conclude that $\lambda_1(\tilde{L}_{n,\lfloor \frac{n}{2} \rfloor}) > \lambda_1^* = \lambda_1\left(\tilde{P}_{n,\lfloor \frac{n}{2} \rfloor}\left(\frac{1}{3}\right)\right)$ which com-

pletes the proof for n odd.

Case (ii) n **is even:** From Theorem 4.5 and Corollary 4.6, note that we have $\mu(L_{n,n}) = \lambda_1(L_{n,\frac{n}{2}} - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^\top)$ and $\mu(P_{n,n}\left(\frac{1}{3}\right)) = \lambda_1\left(P_{n,\frac{n}{2}}\left(\frac{1}{3}\right) - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^\top\right)$. Therefore, mixing rate comparison of $L_{n,n}$ and $P_{n,n}\left(\frac{1}{3}\right)$ can be done by comparing the largest eigenvalues of $L_{n,\frac{n}{2}} - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^\top$ and $P_{n,\frac{n}{2}}\left(\frac{1}{3}\right) - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^\top$. The same procedure above can be applied to show that $\lambda_1(\tilde{L}_{n,\frac{n}{2}} - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^\top) > \lambda_1\left(\tilde{P}_{n,\frac{n}{2}}\left(\frac{1}{3}\right) - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^\top\right)$ which completes the proof for n even. □

4.7. Chapter Summary

In this chapter, we have proposed a new formulation for optimizing the mixing rates of Markov chains on a path. By deriving the analytical solution, we have proved that the proposed solution outperforms that of [27] and the Markov chain on a path with equal weights. We have also examined the effect of the number of vertices on the SLEM, and have shown that the mixing rates for different scenarios under investigation in this chapter are strictly decreasing as the length of the path increases.

5. THE EFFECT OF NON-UNIFORM DELAY ON THE CONVERGENCE SPEED OF CONSENSUS ALGORITHMS

In Chapters 3 and 4, we have investigated the mixing rates of Markov chains that correspond to the convergence rates of the consensus algorithm for networks without delay. In this chapter, we study the convergence properties and rates for delayed networks and derive some results on possible non–detrimental effect of delay on the convergence rate.

5.1. Delayed Consensus Model and Mathematical Preliminaries

In communication networks where communication delay is unavoidable, the distributed consensus algorithm can be mathematically expressed as

$$x_i(t+1) = w_{ii}(t)x_i(t) + \sum_{j=1, j\neq i}^{n} w_{ij}(t)x_j(t-\tau_{ij}) \tag{5.1}$$

where $w_{ij}(t) \geq 0$ is the averaging coefficient; and τ_{ij} is the amount of delay in data transmission between nodes j and i. The following conditions are assumed throughout the chapter [5]:

Assumption 5.1. (i) $w_{ii}(t) > \xi$, $\forall\ i = 1, 2, \ldots, n$, $\forall\ t \geq 0$ for some constant $\xi > 0$.
(ii) $w_{ij}(t) \in [\xi, 1] \cup \{0\}$, $\forall\ i, j = 1, 2, \ldots, n$, $\forall\ t \geq 0$.
(iii) $\sum_{j=1}^{n} w_{ij}(t) = 1$, $\forall\ i = 1, 2, \ldots, n$, $\forall\ t \geq 0$.
(iv) $\tau_{ij} \leq \tau_{max}$, $\forall\ i, j = 1, 2, \ldots, n$.

Article (i) in Assumption 5.1 guarantees that nodes use their own data in their updates whereas the second condition requires that the received data from a neighbor should be used with strictly positive weighting. Article (iii) in Assumption 5.1 induces that the sum of the weighting coefficients for each node is equal to one. Finally Article (iv) excludes unbounded delay in the context of this chapter by restricting the maximum delay amount to be τ_{max}.

When there is no communication delay, i.e. $\tau_{ij} = 0$, the matrix representation of (5.1) is given by

$$x(t+1) = W(t)x(t) \tag{5.2}$$

where the system matrix $W(t) \in \mathbb{W}$ belongs to a special class of so called row–stochastic averaging matrices under the conditions of Articles (i)–(iii) in Assumption 5.1 that are assumed to hold throughout this chapter.

Delay is well known to deteriorate the performance of a system in general; furthermore it can even lead to instability if not compensated properly. The focus in this chapter is to study the effect of non–uniform delay on the convergence rate of the distributed consensus algorithm in (5.1) using mathematical tools of stochastic matrices, ergodicity and graph theory. In the rest of this section, we examine the convergence properties and the effect of delay on convergence rate.

5.2. Convergence Analysis for Uniform Delay and Fixed Topology

Suppose that there exists a uniform and fixed amount of delay between all nodes, i.e., $\tau_{ij}(t) = \tau$, $\forall i,\ j,\ t$. Then, (5.1) is re–expressed as

$$x(t+1) = W_D x(t) + (W - W_D)x(t-\tau) \tag{5.3}$$

where $W_D = diag\{w_{11}, w_{22}, \ldots, w_{nn}\}$. Let the augmented state vector $\hat{x}$ be defined as

$$\hat{x}(t) = [x(t+\tau)^\top, x(t+\tau-1)^\top, ..., x(t)^\top]^\top.$$

Then (5.3) can be re-written in the form

$$\hat{x}(t+1) = \hat{W}(t)\hat{x}(t) \tag{5.4}$$

where $\hat{W}$ is given by

$$\hat{W} = \begin{bmatrix} W_D & 0 & \dots & 0 & W - W_D \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \dots & I & 0 \end{bmatrix} \tag{5.5}$$

Lemma 5.2. *Suppose that at least one node exists in the network from which the rest of the nodes are accessible. Then consensus in the network is achieved despite the existence of bounded uniform delay τ.*

Proof. Let there be a node in the network from which the rest of the nodes are accessible. From [2], W^{n-1} is known to be scrambling. We now show that the network achieves consensus by proving that the $((\tau+1)n-1)$–th power of $\hat{W}$ is scrambling. For mathematical convenience, denote the (i,j)–th block of $\hat{W}^\tau$ by $\hat{W}_{\tau,ij}$ that can be computed as

$$\hat{W}_{\tau,ij} = \begin{cases} \bar{f}_{i,j}(W) & i < j \\ f_{i,1}(W_D) & j = 1,\, i \leq \tau \\ 0 & j \leq i,\, j \neq 1 \\ I & j = 1,\, i = \tau + 1 \end{cases} \tag{5.6}$$

where $f_{i,j}(W)$ and W have equal adjacency matrices; and $\bar{f}_{i,j}(W)$ and W have equal adjacency matrices except that $\bar{f}_{i,j}(W)$ has zero diagonal elements. In the sequel, $g_{i,j}(W)$, $\bar{g}_{i,j}(W)$, $h_{i,j}(W)$ and $\bar{h}_{i,j}(W)$, are defined similarly.

Using (5.6), $\hat{W}^{2\tau+1} = [\hat{W}_{2\tau+1,ij}]$ is computed as

$$\hat{W}_{2\tau+1,ij} = \begin{cases} \bar{g}_{i,j}(W^2) & i < j \\ g_{i,1}(W) & j = 1 \\ \bar{g}_{i,j}(W) & j \neq 1,\, j \leq i \end{cases} \tag{5.7}$$

From (5.7), one can deduce that $\hat{W}^{2\tau+1}$ is scrambling given W is.

Suppose that W is non–scrambling. Then the sub–blocks of the $((\tau+1)n-1)$–th power of $\hat{W}$ are calculated as

$$\hat{W}_{(\tau+1)n-1,ij} = \begin{cases} \bar{h}_{i,j}^{(n)}(W) & i < j \\ h_{i,1}(W^{n-1}) & j = 1 \\ \bar{h}_{i,j}^{(n-1)}(W) & j \leq i,\, j \neq 1 \end{cases} \tag{5.8}$$

where $\bar{h}_{i,j}^{(k)}(\cdot)$ is a polynomial of order k. One can note that each of the first column components of (5.8) has the same adjacency as W^{n-1}. If at least one node exists in the network from which the rest of the nodes are accessible, the matrix W^{n-1} is scrambling [2]. Hence, from (5.8), the $((\tau+1)n-1)$–th power of $\hat{W}$ is also scrambling. Therefore we can conclude that the network achieves consensus under bounded uniform delay. □

5.3. Convergence Analysis for Uniform Delay and Varying Topologies

Consider a network with N different topologies characterized by averaging matrices $\mathbb{W} = \{W_1, W_2, \ldots, W_N\}$. Suppose that uniform and a fixed amount of delay exists between nodes. Let $\hat{\mathbb{W}} = \{\hat{W}_1, \ldots, \hat{W}_N\}$ consist of the corresponding system matrices for the delayed systems.

Lemma 5.3. *Given the set of averaging matrices, $\mathbb{W} = \{W_1, \ldots, W_N\}$, suppose that the network consists of a (possibly different for each topology) node in the network from which the rest of the nodes are accessible. Then consensus is achieved in the delayed*

network for $\tau_{ij} = \tau$ and arbitrary switching of the topologies.

Proof. Let π_k, $k \geq 1$ be sets that consist of all possible k matrix products from the set $\hat{\mathbb{W}}$. It can be seen that all of the elements of π_K are scrambling for some $K \leq (\tau+1)n-1$, and therefore the network achieves consensus. □

5.4. Non–Uniform Delay Varying Topology Networks

Each node may receive data from other nodes with different and varying amounts of delay. Convergence analysis of the consensus algorithm for non–uniform delay networks is crucial since uniform delay does not exist in practical systems.

Theorem 5.4. *Given the set of averaging matrices, $\mathbb{W} = \{W_1, \ldots, W_N\}$, suppose that the network consists of a (possibly different for each topology) node from which the rest of the nodes are accessible. Then consensus is achieved for arbitrarily varying bounded delay τ_{ij}, $i, j = 1, 2, \ldots, n$, $i \neq j$ and arbitrary switching of the topologies.*

Proof. First let us consider the case for non-uniform fixed amount of delay which will later be extended to varying delay and network topologies. Under this setting, the consensus algorithm (5.1) can be put into the following matrix form for networks with non–uniform delay [24]

$$\hat{W} = \begin{bmatrix} W_D + W_{\bar{D},0} & W_{\bar{D},1} & \ldots & W_{\bar{D},\tau_{max}} \\ I & 0 & \ldots & 0 \\ & \ddots & & \\ 0 & \cdots & I & 0 \end{bmatrix} \tag{5.9}$$

where the (i, j)–th element of $W_{\bar{D},l}$, $l = 0, 1, \ldots, \tau_{max}$ is given by

$$[W_{\bar{D},l}]_{ij} = \begin{cases} w_{ij} & \text{if } \tau_{ij} = l \text{ and } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{5.10}$$

so that $W_{\bar{D},0} + \ldots + W_{\bar{D},\tau_{max}} = W - W_D$ satisfied. The minimum power that makes the system matrix scrambling is no more than $(\tau_{max}+1)n-1$ with the equality if delay is uniform and $\tau_{ij} = \tau_{max}$. Since $\hat{W}^{(\tau_{max}+1)n-1}$ is guaranteed to be scrambling under Assumption 5.1, consensus follows for non–uniform fixed amount of delay and the given network topology.

In order to complete the proof for arbitrarily varying topology networks and varying delays, let $\tilde{\pi}_k$, $k \geq 1$ be the sets that consist of all possible matrix products of length k from the set of all matrices of the form (5.9) for all delay combinations and different topologies corresponding to $\mathbb{W} = \{W_1, \ldots, W_N\}$. Note that $\tilde{\pi}_1$ consists of $N(\tau_{max}+1)^{n^2-n}$ matrices. Analogous to the proof of Lemma 3, it can be seen that all of the elements of $\tilde{\pi}_K$ are scrambling for some $K \leq (\tau_{max}+1)n-1$. □

Remark 5.5. Under the conditions of Theorem 5.4, we note that consensus is achieved despite the existence of bounded non–uniform and varying delay. While we rely on properties of scrambling matrices in proving the end result, the stated conditions in the theorem are relatively easy to check, i.e., one needs to ensure that the maximum amount of delay is bounded, and that the network consists of a (possibly different for each topology) node from which the rest of the nodes are accessible.

5.5. The Delay Effect on the Rate of Convergence

In mathematical system theory, delay is well known to deteriorate system performance in general and may even lead to an unstable system if not compensated properly. Although it is noted that algorithm convergence is not adversely affected from bounded delay by Theorem 5.4, its effect on the convergence rate must be examined. To this end, we first consider a simple example to demonstrate that delay does not necessarily deteriorate the rate of convergence.

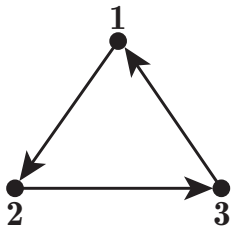


Figure 5.1. A simple network of three nodes.

Example 5.6. Consider the system matrix

$$W = \frac{1}{1000}\begin{bmatrix} 924 & 0 & 76 \\ 52 & 948 & 0 \\ 0 & 507 & 493 \end{bmatrix}. \tag{5.11}$$

for the network given in Figure 5.1.

For the two–step delayed and the nominal systems, we have $\lambda_2(\hat{W}) = 0.7722 < \lambda_2(W) = 0.8613$, which implies that the delayed system converges faster. The simulation results for the nominal and the delayed system are depicted in Figure 5.2 for the initial state values $x(0) = [1,\ 0.9,\ 0.7]^\top$.

Although Example 5.6 illustrates that delay does not necessarily deteriorate the rate of convergence for certain choice of the weighting coefficients, we know that this is not true in general. In the sequel, we examine a class of network topologies for which the convergence rates of the delayed and non–delayed consensus algorithms are equal to each other (regardless of the weighting coefficients). Some examples of the topologies under consideration are depicted in Figure 5.3. The topology in Figure 5.3a arises in the problem of vehicle platooning where each vehicle is trying to follow the preceding at a safe distance. Figure 5.3b depicts the *Master–Slave topology* in which the slaves receive data from the master. This topology is widely used in clock synchronization problems, where the master clock dictates its local clock frequency with its neighbors, namely slave clocks. Figure 5.3c depicts the *hierarchical topology*, where the nodes receive the leading node's data directly or indirectly. This can be thought as a multi–

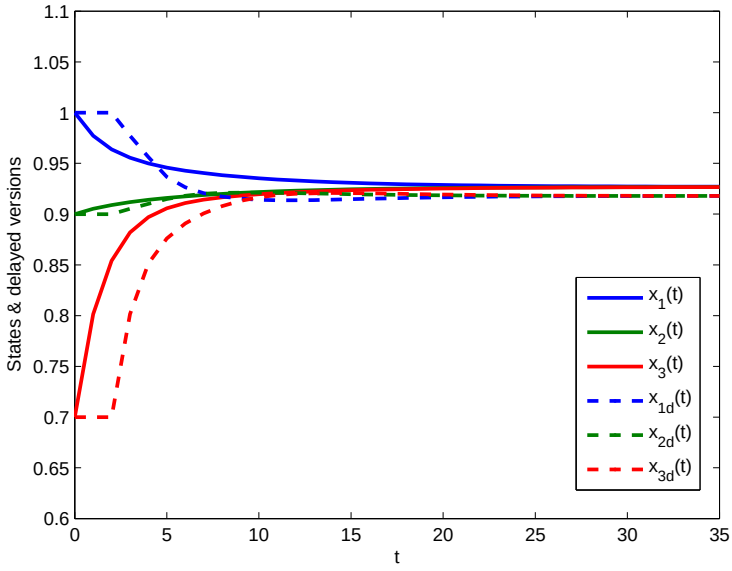


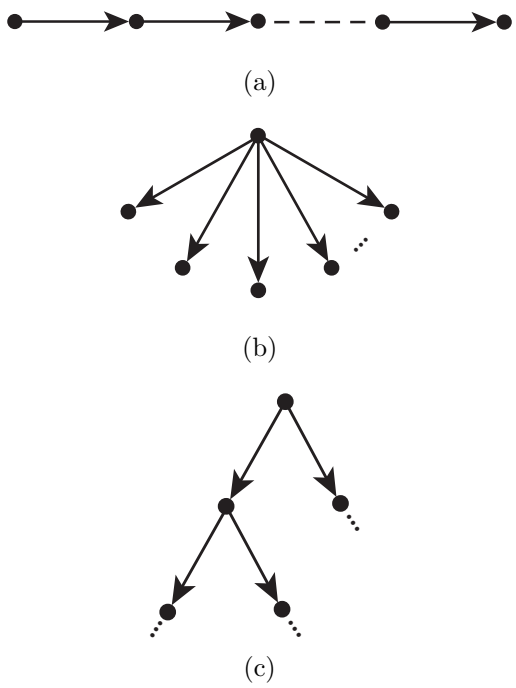
Figure 5.2. State evolution for the nominal and the delayed system in Example 5.6 ($x_i(t)$: nominal system states, $x_{id}(t)$: delayed system states).

level master–slave network where slaves of the n–th level are the masters of (n+1)–th level. The following theorem states that non–uniform delay is not detrimental to the convergence rate for the topologies of interest in this section.

Theorem 5.7. *For directed acyclic graphs, the nominal and delayed consensus algorithms (5.1) have the same non–zero spectra at each iteration step for arbitrarily varying non–uniform bounded delay.*

Proof. The adjacency matrix of a directed acyclic graph can be transformed into an upper (or equivalently lower) triangular matrix with proper vertex numbering. Without loss of generality, it is assumed that the system matrix is lower triangular throughout the proof.

We first consider the non–uniform fixed delay among nodes. In this case, the system matrix can be expressed in the form (5.9) whose characteristic polynomial is



(a)

(b)

(c)

Figure 5.3. Some examples of directed acyclic graphs: (a) chain topology, (b) master–slave topology, (c) hierarchical topology.

given by

$$|\lambda I - \hat{W}| = \begin{vmatrix} \lambda I - (W_D + W_{\bar{D},0}) & -W_{\bar{D},1} & \dots & -W_{\bar{D},\tau_{max}} \\ -I & \lambda I & 0 & 0 \\ & \ddots & & \\ 0 & \cdots & -I & \lambda I \end{vmatrix} \tag{5.12}$$

By using the properties of the determinant for block matrices, (5.12) can be simplified to a lower dimensional matrix determinant:

$$|\lambda I - \hat{W}| = \lambda^n \begin{vmatrix} \lambda I - (W_D + W_{\bar{D},0}) & -W_{\bar{D},1} & \dots & -W_{\bar{D},\tau_{max}-1} - \frac{1}{\lambda} W_{\bar{D},\tau_{max}} \\ -I & \lambda I & 0 & 0 \\ & \ddots & & \\ 0 & \cdots & -I & \lambda I \end{vmatrix} \tag{5.13}$$

By applying the above procedure recursively, we obtain

$$|\lambda I - \hat{W}| = \lambda^{n\tau_{max}} \left| \lambda I - \left(W_D + W_{\bar{D},0} + \frac{1}{\lambda} W_{\bar{D},1} + \ldots + \frac{1}{\lambda^{\tau_{max}}} W_{\bar{D},\tau_{max}} \right) \right|. \tag{5.14}$$

For the network topologies that are covered in the statement of the Theorem, we note that $W_D + W_{\bar{D},0} + \frac{1}{\lambda}W_{\bar{D},1} + \ldots + \frac{1}{\lambda^{\tau_{max}}}W_{\bar{D},\tau_{max}}$ and W (the system matrix without delay) have the same set of eigenvalues. Since the delayed system has the zero eigenvalue with multiplicity $n\tau_{max}$, the nominal and the delayed systems have the same non–zero spectra.

In order to complete the proof for arbitrarily varying delays, let $\tilde{\pi}_k$, $k \geq 1$ be the sets that consist of all possible matrix products of length k from the set of all matrices of the form (5.9) for all delay combinations. Note that $\tilde{\pi}_1$ consists of $(\tau_{max}+1)^{n^2-n}$ matrices. We now consider two subcases: $1 \leq k < \tau_{max}$ and $k \geq \tau_{max}$. For the latter case, any matrix $p_k(\hat{W})$ in $\tilde{\pi}_k$, can be expressed in the following form

$$p_k(\hat{W}) = \begin{bmatrix} W_D^k + * & * & \ldots & * \\ W_D^{k-1} + * & * & \ldots & * \\ \vdots & \vdots & \ddots & \vdots \\ W_D^{k-\tau_{max}} + * & * & \ldots & * \end{bmatrix} \tag{5.15}$$

where $*$ above denotes an upper–triangular matrix with zero diagonal elements. By applying the same recursive procedure as in the fixed delay case, we obtain

$$\left|\lambda I - p_k(\hat{W})\right| = \lambda^{n\tau_{max}} \left|\lambda I - \left(W_D^k + *\right)\right| \tag{5.16}$$

which implies that the non–zero spectra of $p_k(\hat{W})$ and W^k are equal to each other. For $1 \leq k < \tau_{max}$, the first k sub–rows of $p_k(\hat{W})$ has the form

$$\begin{bmatrix} W_D^k + * & * & \ldots & * \\ W_D^{k-1} + * & * & \ldots & * \\ \vdots & \vdots & \ddots & \vdots \\ W_D + * & * & \ldots & * \end{bmatrix} \tag{5.17}$$

whereas the last $\tau_{max} + 1 - k$ sub–rows is equal to $[I_{(\tau_{max}+1-k)n}\ 0]$, where $I_{(\tau_{max}+1-k)n}$ is the identity matrix of length $(\tau_{max} + 1 - k)n$. The same recursive procedure as above

can be applied to $p_k(\hat{A})$ in this case as well to obtain the desired result. □

Remark 5.8. From Theorem 5.7, we note that the non–zero spectra of the nominal system and the version for arbitrarily varying non–uniform bounded delay, are equal to each other at each iteration. Hence, we conclude that bounded non–uniform varying delay does not adversely affect convergence rate for the network topologies under consideration in this chapter, *regardless of the amount of delay and the choice of the weighting coefficients*. As can be expected this is not true for all topologies under which consensus is achieved. To this end, consider the fully connected network where each node uses equal averaging coefficients: $1/n$. It can be easily noted that the system matrix has the zero eigenvalue with multiplicity $n-1$; therefore consensus is achieved in a single iteration (dead–beat system). On the other hand, the delayed system has a non–zero second largest eigenvalue.

5.6. Chapter Summary

In this chapter, the performance of averaging based deterministic consensus under delayed information has been studied. It is proven mathematically that the rate of convergence of the distributed consensus algorithm is not reduced despite delay for directed acyclic graphs such as the master-slave topology, chain topology and hierarchical topology. This result holds regardless of the values of the averaging coefficients so long as they satisfy Assumption 5.1. Although we have focused on the case where delay does not adversely effect the convergence rate, networks topologies where convergence rate is always degraded are to be explored.

6. CONCLUSION & FUTURE WORK

This dissertation is concerned with the convergence rate analysis and optimization of distributed consensus algorithms in networks with and without delay. In the first part of the thesis, the convergence rate of the averaging based consensus algorithm is related to the mixing rate of the Markov chains and we have studied two alternative methods of assigning transition probabilities to a Markov chain in order to optimize its mixing rate. The Markov chains in both formulations are not symmetric, yet they can be transformed into symmetric matrices.

In the first SDP formulation, there is a single transition probability parameter to be optimized (which is the holding probability of vertices) that yields to faster computation than the more general reversible Markov chain formulation corresponding to a stationary distribution that is proportional to the vertex degrees. Exact analytical results verifiy that the both proposed formulations lead to faster mixing for a star, a complete bipartite graph and a wheel graph than the optimal symmetric Markov chain that is studied in the literature. For distance transitive graphs, the symmetric SDP formulation and the ones considered in the thesis lead to the same FMMC solution. Although there is a single optimization parameter, the results of the identical holding probability SDP formulation results in faster mixing than the optimal symmetric $K_n - K_n$ graph for $2 \leq n \leq 4$, whereas it is slower for $n \geq 5$. Even in this case, it has been analytically verified that the reversible Markov chain formulation corresponding to the stationary distribution that is proportional to the vertex degrees outperforms that of the fastest mixing symmetric Markov chain. Moreover, the single parameter SDP formulation has been extended for the $K_n - K_n$ graph resulting in a Markov chain whose SLEM is equal to $1/\sqrt{2}$ and is independent of the graph size. We also note that the single parameter solution and the more general reversible formulation corresponding to the equilibrium distribution in (3.39) results in the same solution. From all of the analysis, it is noted that we obtain better or equal mixing for the graphs under consideration with the reversible Markov chain formulation corresponding to a stationary distribution that is proportional to the degree of vertices. Although

this seems to be valid for many other connected graphs (including all with number of vertices less than six), it can not be generalized. To this end, a numerical example has been given to emphasise that fastest mixing symmetric Markov chain has a better mixing rate. Therefore, we conclude that there is no single formulation that yields to the best results for all topologies. Future research should concentrate on gaining further insight into why the two proposed formulations tend to yield better analytical results than the symmetric formulation for certain classes of graphs. We hope that our research will also pave the way in developing alternative methods for the solution of the FMMC problem.

Another contribution of this thesis is the analytical solution of the FMMC with identical holding probability problem for the path networks. This expression is compared with the existing solutions in the literature and the Markov chain on a path with equal weights in order to prove that the proposed solution yields to faster mixing than others. The effect of the number of vertices on the mixing rate of Markov chains on a path is also examined and it is proven that the mixing rates for different scenarios under investigation in the thesis are strictly decreasing as the length of the path increases.

We have also studied the consensus algorithm where delay exists in data receptions. The delayed version of the consensus algorithm is introduced and by using properties of scrambling matrices, we have proved that delay does not affect convergence of the consensus algorithm as long as it is bounded. Furthermore, it is shown mathematically that the convergence rate of the distributed consensus algorithm is not reduced in the existence of delay for directed acyclic graphs such as master–slave topology, chain topology and hierarchical topology. This result does not depend on the choice of the averaging coefficients so long as the given basic assumption is satisfied. Despite focusing on the case where delay does not adversely effect the convergence rate, networks topologies where convergence rate is always degraded are to be explored as future work.

APPENDIX A: LEMMAS UTILIZED THROUGHOUT THE THESIS

Lemma A.1. *(Cauchy Interlacing Theorem, [40] Theorem 4.3.8 pp 185–186) Let H be an $n \times n$ Hermitian matrix. Let y be a vector of dimension n and let d be a real number. Let $\hat{H}$ be the $(n+1) \times (n+1)$ Hermitian matrix given as*

$$\hat{H} = \begin{bmatrix} H & y \\ y^\top & d \end{bmatrix}. \tag{A.1}$$

Then the eigenvalues of H and $\hat{H}$ satisfy

$$\lambda_1(\hat{H}) \geq \lambda_1(H) \geq \lambda_2(\hat{H}) \geq \lambda_2(H) \geq \cdots \geq \lambda_{n-1}(\hat{H}) \geq \lambda_{n-1}(H) \geq \lambda_n(\hat{H}). \tag{A.2}$$

Lemma A.2. *([40] Theorem 4.3.4(b) pp 182–183) Let A be an $n \times n$ Hermitian matrix. Let z be a vector of dimension n. Then*

$$\lambda_k(A) \geq \lambda_{k+1}(A \pm zz^*) \geq \lambda_{k+2}(A),\ k = 1, \ldots, n-2. \tag{A.3}$$

Lemma A.3. *([40] Theorem 8.1.18(b) p. 491) Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two matrices of the same dimension. If $|a_{ij}| \leq b_{ij}$ for all i and j, then $\lambda_1(A) \leq \lambda_1(abs(A)) \leq \lambda_1(B)$ where the matrix $abs(A)$ is of the same dimension as A and defined as $abs(A) = [|a_{ij}|]$.*

APPENDIX B: PROOF OF THEOREM 4.3

Let T_0 be the similarity transformation matrix given by

$$T_0 = diag\{\sqrt{3/2}, 1, \ldots, 1, \sqrt{3/2}\}$$

so that $\tilde{L}_{n,n} = T_0^{-1} L_{n,n} T_0$ is symmetric. Since $L_{n,n}$ is similar to a symmetric matrix, its spectrum is real. For the second part of the proof, we will consider two cases: (i) n is odd, (ii) n is even.

Case (i) n is odd: Let $x = [x_1 \cdots x_n]^\top$ be a vector and $flip(x)$ be defined as the vector which consists of the elements of x in reverse order, i.e., $flip(x) = [x_n \cdots x_1]^\top$. For n odd, note from the eigenvalue–eigenvector equation that $L_{n,n}$ has $(n-1)/2$ eigenvectors of the form

$$v_o = \begin{bmatrix} v_{r_1}^\top & 0 & -flip^\top(v_{r_1}) \end{bmatrix}^\top \tag{B.1}$$

and $(n+1)/2$ eigenvectors of the form

$$\bar{v}_o = \begin{bmatrix} v_{r_2}^\top & flip^\top(v_{r_2}) \end{bmatrix}^\top \tag{B.2}$$

for some $v_{r_1} \in \Re^{\lfloor \frac{n}{2} \rfloor}$ and $v_{r_2} \in \Re^{\lceil \frac{n}{2} \rceil}$. The eigenvalues corresponding to the eigenvectors of the form v_o can be obtained from the reduced order eigenvalue–eigenvector equation

$$L_{n,\lfloor \frac{n}{2} \rfloor} v_{r_1} = \lambda v_{r_1}. \tag{B.3}$$

Therefore we conclude that the spectrum of $L_{n,\lfloor \frac{n}{2} \rfloor}$ is a subset of the spectrum of $L_{n,n}$.

Similarly, the remaining eigenvalues corresponding to eigenvectors of the form $\bar{v}_o$ can be obtained from

$$\left(L_{n,\lceil \frac{n}{2} \rceil} + \frac{1}{3} q_{\lceil \frac{n}{2} \rceil} q_{\lfloor \frac{n}{2} \rfloor}^{\top}\right) v_{r_2} = \lambda v_{r_2}. \tag{B.4}$$

The spectrum of $L_{n,\lceil \frac{n}{2} \rceil} + \frac{1}{3} q_{\lceil \frac{n}{2} \rceil} q_{\lfloor \frac{n}{2} \rfloor}^{\top}$ consists of the remaining eigenvalues of $L_{n,n}$, which completes the proof for n odd.

Case (ii) n is even: For n even, $L_{n,n}$ has $n/2$ eigenvectors of the form

$$v_e = \begin{bmatrix} v_{r_3}^{\top} & -f^{\top}(v_{r_3}) \end{bmatrix}^{\top} \tag{B.5}$$

and $n/2$ eigenvectors of the form

$$\bar{v}_e = \begin{bmatrix} v_{r_4}^{\top} & f^{\top}(v_{r_4}) \end{bmatrix}^{\top}. \tag{B.6}$$

for some $v_{r_3}, v_{r_4} \in \Re^{\frac{n}{2}}$. The eigenvalues corresponding to the eigenvectors of the form v_e can be obtained from

$$\left(L_{n,\frac{n}{2}} - \frac{1}{3} q_{\frac{n}{2}} q_{\frac{n}{2}}^{\top}\right) v_{r_3} = \lambda v_{r_3}. \tag{B.7}$$

Hence, we can conclude that the spectrum of $L_{n,\frac{n}{2}} - \frac{1}{3} q_{\frac{n}{2}} q_{\frac{n}{2}}^{\top}$ is a subset of the spectrum of $L_{n,n}$.

Finally, the remaining eigenvalues corresponding to the other eigenvectors of the form $\bar{v}_e$ can be obtained from

$$\left(L_{n,\frac{n}{2}} + \frac{1}{3} q_{\frac{n}{2}} q_{\frac{n}{2}}^{\top}\right) v_{r_4} = \lambda v_{r_4}. \tag{B.8}$$

The spectrum of $L_{n,\frac{n}{2}} + \frac{1}{3} q_{\frac{n}{2}} q_{\frac{n}{2}}^{\top}$ consists of the remaining eigenvalues of $L_{n,n}$, which completes the proof for n even.

APPENDIX C: PROOF OF THEOREM 4.5

Since $L_{n,n}$ is a row–stochastic matrix, its largest eigenvalue is 1. From Theorem 4.3, its spectrum can be computed from those of two reduced–order matrices, which implies

$$\mu\left(L_{n,n}\right) = \begin{cases} \max\left\{\lambda_1\left(L_{n,\lfloor\frac{n}{2}\rfloor}\right), \lambda_2\left(L_{n,\lceil\frac{n}{2}\rceil} + \frac{1}{3}q_{\lceil\frac{n}{2}\rceil}q_{\lfloor\frac{n}{2}\rfloor}^{\top}\right)\right\}, \text{if } n \text{ is odd} \\ \max\left\{\lambda_1\left(L_{n,\frac{n}{2}} - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^{\top}\right), \lambda_2\left(L_{n,\frac{n}{2}} + \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^{\top}\right)\right\}, \text{if } n \text{ is even} \end{cases} \tag{C.1}$$

We will consider two cases: (i) n is odd, (ii) n is even.

Case (i) n is odd: For n odd, let $T_1 = diag\{\sqrt{3/2}, 1, \ldots, 1\}$ so that $L_{n,\lfloor\frac{n}{2}\rfloor}$ is similar to a symmetric tridiagonal matrix, $\tilde{L}_{n,\lfloor\frac{n}{2}\rfloor} = T_1^{-1}\left(L_{n,\lfloor\frac{n}{2}\rfloor}\right)T_1$. Likewise, let $T_2 = diag\{\sqrt{3/2}, 1, \ldots, 1, \sqrt{2}\}$ so that $L_{n,\lceil\frac{n}{2}\rceil} + \frac{1}{3}q_{\lceil\frac{n}{2}\rceil}q_{\lfloor\frac{n}{2}\rfloor}^{\top}$ is similar to a symmetric tridiagonal matrix, given by $T_2^{-1}\left(L_{n,\lceil\frac{n}{2}\rceil} + \frac{1}{3}q_{\lceil\frac{n}{2}\rceil}q_{\lfloor\frac{n}{2}\rfloor}^{\top}\right)T_2$. Moreover, the former symmetric tridiagonal matrix is a principal sub–matrix of the latter one, i.e.,

$$T_2^{-1}\left(L_{n,\lceil\frac{n}{2}\rceil} + \frac{1}{3}q_{\lceil\frac{n}{2}\rceil}q_{\lfloor\frac{n}{2}\rfloor}^{\top}\right)T_2 = \begin{bmatrix} T_1^{-1}L_{n,\lfloor\frac{n}{2}\rfloor}T_1 & y \\ y^{\top} & 1/3 \end{bmatrix}. \tag{C.2}$$

From Lemma A.1, we have $\lambda_1\left(T_1^{-1}L_{n,\lfloor\frac{n}{2}\rfloor}T_1\right) \geq \lambda_2\left(T_2^{-1}\left(L_{n,\lceil\frac{n}{2}\rceil} + \frac{1}{3}q_{\lceil\frac{n}{2}\rceil}q_{\lfloor\frac{n}{2}\rfloor}^{\top}\right)T_2\right)$. Since similar matrices have the same spectrum, $\mu\left(L_{n,n}\right) = \lambda_1\left(L_{n,\lfloor\frac{n}{2}\rfloor}\right)$ for n odd.

Case (ii) n is even: For n even, both $L_{n,\frac{n}{2}} - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^{\top}$ and $L_{n,\frac{n}{2}} + \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^{\top}$ are similar to tridiagonal symmetric matrices, $T_1^{-1}\left(L_{n,\frac{n}{2}} - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^{\top}\right)T_1$ and $T_1^{-1}\left(L_{n,\frac{n}{2}} + \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^{\top}\right)T_1$, respectively. From Lemma A.2, we have

$$\lambda_1\left(T_1^{-1}(L_{n,\frac{n}{2}} - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^{\top})T_1\right) \geq \lambda_2\left(T_1^{-1}(L_{n,\frac{n}{2}} + \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^{\top})T_1\right)$$

Since similar matrices have the same spectrum, we obtain $\mu\left(L_{n,n}\right) = \lambda_1\left(L_{n,\frac{n}{2}} - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^{\top}\right)$ for n even, which completes the proof.

APPENDIX D: PROOF OF THEOREM 4.7

We will consider two cases: (i) n is odd, (ii) n is even.

Case (i) n is odd: From Theorem 4.5, we have $\mu(L_{n,n}) = \lambda_1\left(L_{n,\lfloor\frac{n}{2}\rfloor}\right)$ and $\mu(L_{n+1,n+1}) = \lambda_1\left(L_{n,\lceil\frac{n}{2}\rceil} - \frac{1}{3}q_{\lceil\frac{n}{2}\rceil}q_{\lceil\frac{n}{2}\rceil}^\top\right)$. Let $H = T_1^{-1}\, L_{n,\lfloor\frac{n}{2}\rfloor}\, T_1$ where

$$T_1 = diag\{\sqrt{3/2}, 1, \ldots, 1\}.$$

Note that H is symmetric. Similarly, $L_{n,\lceil\frac{n}{2}\rceil} - \frac{1}{3}q_{\lceil\frac{n}{2}\rceil}q_{\lceil\frac{n}{2}\rceil}^\top$ is similar to a symmetric matrix $\hat{H}$ given by

$$\hat{H} = \begin{bmatrix} H & y \\ y^\top & d \end{bmatrix} \tag{D.1}$$

where $y = \frac{1}{3}q_{\lfloor\frac{n}{2}\rfloor}$ and $d = 0$. Note from Lemma A.1 that $\lambda_1(\hat{H}) \geq \lambda_1(H)$ holds. In order to prove that strict inequality holds, we proceed by contradiction. To this end, suppose that we have $\lambda_1(\hat{H}) = \lambda_1(H)$.

Let $z = \left[z_1, \ldots, z_{\lfloor\frac{n}{2}\rfloor}\right]^\top$ be the unit length eigenvector of H corresponding to its largest eigenvalue. From the Rayleigh—Ritz Theorem, we have $\lambda_1(H) = z^\top H z$. Let $\hat{z} = \begin{bmatrix} z^\top & 0 \end{bmatrix}^\top$. Note that $\hat{z}$ has unit length and we have

$$\hat{z}^\top \hat{H} \hat{z} = z^\top H z = \lambda_1(H) = \lambda_1(\hat{H}). \tag{D.2}$$

For the above to hold, $\hat{z}$ should be an eigenvector of $\hat{H}$ corresponding to its largest eigenvalue. From $\hat{H}\hat{z} = \lambda_1(\hat{H})\hat{z}$, we first obtain $z_{\lfloor\frac{n}{2}\rfloor} = 0$. By substituting $z_{\lfloor\frac{n}{2}\rfloor} = 0$ in $\hat{H}\hat{z} = \lambda_1(\hat{H})\hat{z}$, we determine that z is a vector of all zeros. Since z is supposed to be an eigenvector of H, this is a contradiction.

Case (ii) n **is even:** From Theorem 4.5, we have $\mu(L_{n,n}) = \lambda_1\left(L_{n,\frac{n}{2}} - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^\top\right)$, and $\mu(L_{n+1,n+1}) = \lambda_1\left(L_{n,\frac{n}{2}}\right)$. Let $H = T_1^{-1}\left(L_{n,\frac{n}{2}} - \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^\top\right)T_1$ and $\hat{H} = T_1^{-1}L_{n,\frac{n}{2}}T_1 = H + \frac{1}{3}q_{\frac{n}{2}}q_{\frac{n}{2}}^\top$ where $T_1 = diag\{\sqrt{3/2}, 1, \ldots, 1\}$. Note that both H and $\hat{H}$ are symmetric and non–negative. From Lemma A.3, we have $\lambda_1(\hat{H}) \geq \lambda_1(H)$. In order to prove that strict inequality holds, we proceed by contradiction. To this end, suppose that we have $\lambda_1(\hat{H}) = \lambda_1(H)$.

Let $z = \left[z_1, \ldots, z_{\frac{n}{2}}\right]^\top$ be the unit length eigenvector of H corresponding to its largest eigenvalue. Due to the structure of the matrices, $z_{\frac{n}{2}} = 0$ must hold in order to have $z^\top\hat{H}z = z^\top Hz$. However, this leads to z being a vector of all zeros, which is a contradiction.

REFERENCES

1. Akar, M. and R. Shorten, "Distributed Probabilistic Synchronization Algorithms for Communication Networks", *IEEE Transactions on Automatic Control*, Vol. 53, No. 1, pp. 389–393, 2008.

2. Akar, M. and R. Shorten, "On the Existence of Common Lyapunov Functions for Consensus Algorithms Based on Averaging", *Turkish Journal of Electrical Engineering and Computer Science*, Vol. 19, No. 3, pp. 483–493, 2011.

3. Angeli, D. and P.-A. Bliman, "Stability of Leaderless Discrete-Time Multi-Agent Systems", *Mathematics of Control, Signals and Systems*, Vol. 18, No. 4, pp. 293–322, 2006.

4. Bliman, P.-A. and G. Ferrari-Trecate, "Average Consensus Problems in Networks of Agents with Delayed Communications", *Automatica*, Vol. 44, No. 8, pp. 1985–1995, 2008.

5. Blondel, V. D., J. M. Hendrickx, A. Olshevsky and J. N. Tsitsiklis, "Convergence in Multiagent Coordination, Consensus, and Flocking", *Joint 44th IEEE Conference on Decision and Control and European Control Conference*, pp. 2996–3000, 2005.

6. Tanner, H., A. Jadbabaie and G. Pappas, "Stable Flocking of Mobile Agents, Part I: Fixed Topology", *Proceedings of the 42nd IEEE Conference on Decision and Control*, Vol. 2, pp. 2010–2015, 2003.

7. Tanner, H., A. Jadbabaie and G. Pappas, "Stable Flocking of Mobile Agents, Part II: Dynamic Topology", *Proceedings of the 42nd IEEE Conference on Decision and Control*, Vol. 2, pp. 2016–2021, 2003.

8. Olfati-Saber, R., "Flocking for Multi-Agent Dynamic Systems: Algorithms and Theory", *IEEE Transactions on Automatic Control*, Vol. 51, No. 3, pp. 401–420,

2006.

9. Gazi, V., "Stability of an Asynchronous Swarm With Time-Dependent Communication Links", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 38, No. 1, pp. 267–274, 2008.

10. Jadbabaie, A., J. Lin and A. S. Morse, "Coordination of Groups of Mobile Autonomous Agents Using Nearest Neighbor Rules", *IEEE Transactions on Automatic Control*, Vol. 48, No. 6, pp. 988–1001, 2003.

11. Lee, D. and M. W. Spong, "Agreement with Non–uniform Information Delays", *Proceedings of the 2006 American Control Conference*, pp. 756–761, 2006.

12. Nedić, A. and A. Ozdaglar, "Convergence Rate for Consensus With Delays", *Journal of Global Optimization*, Vol. 47, No. 3, pp. 437–456, 2010.

13. Olfati-Saber, R. and R. M. Murray, "Consensus Problems in Networks of Agents with Switching Topology and Time–delays", *IEEE Transactions on Automatic Control*, Vol. 49, No. 9, pp. 1520–1533, 2004.

14. Ren, W. and R. W. Beard, "Consensus Seeking in Multi-Agent Systems Under Dynamically Changing Interaction Topologies", *IEEE Transactions on Automatic Control*, Vol. 40, No. 5, pp. 655–661, 2005.

15. Ren, W., R. Beard and E. Atkins, "A Survey of Consensus Problems in Multi-Agent Coordination", *Proceedings of the 2005 American Control Conference*, Vol. 3, pp. 1859–1864, 2005.

16. Lin, J., A. Morse and B. Anderson, "The Multi-Agent Rendezvous Problem", *Proceedings of the 42nd IEEE Conference on Decision and Control*, Vol. 2, pp. 1508–1513 Vol.2, 2003.

17. Ren, W., R. W. Beard and T. W. McLain, "Coordination Variables and Consen-

sus Building in Multiple Vehicle Systems", *Cooperative Control*, Lecture Notes in Control and Information Science, pp. 171–188, Springer Berlin Heidelberg, 2005.

18. Tsitsiklis, J. N., *Problems in Decentralized Decision Making and Computation*, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1984.

19. Fax, J. A. and R. M. Murray, "Information Flow and Cooperative Control of Vehicle Formations", *IEEE Transactions on Automatic Control*, Vol. 49, No. 9, pp. 1465–1476, 2004.

20. Lawton, J., R. Beard and B. Young, "A Decentralized Approach to Formation Maneuvers", *IEEE Transactions on Robotics and Automation*, Vol. 19, No. 6, pp. 933–941, 2003.

21. Lin, Z., B. Francis and M. Maggiore, "Necessary and Sufficient Graphical Conditions for Formation Control of Unicycles", *IEEE Transactions on Automatic Control*, Vol. 50, No. 1, pp. 121–127, 2005.

22. Marshall, J., M. Brouke and B. Francis, "A Pursuit Strategy for Wheeled-Vehicle Formations", *Proceedings of the 42nd IEEE Conference on Decision and Control*, Vol. 3, pp. 2555–2560, 2003.

23. Wang, W. and J.-J. E. Slotine, "Contraction Analysis of Time-Delayed Communications and Group Cooperation", *IEEE Transactions on Automatic Control*, Vol. 51, No. 4, pp. 712–717, 2006.

24. Wang, L. and F. Xiao, "A New Approach to Consensus Problems for Discrete-Time Multiagent Systems With Time-Delays", *Proceedings of the 2006 American Control Conference*, pp. 2118–2123, 2006.

25. Xiao, F. and L. Wang, "Consensus Protocols for Discrete-Time Multi-Agent Systems with Time-Varying Delays", *Automatica*, Vol. 44, No. 10, pp. 2577–2582, 2008.

26. Boyd, S., P. Diaconis and L. Xiao, "Fastest Mixing Markov Chain on A Graph", *Siam Review*, Vol. 46, pp. 667–689, 2003.

27. Boyd, S., P. Diaconis, J. Sun and L. Xiao, "Fastest Mixing Markov Chain on a Path", *American Mathematical Monthly*, Vol. 113, pp. 70–74, 2006.

28. Boyd, S., P. Diaconis, P. A. Parrilo and L. Xiao, "Fastest Mixing Markov Chain on Graphs with Symmetries", *SIAM Journal on Optimization*, Vol. 20, pp. 792–819, 2009.

29. Pereira, R. and M. A. Vali, "Inequalities for the Spectra of Symmetric Doubly Stochastic Matrices", *Linear Algebra and its Applications*, Vol. 419, No. 2–3, pp. 643–647, 2006.

30. Jafarizadeh, S. and A. Jamalipour, "Fastest Mixing Markov Chain on Symmetric K–partite Sensor Networks", *Proceedings of the 7th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp. 1883–1888, 2011.

31. da Fonseca, C., "Remarks and Open Problems on the Minimum Permanent of Acyclic Doubly Stochastic Matrices", *Linear and Multilinear Algebra*, Vol. 61, No. 3, pp. 281–293, 2013.

32. Allison, M. and B. L. Shader, "Fastest Mixing Markov Chain Problem for the Union of Two Cliques", *Linear and Multilinear Algebra*, Vol. 59, No. 7, pp. 801–823, 2011.

33. Grant, M. and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, version 2.0 beta", http://cvxr.com/cvx, Sep. 2013.

34. Grant, M. and S. Boyd, "Graph implementations for nonsmooth convex programs", *Recent Advances in Learning and Control*, pp. 95–110, Springer-Verlag Limited, 2008.

35. Vicsek, T., A. Czirók, E. Ben-Jacob, I. Cohen and O. Shochet, "Novel Type of

Phase Transition in a System of Self-Driven Particles", *Physical Review Letters*, Vol. 75, pp. 1226–1229, 1995.

36. Hartfiel, D. J., *Markov Set-Chains*, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1998.

37. Tsitsiklis, J. N., D. P. Bertsekas and M. Athans, "Distributed Asynchronous Deterministic and Stochastic Gradient Optimization Algorithms", *IEEE Transactions on Automatic Control*, Vol. 31, No. 9, pp. 803–812, 1986.

38. Li, S. and H. Wang, "Multi-agent coordination using nearest neighbor rules: revisiting the Vicsek model", *CoRR*, Vol. cs.MA/0407021, 2004.

39. Moreau, L., "Stability of Multiagent Systems with Time-Dependent Communication Links", *IEEE Transactions on Automatic Control*, Vol. 50, No. 2, pp. 169–182, 2005.

40. Horn, R. A. and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, NY, USA, 1990.

41. Brouwer, A. E. and W. H. Haemers, *Spectra of Graphs*, Springer, New York, NY, USA, 2012.

42. Xiao, L. and S. Boyd, "Fast Linear Iterations for Distributed Averaging", *Systems and Control Letters*, Vol. 53, pp. 65–78, 2003.

43. Boyd, S., P. Diaconis, P. Parrilo and L. Xiao, "Symmetry Analysis of Reversible Markov Chains", *Internet Mathematics*, Vol. 2, No. 1, pp. 31–71, 2011.