Departament de Teoria del Senyal i Comunicacions
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Distributed Consensus Algorithms for Wireless Sensor Networks

Convergence Analysis and Optimization

Ph.D. Thesis Dissertation

by

Silvana Silva Pereira

Submitted to the Universitat Politècnica de Catalunya - Barcelona Tech
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Supervised by Alba Pagès-Zamora

Barcelona, December 2011

A mi madre,

Abstract

Wireless sensor networks are developed to monitor areas of interest with the purpose of estimating physical parameters or/and detecting emergency events in a variety of military and civil applications. A wireless sensor network can be seen as a distributed computer, where spatially deployed sensor nodes are in charge of gathering measurements from the environment to compute a given function. The research areas for wireless sensor networks extend from the design of small, reliable hardware to low-complexity algorithms and energy saving communication protocols.

Distributed consensus algorithms are low-complexity iterative schemes that have received increased attention in different fields due to a wide range of applications, where neighboring nodes communicate locally to compute the average of an initial set of measurements. Energy is a scarce resource in wireless sensor networks and therefore, the convergence of consensus algorithms, characterized by the total number of iterations until reaching a steady-state value, is an important topic of study.

This PhD thesis addresses the problem of convergence and optimization of distributed consensus algorithms for the estimation of parameters in wireless sensor networks. The impact of quantization noise in the convergence is studied in networks with fixed topologies and symmetric communication links. In particular, a new scheme including quantization is proposed, whose mean square error with respect to the average consensus converges.

The limit of the mean square error admits a closed-form expression and an upper bound for this limit depending on general network parameters is also derived.

The convergence of consensus algorithms in networks with random topology is studied focusing particularly on convergence in expectation, mean square convergence and almost sure convergence. Closed-form expressions useful to minimize the convergence time of the algorithm are derived from the analysis.

Regarding random networks with asymmetric links, closed-form expressions are provided for the mean square error of the state assuming equally probable uniform link weights, and mean square convergence to the statistical mean of the initial measurements is shown. Moreover, an upper bound for the mean square error is derived for the case of different probabilities of connection for the links, and a practical scheme with randomized transmission power exhibiting an improved performance in terms of energy consumption with respect to a fixed network with the same consumption on average is proposed. The mean square error expressions derived provide a means to characterize the deviation of the state vector with respect to the initial average when the instantaneous links are asymmetric.

A useful criterion to minimize the convergence time in random networks with spatially correlated links is considered, establishing a sufficient condition for almost sure convergence to the consensus space. This criterion, valid also for topologies with spatially independent links, is based on the spectral radius of a positive semidefinite matrix for which we derive closed-form expressions assuming uniform link weights. The minimization of this spectral radius is a convex optimization problem and therefore, the optimum link weights minimizing the convergence time can be computed efficiently. The expressions derived are general and apply not only to random networks with instantaneous directed topologies but also to random networks with instantaneous undirected topologies. Furthermore, the general expressions can be particularized to obtain known protocols found in literature, showing that they can be seen as particular cases of the expressions derived in this thesis.

Resumen

Las redes de sensores inalámbricos se utilizan para monitorizar zonas de interés con el propósito final de estimar parámetros físicos y/o detectar situaciones de emergencia en gran variedad de aplicaciones militares y civiles. Una red de sensores inalámbricos puede ser considerada como un método de computación distribuido, donde nodos provistos de sensores toman medidas del entorno para calcular una función que depende de éstas. Las áreas de investigación comprenden desde el diseño de dispositivos hardware pequeños y fiables hasta algoritmos de baja complejidad o protocolos de comunicación de bajo consumo energético.

Los algoritmos de consenso distribuidos son esquemas iterativos de baja complejidad que han suscitado mucha atención en diferentes campos debido a su gran espectro de aplicaciones, en los que nodos vecinos se comunican para calcular el promedio de un conjunto de medidas iniciales de la red. Dado que la energía es un recurso escaso en redes de sensores inalámbricos, la convergencia de dichos algoritmos de consenso, caracterizada por el número total de iteraciones hasta alcanzar un valor estacionario, es un importante tema de estudio.

Esta tesis doctoral aborda problemas de convergencia y optimización de algoritmos de consenso distribuidos para la estimación de parámetros en redes de sensores inalámbricos. El impacto del ruido de cuantización en la convergencia se estudia en redes con topología fija y enlaces de comunicación simétricos. En particular, se propone un nuevo esquema que

incluye el proceso de cuantización y se demuestra que el error cuadrático medio respecto del promedio inicial converge. Igualmente, se obtiene una expresión cerrada del límite del error cuadrático medio, y una cota superior para este límite que depende únicamente de parámetros generales de la red.

La convergencia de los algoritmos de consenso en redes con topología aleatoria se estudia prestando especial atención a la convergencia en valor esperado, la convergencia en media cuadrática y la convergencia casi segura, y a partir del análisis se derivan expresiones cerradas útiles para minimizar el tiempo de convergencia.

Para redes aleatorias con enlaces asimétricos, se obtienen expresiones cerradas del error cuadrático medio del estado suponiendo enlaces con probabilidad idéntica y con pesos uniformes, y se demuestra la convergencia en media cuadrática al promedio estadístico de las medidas iniciales. Se deduce una cota superior para el error cuadrático medio para el caso de enlaces con probabilidades de conexión diferentes y se propone, además, un esquema práctico con potencias de transmisión aleatorias, que mejora el rendimiento en términos de consumo de energía con respecto a una red fija. Las expresiones para el error cuadrático medio proporcionan una forma de caracterizar la desviación del vector de estado con respecto del promedio inicial cuando los enlaces instantáneos son asimétricos.

Con el fin de minimizar el tiempo de convergencia en redes aleatorias con enlaces correlados espacialmente, se considera un criterio que establece una condición suficiente que garantiza la convergencia casi segura al espacio de consenso. Este criterio, que también es válido para topologías con enlaces espacialmente independientes, utiliza el radio espectral de una matriz semidefinida positiva para la cual se obtienen expresiones cerradas suponiendo enlaces con pesos uniformes. La minimización de dicho radio espectral es un problema de optimización convexa y, por lo tanto, el valor de los pesos óptimos puede calcularse de forma eficiente. Las expresiones obtenidas son generales y aplican no sólo para redes aleatorias con topologías dirigidas, sino también para redes aleatorias con topologías no dirigidas. Además, las expresiones generales pueden ser particularizadas para obtener protocolos conocidos en la literatura, demostrando que éstos últimos pueden ser considerados como casos particulares de las expresiones proporcionadas en esta tesis.

Agradecimientos

Quiero expresar mi más sincera gratitud a mi directora de tesis, Alba Pagés-Zamora, cuyo apoyo y dedicación han sido claves para hacer posible la realización de este trabajo. Saber que siempre ha estado disponible y que he podido contar con ella para lo que fuera me ha dado la fuerza y la confianza necesarias para llegar a la meta. Estos años de trabajo y las vivencias compartidas fuera del ámbito laboral me han hecho aprender mucho y le estoy agradecida de todo corazón. Quiero aprovechar la oportunidad para agradecer también a todas las personas que me han alentado a lo largo del camino.

Jeg vil gjerne takke Tobias Dahl for hans støtte og oppmuntring for å få meg til å fortsette. Jeg husker fortsatt delen av en samtale foran Frognerparken der han sa “du må ta doktorgraden!”. Jeg er også takknemlig for at professor David Gesbert satte meg i kontakt med UPC i Barcelona.

I would also like to thank professor Sergio Barbarossa for the collaboration during my stay in Rome. I really enjoyed the time I spent at the INFOCOM Department and, although short, I got inspired for my further work.

Agradezco en especial a mis amigos, los que están cerca y los que están lejos, con quienes sé que siempre puedo contar y que conforman una hinchada maravillosa. A Vero, mi Tintan, hermana del alma, capaz de arrancarme una carcajada en cualquier circunstancia. A mi Pato querida, por recibirme en su casa y por darme un curso intensivo de cariño y convivencia.

A las amistades que se han forjado en la UPC, y que seguirán siendo importantes cuando esto se haya terminado, en especial a Leo y a Maribel. A Luis García por leerse el resumen y hacer muy buenos comentarios. A Luis Segura porque también se lo leyó y con quien, además de tertulias, hemos sabido compartir tribuna para disfrutar de la celeste.

Por último quiero agradecer a mi mamá por haberme dado todo y a mis hermanos por darme sobrinos maravillosos. Y a Ossi, por creer en este proyecto y quererme tanto.

Barcelona, diciembre de 2011.

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia a través del programa de Formación de Profesorado Universitario (F.P.U.) mediante la beca AP2005-4895 y el proyecto CONSOLIDER CSD2008-00010 COMONSENS.

Contents

Notation and Abbreviations

Vectors and matrices

$x, \mathbf{x}, \mathbf{X}$	a scalar, a vector and a matrix
$\mathbf{x}_i$	the entry of the i^{th} row of vector $\mathbf{x}$
$\mathbf{X}_{ij}, x_{ij}$	the entry of the i^{th} row and j^{th} column of $\mathbf{X}$
$\mathbf{x}^T, \mathbf{X}^T$	the transpose of a vector $\mathbf{x}$, the transpose of a matrix $\mathbf{X}$
$\mathbf{X}^{-1}$	the inverse of a square matrix $\mathbf{X}$
$\text{tr}(\mathbf{X})$	the trace of a matrix $\mathbf{X}$
$rank(\mathbf{X})$	the rank of a matrix $\mathbf{X}$, or the dimension of its column space
$diag(\mathbf{x})$	a diagonal matrix whose entries are the elements of vector $\mathbf{x}$
$span\{\mathbf{x}, \mathbf{y}\}$	the subspace spanned by vectors $\mathbf{x}$ and $\mathbf{y}$
$\\|\mathbf{x}\\|_2$	the 2-norm of a vector $\mathbf{x}$
$\mathbf{X} \succ 0$	$\mathbf{X}$ is positive definite
$\mathbf{X} \succeq 0$	$\mathbf{X}$ is positive semidefinite
$\mathbf{X} \succeq \mathbf{Y}$	$\mathbf{X} - \mathbf{Y}$ is positive semidefinite
$\mathbf{X} \odot \mathbf{Y}$	the Schur product between $m \times n$ matrices $\mathbf{X}$ and $\mathbf{Y}$
	i.e., the element-wise multiplication of their elements.
$\mathbf{X} \otimes \mathbf{Y}$	the Kronecker product between matrices $\mathbf{X}$ and $\mathbf{Y}$

Sets

$\mathcal{X}$	a finite nonempty set of elements
$x \in \mathcal{X}$	x belongs to the set $\mathcal{X}$
$\mathcal{X} \subseteq \mathcal{Y}$	$\mathcal{X}$ is a subset of $\mathcal{Y}$
$\lvert\mathcal{X}\rvert$	cardinality of the set $\mathcal{X}$
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^+$	the set of positive real numbers
$\mathbb{S}$	the set of symmetric matrices
$\mathbb{S}^+$	the set of positive semidefinite matrices

Operators and relations

$\triangleq$	defined as
$\approx$	approximately
$\equiv$	is equivalent
$\lim$	limit
$\max$	maximum
$\min$	minimum
$\lvert x\rvert$	the absolute value of x
$\mathbb{E}[x]$	the expected value of x
$\bar{x}$	the expected value or mean of x
$\Pr\{x\}$	the probability of x
$\mathrm{Re}[x]$	the real part of x
δ_{ij}	Kronecker delta function

Common notations

$x_i(0)$	initial state of node i
$x_i(k)$	state of node i at time k
$\mathbf{x}(0)$	initial state vector
$\mathbf{x}(k)$	state vector at time k

$\mathbf{x}_{ave}$	$\frac{1}{N}\mathbf{1}^T\mathbf{x}(0)\mathbf{1}$, average consensus vector
$\bar{\mathbf{x}}_{ave}$	$\frac{1}{N}\mathbf{1}^T\mathbb{E}[\mathbf{x}(0)]\mathbf{1}$, mean average consensus vector
$\mathbf{0}$	the zero vector
$\mathbf{1}$	the all-ones vector
$\mathbf{I}$	the identity matrix
$\mathbf{e}_i$	the i^{th} vector of the matrix $\mathbf{I}$
$\mathbf{J}$	the all-ones matrix
$\mathbf{J}_N$	normalized all-ones matrix
$\mathbf{P}$	the connection probability matrix
$\rho(\mathbf{X})$	the spectral radius of $\mathbf{X}$
$\lambda_i(\mathbf{X})$	i^{th} eigenvalue of $\mathbf{X}$
$\lambda_1(\mathbf{X})$	the largest eigenvalue in magnitude of $\mathbf{X}$
$\lambda_N(\mathbf{X})$	the smallest eigenvalue in magnitude of $\mathbf{X}$
x_0	statistical mean of the measurements
σ_0^2	variance of the measurements
σ_q^2	variance of the quantization noise
ϵ, ϵ^*	link weight and optimum link weight

Graph theory terms

$\mathcal{G}$	a time-invariant graph
$\mathcal{G}(k), \bar{\mathcal{G}}$	a time-varying graph, the expected value of $\mathcal{G}(k)$
$\mathcal{V}$	the time-invariant set of nodes of $\mathcal{G}$
$\mathcal{E}, \mathcal{E}(k)$	the time-invariant/time-varying set of links of $\mathcal{G}$
e_{ij}	edge (link) from node j to node i
$\mathcal{N}_i, \mathcal{N}_i(k)$	the time-invariant/time-varying set of neighbors of node i
$\mathbf{A}, \mathbf{D}, \mathbf{L}$	the adjacency matrix, the degree matrix, the Laplacian matrix
$\mathbf{A}(k), \bar{\mathbf{A}}$	the adjacency matrix at time k, the expected adjacency matrix
$\mathbf{D}(k), \bar{\mathbf{D}}$	the degree matrix at time k, the expected degree matrix
$\mathbf{L}(k), \bar{\mathbf{L}}$	the Laplacian matrix at time k, the expected Laplacian matrix

Acronyms

BLUE	best linear unbiased estimator
FC	fusion center
ML	maximum likelihood
MAC	medium access control
MSE	mean square error
WSN	wireless sensor network
i.i.d.	independent identically distributed
r.v.	random variable

1

Introduction

Wireless sensor networks (WSNs) are developed with the purpose of estimating physical parameters or detecting emergency events in a variety of military and civil applications like battlefield surveillance, target tracking, environmental monitoring for detection of fire hazards, gas leakages or landslides, home automation and health care applications [Cho03, Aky02]. A WSN is composed of multiple units called sensor nodes which are deployed in the area of observation. Depending on the application and on the coverage area, the deployment can consist of a reduced number of sensor nodes or on the contrary, it can be a large-scale deployment composed of hundreds of units. In large-scale WSN applications, low-cost sensor nodes are preferred rather than expensive ones, although they are expected to satisfy minimum requirements so the quality of the measurements is not jeopardized.

A standard sensor node is usually composed of a transducer, in charge of sensing the physical parameters, a radio transceiver for wireless communications, a low complexity processing unit and a power supply, normally in form of a battery. The sensors gather measurements from the environment and eventually make simple processing of the sensed data. The data can be transferred to a central node in a centralized network, or it can be locally processed instead in a decentralized network. A microwave link or a satellite link can be finally used to extract the information from the WSN to take further actions.

Relevant topics of study for WSN applications include the design of low-cost simple devices able to perform simple tasks, possible energy harvesting from the environment, design of energy efficient communication protocols, self-organization of the network, management of node failures and information fusion, among others. A fundamental requirement to guarantee a proper information flow is that the network is connected, that is, a path connecting each node with the network is needed. In a WSN, the existence of a connection between a pair of nodes depends on the transmission power applied and on their geographical location, which can be decided according to optimization methods or it can be completely random; for instance, in natural areas with difficult or restricted access, the nodes might be spread out from a plane. Higher levels of transmission power results typically in more connections among the nodes. Therefore, the transmission power and the location of the nodes determine the connectivity or topology of the network.

If the nodes in a WSN are not plugged into any form of power supply, energy becomes a scarce resource and must be properly administrated. In situations where reaching the devices is practically impossible or when the cost of mobilizing personnel is high, used-up batteries may not be replaced. Although a good option is to provide the sensor nodes with self-rechargeable batteries using for instance solar energy, a low-power consumption is essential to guarantee a longer lifetime for the entire network. Energy saving communication protocols, periodic sleeping intervals and low-complexity programming are some approaches to reduce the overall energy consumption of a WSN.

In summary, the nodes of a -possibly large-scale- WSN are therefore required to be simple and preferably have a reduced size, be inexpensive yet reliable; these devices should be able to perform simple computations and to implement low-complexity protocols.

In typical centralized deployments, the nodes convey their measurements to a more complex and intelligent unit denoted fusion center (FC), in charge of collecting the data of the network and making the final computations. Centralized networks require a proper organization of the nodes and the implementation of medium access control (MAC) as well as routing protocols to forward the data to the FC. In event-driven applications or for instance when an emergency situation arises, the information flow to the FC can become particularly high and create congestion. Moreover, a re-organization of the MAC and routing protocols is required every time a node falls down, turns to sleeping mode or is added to the network. The hardware requirements for wireless communications may also lead to an increase in the cost of the devices and thus, a higher overall cost of the network, specially when the number of nodes becomes large. For these reasons, a centralized WSN can be highly inefficient and expensive in terms of both energy consumption, scalability and response to event-driven applications.

An alternative is a decentralized network, an architecture where all the nodes have the same capabilities and are able to perform the same tasks. The principle of decentralized systems is that the nodes organize themselves interacting locally and carry out the computations without the necessity of conveying the information to a FC. Each node communicates with neighboring nodes -often located within a small range- to exchange their information and make decisions. A decentralized network is expected to provide reliable results that approach a globally optimal solution and in some cases, the global information is expected to be available at each node. A decentralized network can be also organized in clusters, where a node within a cluster and denoted cluster-head functions as a local FC, establishing a connection with other cluster-heads and forming an upper layer of the network connected for instance to a computer with the final application. This is an example of a decentralized network with a hierarchical structure [Gir05, Gir06]. In this thesis however, decentralized WSNs refers to architectures without neither a central node nor clusters-heads.

While in a centralized WSN the FC is in charge of the computations, in decentralized architectures the computations are carried out in a distributed manner, giving rise to distributed algorithms. A WSN can be therefore seen as a distributed computer aimed at computing a function of the data collected by the sensor nodes, and can be designed using tools of parallel and distributed optimization [Ber97,Rab04]. A distributed algorithm may require global information, that is, information which can not be computed through local interactions, or conversely it may only use data gathered by the nodes themselves along with data received from one-hop neighbors. Whenever required, global parameter values can be made available at each node through broadcasting or routing.

Some of the most critical aspects to take into account while designing distributed algorithms for WSNs are the scarcity of resources, the reliability of the final decision as well as the robustness to node failures or sleeping mode periods [Lyn96, Gir06]. Smart hardware design, energy saving algorithms and communication protocols are therefore necessary to build robust and reconfigurable WSNs optimally exploiting the available resources.

1.1 Distributed Consensus Algorithms

Distributed consensus algorithms are low-complexity iterative algorithms where neighboring nodes communicate with each other to reach an agreement regarding a function of the measurements, without the necessity of forwarding any information to a FC. In particular, the average consensus algorithm computes the average of an initial set of measurements [OS04]. In a digital implementation, each node in the network programs a discrete dynamical system whose state is initialized with the value of one or several measurements and updated iteratively using a linear combination of its previous state value and the information received from its neighbors.

Also known as the alignment or the agreement problem [Bor82], consensus was early studied by Tsitsiklis [Tsi84, Tsi86] and has received increased attention in different fields due to its wide range of applications such as load balancing in parallel computing [Cyb89, Ber97, Die97], coordination of autonomous agents [Jad03, Fax04, Lin05, Mor05, Ren05a, Ols06], distributed control [Xia03,OS04,Wu05], data fusion problems [Zha03,Sch04,Spa05, Xia05, Moa06, Sch08], or flocking in dynamical systems [Blo05, OS06]. Consensus could be also seen as a form of self-synchronization of coupled oscillators [Vic95, Mir90, Bel04, Hon05, Bar05a, Bar05b, Por08].

Linear consensus algorithms and nonlinear consensus algorithms are studied in continuous or in discrete form, and can converge either on the state, meaning that the state of the dynamical system reaches a consensus, or on the state derivative, meaning that the derivative of the state reaches a steady-state value. Algorithms converging on the state are robust to changes in the connectivity of the network and have a bounded state value [OS04], while algorithms converging on the state derivative are resilient to propagation delays or to coupling noise. Detailed surveys for continuous-time and for discrete-time implementations of consensus algorithms converging on the state can be found in [Ren05b, OS07]. Consensus algorithms converging on the state derivative are thoroughly studied in [Bar05a, Bar05b, Pes06, Bar06, Bar07b]. [Bar07a] derives conditions on the coupling mechanism, shows globally asymptotically stability of the synchronized state and studies the impact of changes in the topology. Further, the impact of coupling noise is studied in [Clo07] while the effect of propagation delay on both the synchronization capability of the system and on the final estimate is studied in [Scu06].

Consensus algorithms can be synchronous, meaning that the nodes update their state at the same time instant, or on the contrary they can be asynchronous, meaning that the nodes update their state at different time instants. An example of an asynchronous consensus algorithm is the random gossip algorithm, where we can distinguish between three different forms of implementation: the pair-wise gossip algorithm [Boy05, Boy06, Gir06], the geographic gossip algorithm [Dim08] and the broadcast gossip algorithm [Ays09]. The gossip algorithm can be resumed as follows. A node wakes up randomly and either establishes a bidirectional communication link with a randomly chosen node to exchange the state values in pair-wise and geographic gossiping, or it broadcasts its state to the neighboring nodes within connectivity range in broadcast gossiping. The two first models converge to the average of the initial values due to the symmetry of the communication links, whereas the last one converges to a value different from the average consensus due to the non-symmetric nature of the links. The difference between gossip and standard consensus is that in the former, only one link/node is active at each iteration, whereas in the latter, several nodes are transmitting at the same time. Gossip algorithms can be therefore seen as asynchronous versions of the consensus algorithm with spatially correlated links, and important contributions for gossip are useful for the convergence analysis of consensus algorithms.

1.2 Convergence of Consensus Algorithms

Due to its iterative nature, the convergence of the consensus algorithm is determined by the total number of iterations until reaching a steady-state value. A smaller number of iterations until reaching a consensus is therefore interpreted as a faster convergence of the algorithm. In particular, a reduction in the total number of iterations until convergence in a WSN can lead to a reduction in the total amount of energy consumed by the network, a desirable result since energy is a scarce resource.

The convergence analysis of consensus algorithms has a vast literature, specially for networks with time-invariant topologies and communication links exhibiting symmetry. Regarding more general network characteristics, the number of contributions is more limited. Early results on consensus focus on fixed topologies [Xia03, Sch04], that is, networks where the nodes and the communication links are assumed constant throughout time. Regarding time-varying networks, [OS04] introduces the concept of switching topology which refers to a deterministically time-varying model where at each time instant, the network adopts a topology from a finite known set. In that contribution, convergence conditions for networks with such topologies are derived. Furthermore, the effect of time delays is studied in [OS04, Fan05, Xia06a, Lin09] whereas the effect of additive noise is addressed in e.g. [Xia05, Xia06b, Hat05, Kar09, Ays10].

When the links are symmetric or bidirectional, the topology of the network is denoted undirected. On the other hand, when the links are asymmetric or directional, the topology of the network is denoted directed. In the presence of random failures caused by for instance changes in the environment, mobility of the nodes, asynchronous sleeping periods or randomized communication protocols, the topology of a WSN varies randomly with time and the convergence can be characterized in probabilistic terms. Important contributions for undirected random networks with independent links include [Hat04, Hat05, Ols06, Pat07, Kar07] whereas for directed random networks with independent links important contributions include [Wu06, Por07, TS08, Por08, Zho09, Pre10]. The works in [Aba10, Jak10] provide results assuming correlated random topologies, whereas [Boy06, Pic07, Fag08, Ays09, Ays10] focus on random gossip algorithms.

This PhD thesis focuses on the convergence analysis and optimization of distributed consensus algorithms for wireless sensor networks. The impact of quantization noise in the performance of the algorithm in a network with fixed topology is also studied, where a new scheme including quantization is proposed. The convergence in networks with random topology is studied, focusing particularly on convergence in expectation, mean square convergence and almost sure convergence.

1.3 Thesis Outline and Contributions

The work described in this PhD thesis is the result of the research on discrete-time consensus algorithms converging on the state for WSNs and, in particular, on the average consensus algorithm for the estimation of physical parameters like temperature, humidity, pressure, etc.

The organization of the thesis and the main contributions from each chapter are described below, along with the publications which resulted from the research. The use of graph theory terms and concepts is widely used throughout this thesis, and for that reason, the next chapter is devoted to introduce graph theory concepts. The third chapter is devoted to consensus algorithms while the last three chapters present the main results of this thesis, concerning the effect of quantization noise in fixed topologies and probabilistic convergence in networks with random topologies.

Chapter 2

This chapter presents a review of fundamental concepts of algebraic graph theory and the notation used in subsequent chapters. Notions of connectivity for undirected and for directed graphs, definitions of subgraphs and trees and an overview of common graph topologies are presented. The matrices associated with a graph, with special focus on the Laplacian matrix along with its spectral properties are described in detail. Definitions for deterministic time-varying and for random time-varying graphs are finally presented.

Chapter 3

This chapter presents the state of the art on discrete-time consensus algorithms converging on the state for applications aimed at estimating one or several parameters. We start presenting the time-invariant consensus model and review the conditions for convergence to a common value in directed and undirected topologies, as well as the conditions to reach the average of the initial set of measurements. Then, we introduce consensus in time-varying topologies, differentiating between the deterministic case and the random case, and define different forms for random convergence of the state vector. We review some commonly used designs for the weight matrix of the consensus algorithm, with special focus on the uniform weights model and its convergence conditions. Finally, we discuss approaches to reduce the convergence time of the consensus algorithm in networks with time-invariant topology.

Chapter 4

The focus of this chapter is on consensus algorithms with quantized information exchange. A detailed review of existing contributions is provided as well as the quantization noise model used in the chapter. A new approach which results from a modification of the well-known discrete-time consensus model by Olfati-Saber and Murray [OS04] is presented, and its performance is evaluated by analyzing the mean square error of the state with respect to the average of the initial values, as well as its asymptotic behavior. Conversely to existing models that include quantization, the mean square error of the state for the proposed model converges and its limit admits a closed-form expression. An upper bound for the limit of the mean square error which depends on general network parameters is derived as well.

The work of this chapter has led to the publication of one article in an international conference.

[Sil08] S. Silva Pereira and A. Pagès-Zamora, "Distributed consensus in wireless sensor networks with quantized information exchange", *Proceedings of the 9th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC'08)*, pp. 241–245, Recife, Brasil, July 2008.

Chapter 5

This chapter presents the study of mean square convergence of consensus algorithms in networks with random directed topologies, where the mean square error with respect to the statistical mean of the initial values is analyzed. For the case of random links with equal probability, closed-form expressions for the mean square error of the state and for the asymptotic mean square error are derived, as well as the dynamic range and the value of the link weights minimizing the convergence time. For the case of random links with different probabilities, an upper bound for the mean square error of the state is derived and the asymptotic mean square error is studied. Additionally, an approach to find the optimum link weights minimizing the convergence time of the upper bound is provided. Finally, a practical scheme of randomized transmission power intended to reduce the overall power consumption of the network is proposed, where the results of the chapter are used to minimize the convergence time of the algorithm in the mean square sense.

The technical contributions of this chapter have been published in the proceedings of two international conferences and in one journal paper.

[Sil09a] S. Silva Pereira and A. Pagès-Zamora, "Fast mean square convergence of consensus algorithms in WSNs with random topologies", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'09)*, pp. 2213–2216, Taipei, Taiwan, April 2009.

[Sil09b] S. Silva Pereira and A. Pagès-Zamora, "Randomized transmission power for accelerated consensus in asymmetric WSNs", *Proceedings of the 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP'09)*, pp. 348–351, Aruba, Dutch Antilles, December 2009.

[Sil10a] S. Silva Pereira and A. Pagès-Zamora, "Mean square convergence of consensus algorithms in random WSNs", *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2866–2874, May 2010.

A contribution partially using the results of this chapter and not included in this PhD thesis, is published in the proceedings of an international conference.

[Sil10b] S. Silva Pereira, S. Barbarossa, and A. Pagès-Zamora, "Consensus for distributed EM-based clustering in WSNs", *Proceedings of the 6th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM'10)*, pp. 45–48, Ma'ale Hahamisha, Israel, October 2010.

Chapter 6

This chapter studies the convergence of consensus algorithms in random networks assuming spatially correlated links, where an optimization criterion that establishes a sufficient condition for almost sure convergence is considered. The convergence is related to the spectral radius of a positive semidefinite matrix for which we derive closed-form expressions for both directed and undirected topologies. The minimization of this spectral radius can be obtained as the solution of a convex optimization problem and the general formulations derived subsume known protocols found in literature. Additional closed-form expressions for the dynamic range and the optimum link weights for particular cases of links with equal probability of connection are also provided. The analytical results are further validated with computer simulations of a general case with different probabilities of connection for the links and different correlations among pairs of links. Simulations of a small-world network and simulations of a randomized transmission power network are also provided.

The technical contributions of this chapter have been published in the proceedings of two international conferences and in one journal paper.

[Sil11a] S. Silva Pereira and A. Pagès-Zamora, "Consensus in random WSNs with correlated symmetric links", *Proceedings of the 12th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC'11)*, pp. 136–140, San Francisco, USA, June 2011.

[Sil11b] S. Silva Pereira and A. Pagès-Zamora, "When gossip meets consensus: convergence in correlated random WSNs", *International Conference on Wireless Technologies for Humanitarian Relief (ACWR2011), Invited Paper*, Kochi, India, December 2011.

[Sil11c] S. Silva Pereira and A. Pagès-Zamora, "Consensus in correlated random wireless sensor networks", *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 6279–6284, December 2011.

Chapter 7

This chapter summarizes the results of this PhD thesis and discusses open problems.

2

Graph Theory Concepts

The information flow among the nodes of a WSN is usually described using graphs. A graph is a mathematical abstraction used to represent binary relations among the elements of a set. These elements are called vertices and the relations are described by edges between pairs of vertices. In terms of a WSN, the vertices represent the sensor nodes and the edges represent the possibly time-varying wireless communication links among these nodes.

The connectivity of a WSN, described by the Laplacian matrix of its underlying graph model, determines the capacity of the network to reach a consensus and characterizes the convergence rate of the consensus algorithm [Fax04]. In particular, a symmetric Laplacian matrix, associated with communication links exhibiting symmetry, is commonly assumed in the study of consensus algorithms converging to the average of the initial values. The spectral properties of the Laplacian matrix play an important role on the convergence analysis of consensus algorithms, since the stability of the system is determined by the location of its eigenvalues.

This chapter reviews important concepts on algebraic graph theory and presents the notation used in the forthcoming chapters for the convergence analysis of consensus algorithms [God01]. We start defining fundamental concepts and the matrices associated with a graph in section 2.1. Notions of connectivity and the concepts of subgraphs and trees, important to understand the results presented in the state of the art of consensus algorithms, are given in section 2.2 and 2.3 respectively. Typical graph topologies are presented in section 2.4, and the Laplacian matrix associated with a graph along with its spectral properties are described in section 2.5. Finally, section 2.6 introduces the terminology used for deterministic time-varying graphs and random graphs, where we have borrowed concepts from [Bol01, Erd60].

2.1 Fundamental Concepts of Graph Theory

A graph is defined as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is the set of *vertices* indexed with $i = \{1, \cdots, N\}$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is an unordered set of pairs of vertices from $\mathcal{V}$ called *edges* representing a connection between two vertices, where the total number of vertices is $M \leq N^2$. The edge between two vertices i and j is denoted e_{ij} and refers to the information flowing from vertex j to vertex i.

If a direction is assigned to the edges, the relations are asymmetric and the graph is called a *directed* graph, or a *digraph*. For a directed edge e_{ij}, j is called the *head* and i is called the *tail* of edge e_{ij}. On the other hand, if no direction is assigned to the edges, then $e_{ij} \in \mathcal{E} \Leftrightarrow e_{ji} \in \mathcal{E}$ for all pairs $\{i, j\} \in \mathcal{V}$ and the graph is called an *undirected* graph. Graphs are graphically visualized using diagrams where vertices are depicted with points and edges are depicted with lines from one vertex to another, while directed edges are depicted with arrows. Examples of a graph composed of $N = 5$ vertices and $M = 6$ edges, with undirected and directed edges are depicted respectively in Fig. 2.1 and Fig. 2.2.

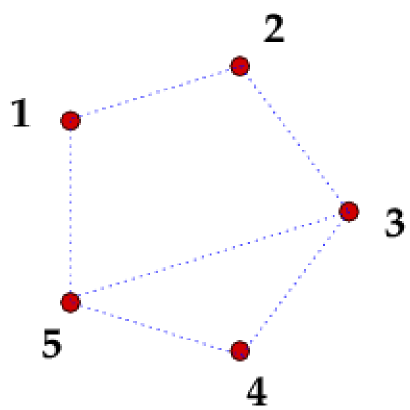


Figure 2.1: Undirected graph with $N = 5$ vertices and $M = 6$ edges.

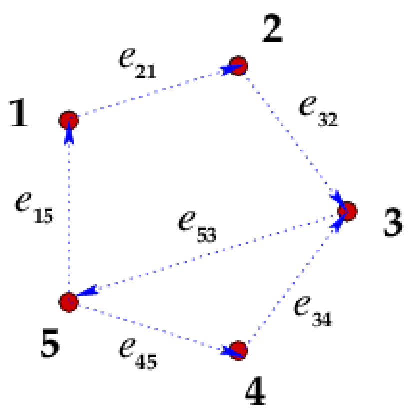


Figure 2.2: Directed graph with $N = 5$ vertices and $M = 6$ edges.

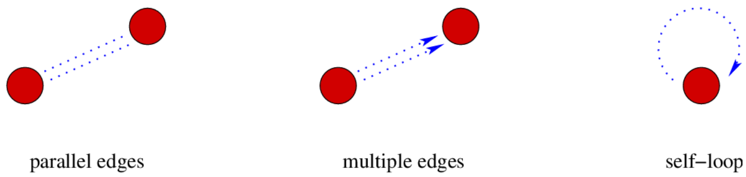


Figure 2.3: Three diagrams depicting different binary relations.

Two undirected edges with the same end vertices are called *parallel* edges, while two directed edges with the same head and tail are called *multiple* edges. An edge that connects a vertex with itself is called a *self-loop*. These particular cases are depicted in Fig. 2.3.

An *oriented* graph is a directed graph without loops or multiple edges. An oriented graph can be also seen as the result of assigning an arbitrary direction to each edge of an undirected graph. Therefore, the set of edges of an oriented graph has one and only one of the two edges $\{e_{ij}, e_{ji}\}$. For instance, the example graph in Fig. 2.2 is also an oriented version of the undirected graph in Fig. 2.1.

A graph is called *weighted* if a weight is associated with every edge according to a proper map $W : \mathcal{E} \to \mathbb{R}$, such that if $e_{ij} \in \mathcal{E}$, then $W(e_{ij}) \neq 0$, otherwise $W(e_{ij}) = 0$.

Two vertices joined by an edge are called the *endpoints* of the edge. If vertex i and vertex j are endpoints of the same edge, then i and j are said to be *adjacent* to each other. An edge is said to be *incident* on a vertex if the vertex is one endpoint of the edge. Two edges are *adjacent* if they have a common endpoint. The *outgoing* edges of a vertex i are the directed edges whose origin is vertex i. The *incoming* edges of a vertex i are the directed edges whose destination is vertex i.

In undirected graphs, vertices that are adjacent to a vertex i are called the *neighbors* of i. In directed graphs, the neighbors of a vertex i are those vertices that have an outgoing edge to i.

The set of all neighbors of a vertex i is defined as

$$\mathcal{N}_i \triangleq \{j \in \mathcal{V} \,:\, e_{ij} \in \mathcal{E}\}\,.$$

The edge structure of a graph $\mathcal{G}$ with N nodes is described by means of an $N \times N$

matrix. The *adjacency matrix* $\mathbf{A}$ of $\mathcal{G}$ is the matrix with entries a_{ij} given by

$$a_{ij} = \begin{cases} 1 & \text{if } e_{ij} \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

i.e., the $\{ij\}^{th}$ entry of $\mathbf{A}$ is 1 only if vertex j is a neighbor of vertex i. If $\mathcal{G}$ is weighted, then $a_{ij} = W(e_{ij})$ for all $e_{ij} \in \mathcal{E}$. Further, if $\mathcal{G}$ has no self-loops $a_{ii} = 0$, i.e., the diagonal entries of the adjacency matrix are all equal to 0. If $\mathcal{G}$ is undirected, $a_{ij} = a_{ji}$, i.e., $\mathbf{A}$ is symmetric. For instance, the adjacency matrices for the undirected and for the directed example graphs in Fig. 2.1 and Fig. 2.2 are given respectively by

$$\mathbf{A}_U = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}, \quad \mathbf{A}_D = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

If we consider a given ordering $\{1, 2, \cdots, M\}$ of the edge set $\mathcal{E}$, the *incidence matrix* $\mathbf{B}$ of an undirected graph $\mathcal{G}$ is the $N \times M$ matrix with entries

$$b_{il} = \begin{cases} 1 & \text{if vertex } i \text{ is incident with edge } l \\ 0 & \text{otherwise} \end{cases}.$$

In other words, b_{il} is 1 if vertex i is in the edge l, or equivalently edge l is incident to vertex i. For digraphs, the incidence matrix is a $(0, \pm 1)$-matrix such that

$$b_{il} = \begin{cases} 1 & \text{if vertex } i \text{ is the tail of edge } l \\ -1 & \text{if vertex } i \text{ is the head of edge } l \\ 0 & \text{otherwise} \end{cases}. \tag{2.2}$$

The incidence matrix for the example graphs in Fig. 2.1 and Fig. 2.2 are given respectively by

$$\mathbf{B}_U = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{B}_D = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{bmatrix}, \tag{2.3}$$

where the rows correspond to the ordering of the vertices and the columns correspond to the following order of the edges

$$e_{21} = 1, \quad e_{32} = 2, \quad e_{34} = 3, \quad e_{45} = 4, \quad e_{15} = 5, \quad e_{53} = 6.$$

Due to the arbitrary ordering of the edges, the incidence matrix is not unique. However, the different versions of an incidence matrix for a given set of nodes vary only by column permutation.

The in-degree and out-degree of a vertex i are determined by the sums of the weights of the outgoing and the incoming edges respectively, i.e.,

$$d_i^{in} = \sum_{j=1}^{N} a_{ji}, \quad \text{and} \quad d_i^{out} = \sum_{j=1}^{N} a_{ij}.$$

A vertex i is said to be *balanced* if its in-degree and out-degree are equal, i.e., $d_i^{in} = d_i^{out}$. A digraph $\mathcal{G}$ is called *balanced* if all its vertices are balanced. Therefore, all undirected graphs are balanced graphs.

The *degree matrix* $\mathbf{D}$ of $\mathcal{G}$ is the $N \times N$ diagonal matrix with $\{ij\}^{th}$ entry given by

$$\mathbf{D}_{ij} = \begin{cases} d_i^{out} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

where d_i^{out} is the out-degree of vertex i. For graphs with unit weights, that is $a_{ij} = 1$ for each $e_{ij} \in \mathcal{E}$, the diagonal entries of $\mathbf{D}$ coincide with the number of incoming edges for each vertex, i.e., $\mathbf{D}_{ii} = \mathcal{N}_i$ for all $i \in \mathcal{V}$. The entries of the degree matrix are equal to the row sums of the adjacency matrix, that is

$$\mathbf{D} = diag(\mathbf{A} \cdot \mathbf{1})$$

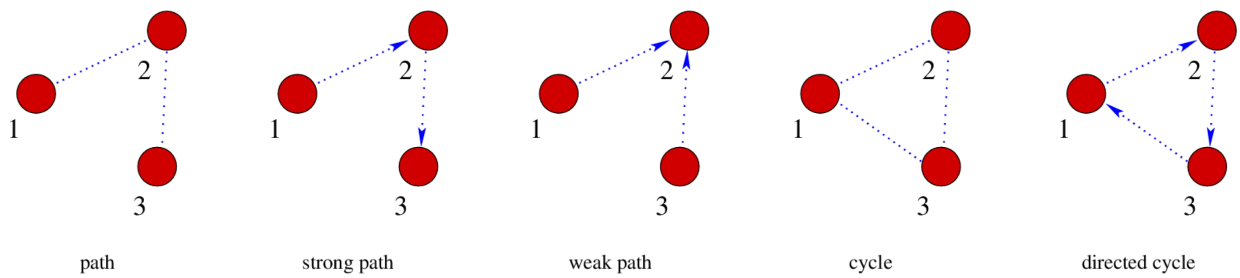
where $\mathbf{1} \in \mathbb{R}^{N \times 1}$ is the vector of all ones and $diag(\mathbf{v})$ refers to the $N \times N$ diagonal matrix whose entries are the elements of a vector $\mathbf{v} \in \mathbb{R}^{N \times 1}$.

2.2 Connectivity in Undirected and Directed Graphs

This section provides important definitions regarding the connectivity of a graph having either undirected or directed edges. We start defining the concept of path, which leads to the concept of connectivity.

A *path* from a vertex i to a vertex j is a sequence of distinct vertices starting with vertex i and ending with vertex j such that consecutive vertices are adjacent. A *simple path* is a path with no repeated vertices. A *directed path* is a path with directed edges. A *strong path* in a digraph is a sequence of distinct vertices with consecutive order $1, \cdots, q \in \mathcal{V}$ such that $e_{i,i-1} \in \mathcal{E}$, $\forall i = 2, \cdots, q$. A *weak* path is a sequence of distinct vertices with consecutive order $1, \cdots, q \in \mathcal{V}$ such that either $e_{i-1,i} \in \mathcal{E}$ or $e_{i,i-1} \in \mathcal{E}$.

A *cycle* is a closed path that starts and ends at the same vertex, and visits each other vertex only once. A *directed cycle* is a cycle where all the edges are directed.



For instance, for the directed graph example in Fig. 2.2, the sequence of vertices $\{1, 2, 3\}$ with edge set $\{e_{21}, e_{32}\}$ is a strong path, the sequence $\{2, 3, 4\}$ with edge set $\{e_{32}, e_{34}\}$ is a weak path and the sequence $\{5, 4, 3\}$ with edge set $\{e_{45}, e_{34}, e_{53}\}$ is a directed cycle.

In an undirected graph $\mathcal{G}$, two vertices i and j are *connected* if there is a path from i to j, or equivalently from j to i. An undirected graph is therefore connected if for any two vertices in $\mathcal{G}$ there is a path between them. Conversely, two vertices i and j in an undirected graph are *disconnected* if there is no path from i to j. An undirected graph is disconnected if we can partition its vertices into two nonempty sets $\mathcal{X}$ and $\mathcal{Y}$ such that no vertex in $\mathcal{X}$ is adjacent to a vertex in $\mathcal{Y}$.

Whereas undirected graphs are either connected or disconnected, we can differentiate between different forms for connectivity in digraphs. A digraph is *strongly connected* if any ordered pair of distinct vertices can be joined by a strong path. A digraph is *quasi-strongly connected* if for every ordered pair of vertices i and j, there exists another vertex that can reach either i or j by a strong path. A digraph is *weakly connected* if any ordered pair of distinct vertices can be joined by a weak path. An alternative definition of weakly connection is as follows: a digraph is weakly connected if the equivalent undirected graph, i.e., the graph with no direction assigned to the edges, is connected. A digraph is *disconnected* if it is not weakly connected, i.e., if the equivalent undirected graph is disconnected. Strong connectivity in digraphs implies quasi-strong connectivity, and quasi-strong connectivity implies weak connectivity, but the converse does not hold in general.

A graph is called a *regular graph* if all the vertices have the same number of neighbors. The graph is denoted n*-regular* if the number of neighboring vertices is n for all vertices. A graph is said to be *complete* if every pair of vertices has an edge connecting them. In other words, every vertex is adjacent to every other, such that the number of neighbors for each vertex is equal to $N - 1$. A complete graph is also known as a *fully-connected* graph.

2.3 Subgraphs and Trees

Consider a graph $\mathcal{G}=\{\mathcal{V}, \mathcal{E}\}$ with vertex set $\mathcal{V}$ and edge set $\mathcal{E} = \{e_{ij} : i, j \in \mathcal{V}\}$:

A *subgraph* $\mathcal{G}_s=\{\mathcal{V}_s, \mathcal{E}_s\}$ of $\mathcal{G}$ is a graph whose vertices and edges are subsets of the vertices and edges of $\mathcal{G}$ respectively, such that $\mathcal{V}_s \subseteq \mathcal{V}$ and $\mathcal{E}_s \subseteq \mathcal{E}$. The graph $\mathcal{G}$ is then called the *supergraph* of $\mathcal{G}_s$.

A *spanning subgraph* of $\mathcal{G}$ is a subgraph $\mathcal{G}_s=\{\mathcal{V}_s, \mathcal{E}_s\}$ that contains all the vertices of $\mathcal{G}$, i.e., $\mathcal{V}_s \equiv \mathcal{V}$, and $\mathcal{E}_s \subseteq \mathcal{E}$. Any spanning subgraph of $\mathcal{G}$ can be obtained by deleting some of the edges from $\mathcal{E}$.

An *induced subgraph* of $\mathcal{G}$ is a subgraph $\mathcal{G}_s=\{\mathcal{V}_s, \mathcal{E}_s\}$ such that for any pair of vertices i and j of $\mathcal{V}_s \subseteq \mathcal{V}$, $e_{ij} \in \mathcal{E}_s \Leftrightarrow e_{ij} \in \mathcal{E}$. Any induced subgraph of $\mathcal{G}$ can be obtained by deleting some vertices in $\mathcal{V}$ along with any edges in $\mathcal{E}$ incident to a deleted vertex.

A *maximal subgraph* of $\mathcal{G}$ is a subgraph $\mathcal{G}_s$ such that there are no other node in $\mathcal{G}$ that could be added to $\mathcal{G}_s$, and all the nodes in the subgraph would still be connected.

A *cliqué* C of $\mathcal{G}$ is a complete subgraph of $\mathcal{G}$ which is also an induced graph. A cliqué C is maximal if no vertex of $\mathcal{G}$ outside of C is adjacent to all members of C.

A *connected component* of an undirected graph $\mathcal{G}$ is an induced subgraph that is connected and maximal. A connected graph has exactly one connected component, whereas a disconnected graph is the *disjoint union* of two or more connected components. A *strongly connected component* of a digraph is a subgraph that is strongly connected.

A particular case of a quasi-strongly connected digraph is the *tree*, where only one vertex has a directed path to every other vertex in the graph. This distinguished vertex is called the *root* and has no incoming edges, whereas every vertex i distinct from the root, has only one incoming edge[1]. A n-ary tree is a tree in which every internal vertex has n outgoing edges. In a *binary* tree, each vertex has at most two outgoing edges.

A subgraph $\mathcal{G}_s$=$\{\mathcal{V}_s, \mathcal{E}_s\}$ of $\mathcal{G}$ is a *spanning tree* if it is a tree and a spanning subgraph.

A graph is a *forest* if it consists in one or more trees with no vertices in common.

A subgraph $\mathcal{G}_s$=$\{\mathcal{V}_s, \mathcal{E}_s\}$ of $\mathcal{G}$ is a *spanning forest* if it is a forest and $\mathcal{V}_s \equiv \mathcal{V}$.

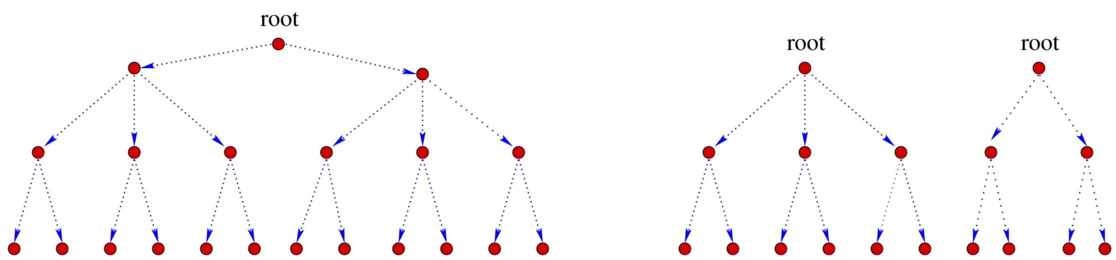
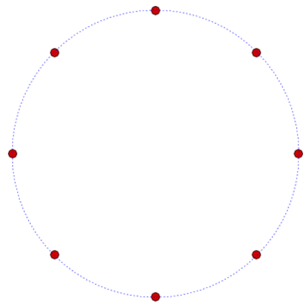


Figure 2.4: Example tree composed of 21 vertices and example forest composed of two trees of 10 and 7 vertices respectively.

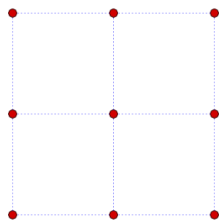
[1] For undirected graphs, a tree is a graph where any two vertices are connected by exactly one simple path, i.e., a graph with no cycles.

2.4 Examples of Common Graph Topologies

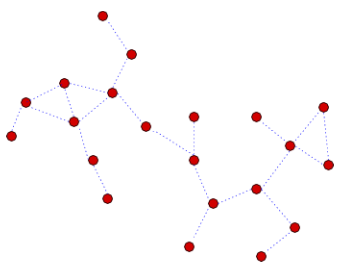
We overview some common topologies assumed in the study of graphs; some of them are used throughout the analysis carried out in this thesis. The following models apply to both undirected and directed graphs.



The ring A *ring* is a one-dimensional grid where the vertices are spatially distributed forming a circle. The ring is a 2-regular graph, since every node has exactly two neighbors.

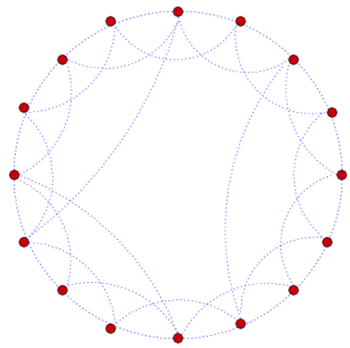


The lattice A *lattice* is a topology where the vertices are spatially distributed according to a two-dimensional grid. The number of neighbors for an internal vertex is 4, whereas for an external vertex is 2.

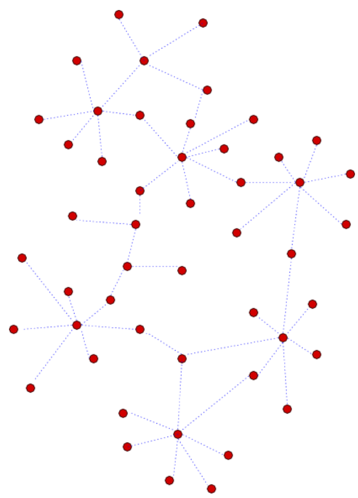


Random geometric graph A *random geometric* graph consists in a set of vertices randomly spread in a 2-dimensional area where every pair of vertices are connected whenever the euclidean distance between them is smaller than a given radius [Pen03]. In such topologies, the radius must be asymptotically larger than $\sqrt{\log N/N}$ to guarantee the connectivity of the graph [Gir05].

Small-world network In a *small-world* network, most of the vertices are not neighbors of each other, but most vertices can be reached from every other by a small number of hops [Wat98]. Starting from a regular grid, random connections can be established between vertices by rewiring existing edges or by adding new ones with nonzero probability.



Scale-free network In a *scale-free* network, the distribution of the vertex degrees follows a power law [Cal07]. In these graphs, some vertices are highly connected but most vertices have a low number of connections. Scale-free networks have a number of vertices which can sum up to some millions, and therefore statistical distributions are used for descriptions rather than quantities. Examples of these structures are commonly found in nature and in technology, like for instance the *Internet* or the *World Wide Web*.



2.5 The Laplacian Matrix

This section introduces an important matrix associated with a graph, known as the *connectivity* matrix or the *Laplacian* matrix, used for mathematical convenience to describe the connectivity in a more compact form. In general, the spectral properties of the Laplacian are of prime importance in the convergence analysis of the consensus algorithm, as we will see in Chapter 3.

The Laplacian matrix of a graph $\mathcal{G}$ has $\{ij\}^{th}$ entry given by

$$\mathbf{L}_{ij} = \begin{cases} d_i^{out} & i = j \\ -a_{ij} & i \neq j \end{cases} .$$

This definition can be expressed in matrix form as follows

$$\mathbf{L} = \mathbf{D} - \mathbf{A}$$

where $\mathbf{D}$ and $\mathbf{A}$ are the degree and the adjacency matrix of $\mathcal{G}$ with entries defined in (2.4) and (2.1) respectively. Recalling the example graphs in Fig. 2.1 and Fig. 2.2, the corresponding Laplacian matrices are given respectively by

$$\mathbf{L}_U = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & -1 & 3 \end{bmatrix}, \quad \mathbf{L}_D = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 0 & 1 \end{bmatrix}.$$

If $\mathcal{G}$ is undirected, $\mathbf{L}$ can be also defined as the $N \times N$ matrix

$$\mathbf{L} = \mathbf{B}_D \mathbf{B}_D^T$$

where $\mathbf{B}_D$ is the incidence matrix of $\mathcal{G}$ with an arbitrary orientation, and build using (2.2). As an example, note that the Laplacian $\mathbf{L}_U$ of the undirected graph in Fig. 2.1 can be also obtained using the incidence matrix $\mathbf{B}_D$ in (2.3) corresponding to the directed case depicted in Fig. 2.2.

By construction, the Laplacian matrix of an undirected graph is always symmetric, whereas the Laplacian of a digraph is not.

2.5.1 Spectral Properties of the Laplacian

In this section we review the spectral properties of the Laplacian matrix associated with a graph. According to the Geršgorin circle theorem [Hor06], the eigenvalues of the Laplacian $\mathbf{L}$ of a graph $\mathcal{G}$ are located inside the discs in the complex plane with centers in $\mathbf{L}_{ii}$ and radii given by the row-sums $\sum_{\substack{j=1 \\ j \neq i}}^{N} |\mathbf{L}_{ij}|$ for each i, where $|\cdot|$ denotes absolute value. Since by definition the diagonal entries of $\mathbf{L}$ are nonnegative and all row-sums are equal to zero, the Geršgorin circles are tangent to the imaginary axis at zero. Fig. 2.5 visualizes an example of Geršgorin circles for the Laplacian in the complex plane.

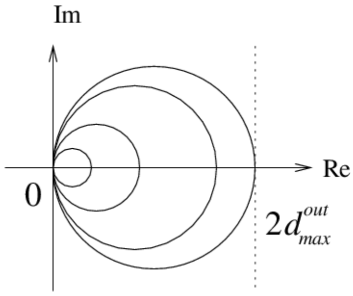


Figure 2.5: Geršgorin circles for the eigenvalues of the Laplacian matrix.

Therefore, the eigenvalues of $\mathbf{L}$ have nonnegative real parts and are all inside a circle of radius $2d_{max}^{out}$ where

$$d_{max}^{out} = \max_{i \in \mathcal{V}} d_i^{out}.$$

Since $\mathbf{L} \cdot \mathbf{1} = \mathbf{0}$, where $\mathbf{0}$ is a zero vector of length N, $\mathbf{L}$ has at least one eigenvalue zero with associated right eigenvector $\mathbf{1}$.

In undirected graphs, the associated Laplacian is positive semidefinite and its eigenvalues can be arranged in non-increasing order as follows

$$2d_{max}^{out} \geq \lambda_1(\mathbf{L}) \geq \cdots \geq \lambda_N(\mathbf{L}) = 0.$$

In addition, both the left and the right eigenvector associated with $\lambda_N(\mathbf{L})$ is $\mathbf{1}$ such that $\mathbf{1}^T\mathbf{L} = \mathbf{0}^T$. In weighted digraphs however, a necessary and sufficient condition for $\mathbf{1}$ to be the left eigenvector associated with $\lambda_N(\mathbf{L})$ is that the digraph is balanced [OS04]. The second smallest eigenvalue $\lambda_{N-1}(\mathbf{L})$ is known as the *algebraic connectivity* and reflects the degree of connectivity of the graph [Fie73]. The algebraic connectivity can be used to define the *spectral gap*, a quantity useful to get insight into important properties of the graph like expansion[2] and the mixing time of random walks [Hoo06]. In some cases, the term spectral gap is directly used to refer to $\lambda_{N-1}(\mathbf{L})$.

As it will be seen in Chapter 3, the conditions to reach a consensus are related to the algebraic multiplicity of the eigenvalue zero of $\mathbf{L}$. A necessary and sufficient condition for $\lambda_N(\mathbf{L})$ to have algebraic multiplicity one in undirected graphs, i.e., the algebraic connectivity is nonzero, is that the graph is connected [Chu97]. If the algebraic multiplicity of

[2] Expansion of a graph requires that it is simultaneously sparse and highly connected.

$\lambda_N(\mathbf{L})$ is one, $\mathbf{L}$ is an irreducible matrix[3]. For digraphs however, a necessary and sufficient condition for $\lambda_N(\mathbf{L})$ to have algebraic multiplicity one is that the associated digraph has a spanning tree, i.e., there is at least one vertex that can communicate with any other vertex in the network [Ren05a]. Furthermore, the Laplacian of a digraph is irreducible if and only if the digraph is strongly connected [Chu97]. Therefore, whereas in undirected graphs a necessary and sufficient condition for the eigenvalue zero of $\mathbf{L}$ to have algebraic multiplicity one is that $\mathbf{L}$ is irreducible, in directed graphs irreducibility of $\mathbf{L}$ is only a sufficient condition.

If the geometric multiplicity of the eigenvalue zero is one, then $rank(\mathbf{L}) = N - 1$. In general, if the graph has c connected components then $rank(\mathbf{L}) = N - c$. The proof of this property for undirected graphs can be found in [God01], whereas the proof of this property for digraphs is given in [OS04].

2.6 Graphs with Time-Varying Topology

This section introduces the notation to be used for time-varying graphs. We denote by $\mathcal{G}(k) = \{\mathcal{V}, \mathcal{E}(k)\}$ the graph with fixed vertex set $\mathcal{V} = \{1, \cdots, N\}$ and time-varying edge set $\mathcal{E}(k)$, where the edges can vary with time either deterministically or completely random.

The instantaneous set of neighbors of vertex i is denoted $\mathcal{N}_i(k) = \{j \in \mathcal{V} : e_{ij} \in \mathcal{E}(k)\}$ and the adjacency matrix is time-varying and denoted $\mathbf{A}(k)$. The Laplacian matrix is also time-varying and modeled as

$$\mathbf{L}(k) = \mathbf{D}(k) - \mathbf{A}(k)$$

where $\mathbf{D}(k)$ is the instantaneous degree matrix. Analogously to the time-invariant case, the instantaneous Laplacian satisfies $\mathbf{L}(k)\mathbf{1} = \mathbf{0}$ for all k by construction. Additionally, if the instantaneous edges are balanced, $\mathbf{1}^T\mathbf{L}(k) = \mathbf{0}^T$ is satisfied for all k.

[3] A matrix is irreducible if it is not similar to a block upper triangular matrix with two blocks via a permutation [Hor06].

2.6.1 Deterministic Time-Varying Topologies

If the time-varying set of edges $\mathcal{E}(k)$ belongs to a known finite set $\{\mathcal{E}_1, \mathcal{E}_2, \cdots, \mathcal{E}_K\}$ with $K < \infty$, the topology of the graph is called a *switching topology.* This time-varying yet deterministic topology was defined in [OS04] to study the convergence of consensus algorithms in directed networks. In that contribution, the instantaneous topologies are assumed strongly connected and balanced for all k, i.e., $\mathbf{L}(k)$ is irreducible and the eigenvalue zero of $\mathbf{L}(k)$ has $\mathbf{1}$ as both the left and the right associated eigenvectors for all k. Therefore, the total number of elements in the set $\mathcal{E}(k)$ is bounded above with $K \leq N(N-1)$.

Another connectivity concept is the *periodical connectivity*, defined in [Jad03] to reach a consensus in time-varying undirected topologies. Consider a finite collection of graphs $\mathcal{S}_\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_K\}$ for some $K < \infty$ with vertex set $\{\mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_K\}$. The *union* of the graphs in $\mathcal{S}_\mathcal{G}$ is a graph $\mathcal{G}_u = \mathcal{G}_1 \bigcup \mathcal{G}_2 \bigcup \cdots \bigcup \mathcal{G}_K$ with vertex set $\mathcal{V}_u = \mathcal{V}_1 \bigcup \mathcal{V}_2 \bigcup \cdots \bigcup \mathcal{V}_K$ and whose set $\mathcal{E}_u = \mathcal{E}_1 \bigcup \mathcal{E}_2 \bigcup \cdots \bigcup \mathcal{E}_K$ is the union of all the edges of the graphs in $\mathcal{S}_\mathcal{G}$. A collection of graphs $\mathcal{S}_\mathcal{G}$ is said to be *jointly connected* if the union of graphs $\mathcal{G}_u$ is connected. A collection $\mathcal{S}_\mathcal{G}$ can be jointly connected even if none of its members is connected. If a collection $\mathcal{S}_\mathcal{G}$ is jointly connected frequently enough over a finite sequence of intervals, we say that the resulting graph is *periodically connected.*

2.6.2 Random Topologies

A random graph $\mathcal{G}(k)$ is a graph generated by some random process [Bol01]. Typically, the set of vertices $\mathcal{V}$ is assumed constant throughout time whereas the set of edges $\mathcal{E}(k)$ varies randomly with time. A general way of modeling the randomness of the edges consists in assuming a probability of connection between two vertices i and j, such that $e_{ij} \in \mathcal{E}(k)$ with probability $0 \leq p_{ij} \leq 1$. Let's define the $N \times N$ connection probability matrix $\mathbf{P}$ with $\{ij\}^{th}$ entry

$$\mathbf{P}_{ij} = \begin{cases} p_{ij} & i \neq j \\ 0 & i = j \end{cases}. \tag{2.5}$$

The instantaneous adjacency matrix $\mathbf{A}(k) \in \mathbb{R}^{N \times N}$ of $\mathcal{G}(k)$ is random with statistically independent entries over time given by

$$\mathbf{A}_{ij}(k) = \begin{cases} 1 & \text{with probability } p_{ij} \\ 0 & \text{with probability } 1 - p_{ij} \end{cases}. \tag{2.6}$$

Then, a realization $\mathcal{G}(k)$ at time k can be seen as a spanning subgraph of a graph with fixed edge set. Due to the random nature of $\mathbf{A}(k)$, the instantaneous degree matrix $\mathbf{D}(k)$ and the instantaneous Laplacian $\mathbf{L}(k)$ are also random and so are their corresponding eigenvalues. The adjacency matrix has expected value $\mathbb{E}[\mathbf{A}(k)] = \bar{\mathbf{A}} = \mathbf{P}$ whereas the degree matrix has expected value $\bar{\mathbf{D}} = diag(\mathbf{P} \cdot \mathbf{1})$. Analogously, the mean of the Laplacian is given by

$$\bar{\mathbf{L}} = \bar{\mathbf{D}} - \mathbf{P}.$$

Note that $\bar{\mathbf{L}}$ can be seen as the Laplacian of the supergraph with fixed edge set, which can be defined as $\mathbb{E}[\mathcal{G}(k)] = \bar{\mathcal{G}}$.

Erdős-Rényi models

Two well-known models of random graphs were introduced by Erdős and Rényi [Erd60], each with a different way of modeling the randomness of the edges:

1. The Erdős-Rényi model $\mathcal{G}(k) = \{\mathcal{V}, M\}$ refers to a random graph where at each realization there are N vertices and exactly M edges whose endpoints may vary from realization to realization. In other words, at time k a graph $\mathcal{G}(k)$ is chosen uniformly at random from the collection of all graphs which have N vertices and M edges.

2. The Erdős-Rényi model $\mathcal{G}(k) = \{\mathcal{V}, p\}$ refers to a graph with N vertices where each edge exists with nonzero probability p, equal for all the vertices. For this model, any pair of vertices in the network can be connected with the same probability p regardless of their position, and the number of edges varies generally from realization to realization. The adjacency matrix for this model is a particular case of the general adjacency matrix in (2.6) with $p_{ij} = p$ for all $\{i, j\} \in \mathcal{V}$. The connection probability matrix defined in (2.5) can be rewritten for this model as follows

$$\mathbf{P} = p(\mathbf{J} - \mathbf{I})$$

where $\mathbf{J} = \mathbf{1}\mathbf{1}^T$ is an all-ones matrix and $\mathbf{I}$ denotes the identity matrix, both of dimensions $N \times N$. The adjacency matrix $\mathbf{A}(k)$ and the degree matrix $\mathbf{D}(k)$ have mean given respectively by

$$\bar{\mathbf{A}} = p(\mathbf{J} - \mathbf{I}), \quad \bar{\mathbf{D}} = (N-1)p\mathbf{I}.$$

Analogously, the mean of the Laplacian matrix is given by

$$\bar{\mathbf{L}} = p(N\mathbf{I} - \mathbf{J}).$$

Remark that as $p \to 1$, the number of edges increases and with $p = 1$ we obtain a complete graph at every realization. References to Erdős-Rényi graphs made throughout this PhD thesis refer to this particular model.

3

Distributed Consensus Algorithms

3.1 Introduction

Consensus algorithms are iterative schemes where autonomous agents or nodes communicate with each other to reach an agreement regarding a certain value of interest, without the necessity to forward any information to a central node or FC. At each iteration, the nodes exchange information locally such that a common value is asymptotically reached. In particular, the average consensus algorithm computes the average of an initial set of measurements.

Consider a network composed of N nodes indexed with $i = \{1, \cdots, N\}$. Each node has an associated scalar value x_i defined as the *state* of node i -if several variables are considered, a state vector is used instead-. The state is initialized with the value of a measurement and updated iteratively using the information received from its neighbors. We say that nodes i and j have reached a consensus if $x_i = x_j$.

Consensus algorithms can be classified according to the value computed by the algorithm. When the application requires only a global agreement and the value reached is irrelevant, the algorithm is denoted *unconstrained*. On the other hand, when the application requires the computation of a function of the initial measurements, the algorithm is denoted *constrained*. For instance, let $\chi : \mathbb{R}^N \to \mathbb{R}$ be a function of N variables $x_1, \cdots, x_N$ and let $\mathbf{x}(0) = [x_1(0), \cdots, x_N(0)]^T$ be the vector of initial states of the network. Common functions of the initial measurements are

$$\chi(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} x_i(0), \qquad \textit{average consensus}$$

$$\chi(\mathbf{x}) = \frac{\sum_i d_i^{out} x_i(0)}{\sum_i d_i^{out}}, \qquad \textit{weighted average consensus}$$

$$\chi(\mathbf{x}) = \max_i \ x_i(0), \qquad \textit{max-consensus}$$

$$\chi(\mathbf{x}) = \min_i \ x_i(0), \qquad \textit{min-consensus}$$

where d_i^{out} is the out-degree of node i, as defined in Chapter 2 -Section 2.1-.

Consensus algorithms can converge either on the state or on the derivative of the state. Algorithms converging on the state are robust to changes in the topology of the network and have a bounded state value, but are sensitive to propagation delays and additive noise. On the other hand, consensus algorithms converging on the state derivative are resilient to propagation delays or coupling noise, but are sensitive to changes in the topology of the network and have unbounded state value.

In a *continuous-time* implementation, a simple algorithm to reach a consensus regarding the state of N integrator agents can be expressed as an N^{th}-order linear system

$$\dot{x}_i(t) = \sum_{j\in\mathcal{N}_i} a_{ij}(x_j(t) - x_i(t)) \tag{3.1}$$

where a_{ij} is the $\{ij\}^{th}$ entry of the adjacency matrix associated to the underlying graph of the network and $\mathcal{N}_i$ denotes the set of neighbors of node i [OS04]. The state of the network evolves according to the following linear system

$$\dot{\mathbf{x}}(t) = -\mathbf{L}\mathbf{x}(t) \tag{3.2}$$

where $\mathbf{x}(t) = [x_1(t), x_2(t), \cdots, x_N(t)]^T$ is the vector of all states at time t and $\mathbf{L} \in \mathbb{R}^{N\times N}$ is the Laplacian matrix associated to the graph. An equilibrium of the system in (3.1) is a state of the form $\mathbf{x}^* = [c, \cdots, c]^T$, with $c \in \mathbb{R}$, unique for connected graphs and globally exponentially stable [OS07]. For the case of undirected graphs, the algorithm in (3.2) is a gradient-descent algorithm.

A generalized nonlinear extension of (3.1) for self-synchronization of mutually coupled oscillators is given by

$$\dot{x}_i(t) = g_i(y_i) + \frac{K_C}{z_i}\sum_{j\in\mathcal{N}_i} a_{ij} f\Big(x_j(t) - x_i(t)\Big) \tag{3.3}$$

where $g_i(y_i)$ is a function of the local observation y_i, f is a nonlinear odd function, $K_C > 0$ measures the coupling strength and z_i is a local parameter to control the final equilibrium value [Bar07a]. The model in (3.3) converges on the state derivative and coincides with the Kuramoto model [Kur03] when $f(x) = \sin(x)$ and $a_{ij} = z_i = 1$, for all $\{i, j\}$.

A *discrete* implementation of the expression in (3.1) yields the difference equation given by

$$x_i(k+1) = \sum_{j \in \mathcal{N}_i \bigcup \{i\}} w_{ij} x_j(k), \quad k \geq 0 \tag{3.4}$$

where w_{ij} is a nonzero *weight* assigned by node i to the information received from node j, satisfying

$$\sum_{j \in \mathcal{N}_i \bigcup \{i\}} w_{ij} = 1. \tag{3.5}$$

The weight w_{ij} can be seen as the degree of confidence assigned by node i to the information received from node j [Ren05b].

Since discrete-time algorithms can be directly loaded to a digital device, in this thesis we focus on discrete-time implementations of linear consensus algorithms converging on the state, i.e., the consensus of the form (3.4) with link weights satisfying (3.5).

Outline of the Chapter

This chapter is devoted to a detailed review of discrete-time consensus algorithms of the form (3.4). We start presenting the time-invariant consensus model by [OS04] in Section 3.2 and review the convergence conditions to reach a consensus as well as the conditions to reach the average consensus, assuming directed topologies in Sections 3.2.1 and assuming undirected topologies in Section 3.2.2. In Section 3.3 we present the consensus model for networks with time-varying topology, where we distinguish between deterministic time-varying topologies in Section 3.3.1 and random time-varying topologies in Section 3.3.2, and introduce different forms of probabilistic convergence for the state vector. In Section 3.4 we describe common approaches to model the weight matrix with special focus on the uniform weights model and its convergence conditions. In Section 3.5 we discuss some approaches to minimize the convergence time of the consensus algorithm in fixed topologies and in Section 3.6 we resume the conclusions of the chapter.

3.2 Consensus Algorithms: The Time-Invariant Model

Consider a WSN with information flow characterized by a fixed underlying graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ with node set $\mathcal{V} = \{1, \cdots, N\}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. At time k, each node i updates its state using a linear combination of its own previous value and the information received from its neighbors as follows

$$x_i(k) = \mathbf{W}_{ii} x_i(k-1) + \sum_{j \in \mathcal{N}_i} \mathbf{W}_{ij} x_j(k-1), \quad k > 0$$

where $\mathbf{W}_{ij}$ is the $\{ij\}^{th}$ entry of the *weight matrix*, and is the weight associated to the edge e_{ij}. According to the model in (3.4) we have

$$\mathbf{W}_{ij} = \begin{cases} w_{ij} & e_{ij} \in \mathcal{E} \\ 0 & e_{ij} \notin \mathcal{E} \end{cases} \qquad \forall i, j \in \mathcal{V}$$

i.e., a nonzero entry of the weight matrix corresponds to a node i receiving information from node j. Consider the vector of all the states of the network $\mathbf{x}(k)$ at time k. The evolution of $\mathbf{x}(k)$ can be written in matrix form as follows

$$\mathbf{x}(k) = \mathbf{W}\mathbf{x}(k-1), \quad \forall k > 0 \tag{3.6}$$

or equivalently

$$\mathbf{x}(k) = \mathbf{W}^k \mathbf{x}(0), \quad \forall k > 0. \tag{3.7}$$

We say that the state of the nodes reach a consensus asymptotically if

$$\lim_{k \to \infty} \mathbf{x}(k) = c\mathbf{1} \tag{3.8}$$

where $c \in \mathbb{R}$ is the *consensus value* and $\mathbf{1} \in \mathbb{R}^{N \times 1}$ is a vector of all-ones. The consensus vector $c\mathbf{1}$ belongs to the *agreement space* $\mathcal{A}$, defined as

$$\mathcal{A} \subset \mathbb{R}^{N \times 1} \mid \mathcal{A} = span\{\mathbf{1}\}. \tag{3.9}$$

If the consensus value is given by

$$c = \frac{1}{N} \mathbf{1}^T \mathbf{x}(0) \tag{3.10}$$

we say that the nodes asymptotically reach the *average consensus*.

We review the conditions on the weight matrix $\mathbf{W}$ for a network characterized by a time-invariant graph $\mathcal{G}$, so that $\mathbf{x}(k)$ in (3.6) reaches a consensus. We review first the conditions to reach a consensus and the conditions to reach the average consensus in networks with directed topologies in Section 3.2.1, whereas in Section 3.2.2 we address the case of networks with undirected topologies.

3.2.1 Consensus in Directed Topologies

According to the expression in (3.7), the convergence of $\mathbf{x}(k)$ to a vector of the form $c\mathbf{1}$ as in (3.8) depends on the existence of $\lim_{k\to\infty}\mathbf{W}^k$, which is independent of the initial set of values $\mathbf{x}(0)$. Suppose that $\mathbf{W}$ is diagonalizable with eigenvalues $\lambda_i(\mathbf{W})$, $i = 1,\cdots,N$. The weight matrix can be expressed using its eigenvalue decomposition as follows

$$\mathbf{W} = \mathbf{QMQ}^{-1} \tag{3.11}$$

where $\mathbf{Q}$ is a nonsingular matrix whose columns are the right eigenvectors $\{\mathbf{q}_1,\cdots,\mathbf{q}_N\}$, $\mathbf{M}$ is an $N \times N$ diagonal matrix with the eigenvalues of $\mathbf{W}$ arranged in non-increasing order of magnitude, and the columns of $(\mathbf{Q}^{-1})^T$ are the left eigenvectors $\{\mathbf{v}_1,\cdots,\mathbf{v}_N\}$ of $\mathbf{W}$. Substituting $\mathbf{W}$ for its factorization we have

$$\mathbf{W}^k = \sum_{i=1}^{N}\lambda_i(\mathbf{W})^k\mathbf{q}_i\mathbf{v}_i^T \tag{3.12}$$

where $\mathbf{v}_i^T\mathbf{q}_j = \delta_{ij}$ is satisfied for all $\{i,j\}$, and δ_{ij} denotes the Kronecker delta function. Consider now the *spectral radius* of $\mathbf{W}$ defined as

$$\rho(\mathbf{W}) = \max_i|\lambda_i(\mathbf{W})|, \quad i = 1,\cdots,N.$$

If $\rho(\mathbf{W}) > 1$, $\mathbf{W}^k$ does not converge as k increases because the sum in (3.12) grows unbounded, and therefore $\mathbf{x}(k)$ in (3.6) cannot converge to a vector of the form $c\mathbf{1}$. Further, if $\rho(\mathbf{W}) < 1$, $\mathbf{W}^k$ converges to an $N\times N$ zero matrix as k increases and $\mathbf{x}(k)$ converges to the zero vector $\mathbf{0} \in \mathbb{R}^{N\times 1}$. However, the algorithm converges for the case $\rho(\mathbf{W}) = 1$, implying that $|\lambda_i(\mathbf{W})| \leq 1$ must be satisfied for all i with -at least- one equality.

3.2.1.1 Conditions to Reach a Consensus

A consensus of the form (3.8) is asymptotically reached if the spectral radius of $\mathbf{W}$ is equal to 1, i.e., $\rho(\mathbf{W}) = \lambda_1(\mathbf{W}) = 1$ with associated right eigenvector $\mathbf{q}_1 = \mathbf{1}$ and $|\lambda_i(\mathbf{W})| < 1$

for $i = \{2, \cdots, N\}$ such that

$$\lim_{k\to\infty} \mathbf{W}^k = \mathbf{1}\mathbf{v}_1^T \tag{3.13}$$

with $\mathbf{v}_1$ satisfying $\mathbf{1}^T\mathbf{v}_1 = 1$. Then,

$$\lim_{k\to\infty} \mathbf{x}(k) = \mathbf{1}\mathbf{v}_1^T\mathbf{x}(0) \tag{3.14}$$

and the consensus value is $c = \mathbf{v}_1^T\mathbf{x}(0)$. Actually this condition is not only sufficient but also necessary, since if the algebraic multiplicity of $\lambda_1(\mathbf{W})$ is larger than one, a consensus of the form (3.8) cannot be achieved. This is easily explained considering that the algebraic multiplicity of $\lambda_1(\mathbf{W})$ is 2. Then, taking the limit of (3.12) yields

$$\lim_{k\to\infty} \mathbf{W}^k = \mathbf{1}\mathbf{v}_1^T + \mathbf{q}_2\mathbf{v}_2^T.$$

Since the eigenvectors are linearly independent by definition, this sum cannot take the form in (3.13). Similarly, (3.13) is not achieved if $\mathbf{q}_1 \neq \mathbf{1}$. Therefore, $\mathbf{x}(k)$ converges asymptotically to $c\mathbf{1}$ for any set of initial values $\mathbf{x}(0)$ if and only if $\mathbf{W}$ satisfies

$$\begin{aligned} \mathbf{W}\mathbf{1} &= \mathbf{1} \\ \rho(\mathbf{W} - \mathbf{1}\mathbf{v}_1^T) &< 1 \end{aligned} \tag{3.15}$$

where $\rho(\mathbf{W} - \mathbf{1}\mathbf{v}_1^T)$ is the second largest eigenvalue of $\mathbf{W}$ in magnitude. Summing up, the conditions in (3.15) are that the weight matrix has row-sums equal to 1 as in (3.5) and that the algebraic multiplicity of the eigenvalue one is 1. Note that these conditions apply also for weight matrices with negative entries.

Moreover, if the weight matrix is defective, the algebraic multiplicity of at least one of its eigenvalues is different from its geometric multiplicity, meaning that the number of independent eigenvectors associated to that particular eigenvalue is less than its algebraic multiplicity. The eigenvalue matrix $\mathbf{M}$ in (3.11) has therefore the form of a Jordan matrix [Hor06]. According to [Nob88, Theorem 9.30], $\mathbf{W}^k$ is bounded as $k \to \infty$ if and only if

(1) $\rho(\mathbf{W}) \leq 1$ and
(2) if $|\lambda_i(\mathbf{W})| = 1$, then its algebraic multiplicity equals its geometric multiplicity.

In other words, $\mathbf{W}$ may be a defective matrix but not in the eigenvalue 1 to ensure the convergence of $\mathbf{W}^k$. Therefore, if the conditions (1) and (2) above are satisfied, the convergence conditions derived in (3.15) apply also for defective matrices.

The convergence to a consensus has been also studied in terms of the underlying graph model. For instance, assuming that the weight matrix is nonnegative with positive diagonal elements and with row-sums equal to 1, a necessary and sufficient condition for the state vector $\mathbf{x}(k)$ to achieve a consensus asymptotically is that the underlying graph model has at least one spanning tree, meaning that there is at least one node that can communicate with all the other nodes in the network [Ren05a]. Note that nonnegative matrices with row-sums equal to 1 are called row-stochastic matrices, and all row-stochastic matrices satisfy the convergence conditions in (3.15). Therefore, having a row-stochastic $\mathbf{W}$ in (3.6) is sufficient to achieve a consensus. Moreover, if the graph has a quasi-strongly connected topology, the consensus value reached is equal to the weighted average of the entries of $\mathbf{x}(0)$ corresponding to those nodes that have a directed path to all the other nodes in the network, i.e., the roots of the spanning trees contained in the topology [Ren05b].

3.2.1.2 Conditions to Reach the Average Consensus

We review now the conditions to reach a consensus on the average of the initial values, i.e., the conditions such that the consensus value c is given by (3.10). Note that, according to (3.14) the consensus value is given by $c = \mathbf{v}_1^T \mathbf{x}(0)$. In order to compute the average of the initial values, the left eigenvector associated with $\lambda_1(\mathbf{W})$ must be $\mathbf{v}_1 = \mathbf{1}$, such that as k tends to infinity we have

$$\begin{aligned} \lim_{k\to\infty} \mathbf{W}^k &= \frac{1}{N}\mathbf{1}\mathbf{1}^T \\ &\triangleq \mathbf{J}_N. \end{aligned} \tag{3.16}$$

Then,

$$\begin{aligned} \lim_{k\to\infty} \mathbf{x}(k) &= \mathbf{J}_N \mathbf{x}(0) \\ &\triangleq \mathbf{x}_{ave} \end{aligned} \tag{3.17}$$

where $\mathbf{x}_{ave}$ is denoted the *average consensus vector*. Summing up, both the right eigenvector $\mathbf{q}_1$ and the left eigenvector $\mathbf{v}_1$ associated with $\lambda_1(\mathbf{W})$ are all-ones vectors. A right eigenvector $\mathbf{1}$ implies that after reaching a consensus the network will remain in consensus, and a left eigenvector $\mathbf{1}$ implies that the average of the state vector is preserved from iteration to iteration.

Necessary and sufficient conditions for (3.6) to reach the average consensus asymptotically, for any set of initial values $\mathbf{x}(0)$ are therefore

$$\begin{aligned} \mathbf{1}^T\mathbf{W} &= \mathbf{1}^T \\ \mathbf{W}\mathbf{1} &= \mathbf{1} \\ \rho(\mathbf{W}-\mathbf{J}_N) &< 1 \end{aligned} \tag{3.18}$$

or in other words, the matrix $\mathbf{W}$ has largest eigenvalue 1 with algebraic multiplicity one, with associated left and right eigenvector $\mathbf{1}$. The condition $\mathbf{1}^T\mathbf{W} = \mathbf{1}^T$ implies that the topology of the network is balanced.

Analogous to the previous case, the convergence to the average consensus can be related to the connectivity of the underlying graph model. Assuming that the weight matrices are row-stochastic with positive diagonal entries, a necessary and sufficient condition to solve the average consensus problem in weighted directed graphs is that the underlying graph is strongly connected and balanced [OS04]. Remark that a row-stochastic matrix with positive diagonal elements is also a primitive matrix[1]. Due to the Perron-Frobenius theorem [Hor06], the eigenvalue 1 has algebraic multiplicity one and the weight matrix satisfies the convergence conditions in (3.18). Furthermore, since the Laplacian is balanced, so is the weight matrix and the conditions to reach the average consensus in (3.18) are satisfied.

Fig. 3.1 depicts the evolution of the states for an example network composed of $N = 20$ nodes with directed communication links where the consensus value is $c = 20.38$, whereas the average of the initial values is $1/N\mathbf{1}^T\mathbf{x}(0) = 19.91$.

3.2.2 Consensus in Undirected Topologies

For the particular case of undirected topologies, that is when the communication links are bidirectional, the weight matrix $\mathbf{W}$ is symmetric and the left and right eigenvectors associated with the eigenvalue $\lambda_1(\mathbf{W})$ are equal, i.e., $\mathbf{q}_1 = \mathbf{v}_1$. Since $\mathbf{q}_1 = \mathbf{1}$ must be satisfied to reach a consensus, we have that $\mathbf{v}_1 = \mathbf{1}$ by construction. The conditions in

[1] A nonnegative matrix $\mathbf{X}$ is called primitive if there exists a K such that for all $\{i, j\}$, the $\{ij\}^{th}$ entry of $\mathbf{X}^K$ is positive. A sufficient condition for $\mathbf{X}$ to be primitive is to be nonnegative irreducible with positive entries on the main diagonal.

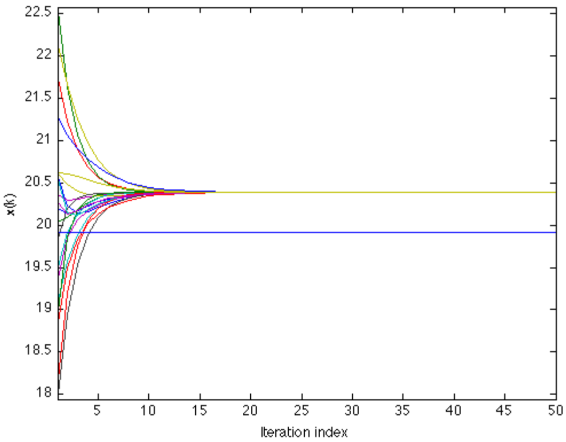


Figure 3.1: Evolution of the states in a directed network with $N = 20$ nodes and average of initial values equal to $1/N\mathbf{1}^T\mathbf{x}(0) = 19.91$.

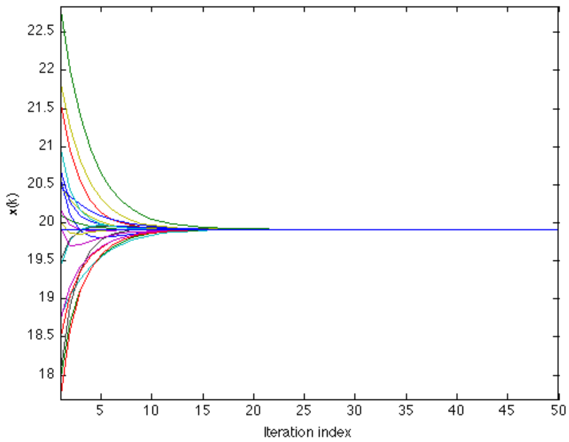


Figure 3.2: Convergence of the states to the average consensus in the same network considering undirected communication links.

(3.18) become therefore equivalent to the conditions in (3.15) and the algorithm in (3.6) asymptotically reaches the average of the initial values given by (3.10).

In terms of connectivity of the underlying graph model, a necessary and sufficient condition to solve the average consensus problem in undirected topologies is that the associated graph is connected [OS07, Lemma 1]. Under these connectivity conditions, the Laplacian matrix and the weight matrix are both irreducible.

Fig. 3.2 depicts the evolution of the states for the example network used in Fig. 3.1 but considering instead undirected communication links. The consensus value c for the undirected case is equal to the average of the initial values, i.e., $c = 1/N\mathbf{1}^T\mathbf{x}(0) = 19.91$.

The models presented so far assume that the topology of the underlying graph model is time-invariant. Although the set of nodes is usually assumed constant throughout time, the communication links can vary with time and therefore, the information flow among the nodes of the network must be described by means of a time-varying graph. The following section presents the consensus algorithm in time-varying topologies.

3.3 Consensus Algorithms: The Time-Varying Model

When the communication links of a network vary with time, the connectivity is characterized by a dynamic graph $\mathcal{G}(k) = \{\mathcal{V}, \mathcal{E}(k)\}$, where the set of edges $\mathcal{E}(k)$ varies with time while the set of nodes $\mathcal{V}$ remains constant. The consensus algorithm in (3.6) for time-varying systems can be rewritten as follows

$$\mathbf{x}(k) = \mathbf{W}(k-1)\mathbf{x}(k-1) \tag{3.19}$$

where

$$[\mathbf{W}(k)]_{ij} = \begin{cases} w_{ij} & e_{ij} \in \mathcal{E}(k) \\ 0 & e_{ij} \notin \mathcal{E}(k) \end{cases} \qquad \forall i,j \in \mathcal{V}$$

is the $\{ij\}^{th}$ entry of the time-varying weight matrix, being nonzero whenever there is information flowing from node j to node i at time k. The evolution of the state vector $\mathbf{x}(k)$ in (3.19) can be rewritten as follows

$$\mathbf{x}(k) = \prod_{l=1}^{k} \mathbf{W}(k-l)\mathbf{x}(0). \tag{3.20}$$

Clearly, the convergence of $\mathbf{x}(k)$ to a vector of the form $c\mathbf{1}$ as in (3.8) will be determined by the convergence of the product $\prod_{l=1}^{k} \mathbf{W}(k{-}l)$ to a rank one matrix, that is, a consensus is reached whenever

$$\lim_{k\to\infty} \prod_{l=1}^{k} \mathbf{W}(k-l) = \mathbf{1}\mathbf{v}_1^T. \tag{3.21}$$

The convergence above is ensured if all the matrices in the set $\{\mathbf{W}(k), \forall k\}$ satisfy the convergence conditions in (3.15). In addition, if all the weight matrices satisfy the convergence conditions in (3.18), $\mathbf{x}(k)$ converges to the average consensus vector $\mathbf{x}_{ave}$ defined in (3.17). However, the states will still converge to a consensus although the condition $\rho(\mathbf{W} - \mathbf{1}\mathbf{v}_1^T) < 1$ is not satisfied for all k, provided that the weight matrices have row-sums equal to one and that the network is connected on average. These conditions will be deeply analyzed in the study of probabilistic convergence of consensus algorithms. The existence of the limit in (3.21) is studied in different ways depending on whether the instantaneous weight matrices belong to a deterministic set or on the contrary, they conform a collection of random matrices. We review existing contributions distinguishing between both cases in the following sections.

3.3.1 Consensus in Deterministic Topologies

In this section we review some of the most relevant contributions regarding consensus in deterministic time-varying topologies, where the convergence is described in terms of the connectivity of the underlying graph model, and where new concepts of connectivity have been defined.

Olfati-Saber and Murray [OS04] introduces the concept of *switching topology* for directed networks, which refers to a case where at each time instant, the instantaneous topology belongs to a finite set of known topologies -see Chapter 2, Section 2.6.1-. In that contribution, the authors assume that the topologies in the set are all strongly connected and balanced, i.e., the instantaneous Laplacian is irreducible and the eigenvalue zero has $\mathbf{1}$ as both the left and the right associated eigenvectors[2]. According to the model in [OS04], this condition is equivalent to say that the weight matrix fulfills the conditions in (3.18) at any time instant. For the particular case of undirected networks, $\mathbf{W}(k)$ is a symmetric matrix and the conditions of strong connectivity and balanced nodes reduce to the condition of a connected network at any time instant.

Jadbabaie *et al.* [Jad03] introduces the concept of *periodical connectivity* -see Chapter 2, Section 2.6.1- to show convergence to a common value in undirected networks assuming row-stochastic and primitive weight matrices, i.e., the instantaneous weight matrices are nonnegative and due to the Perron-Frobenius theorem [Hor06] they satisfy the convergence conditions in (3.18). In that contribution, the properties of products of indecomposable, aperiodic and stochastic matrices which result from Wolfowitz theorem [Wol63] are used to show that a sufficient condition for the state vector to converge to a common value asymptotically, is that the network is periodically connected. Clearly, periodical connectivity of the network is a weaker condition than instantaneous connectivity, as assumed in [OS04]. The value reached by the nodes depends on the set of initial values and on the sequence of weight matrices.

Ren and Beard [Ren05a] generalizes the results in [Jad03] for directed networks with positive weight matrices and shows that a necessary and sufficient condition to achieve

[2] The set of all topologies has therefore at most $N(N-1)$ elements.

a consensus asymptotically assuming time-varying topologies is that the union of the collection of graphs across some time intervals has a spanning tree frequently enough. Then, the states of the nodes converge to the state corresponding to the root of the tree. The requirement of having a spanning tree frequently enough is less strict than the requirement of having a periodically connected network as in [Jad03], but guarantees that the algebraic multiplicity of the eigenvalue zero of the Laplacian matrix is one frequently enough. Recall that the Laplacian of a strongly connected digraph is irreducible, while the Laplacian of a quasi-strongly connected graph is not.

Assuming a noisy model, Xiao *et al.* [Xia05] computes the maximum-likelihood (ML) of the initial values in a network with a time-varying topology using two different *paracontracting matrices*[3] to model the link weights, and shows that both models guarantee convergence to the average consensus provided that the infinitely occurring communication graphs are jointly connected. Analogously to the assumption in [Jad03], the model allows communication links to fail all the time but, the convergence is guaranteed as long as the union of infinitely occurring communication graphs is jointly connected.

3.3.2 Consensus in Random Topologies

When the edges of a graph are added or removed unpredictably from the set at any time, the graph can be seen as the realization of a random process. WSNs are normally exposed to random communication failures caused by packet loss, range problems, mobility of the nodes, nodes being switched off, damaged or turned to stand-by mode. These communication impairments cause abrupt changes in the connectivity of the network, which are described by means of a random graph.

Consider a supergraph $\bar{\mathcal{G}}$ with a fixed number of nodes $\mathcal{V}$ and a fixed edge set $\mathcal{E}$. A realization $\mathcal{G}(k)$ at time k is a spanning subgraph of $\bar{\mathcal{G}}$ whose edge set $\mathcal{E}(k) \subseteq \mathcal{E}$ varies randomly with time. The weight matrix $\mathbf{W}(k)$ under these connectivity conditions is also random and makes the state vector $\mathbf{x}(k)$ in (3.19) become a random process. In addition to the condition imposed to the weight matrix on having row-sums equal to one, conditions

[3] A matrix is paracontracting with respect to the Euclidean norm if and only if all its eigenvalues lie in the interval $(-1, 1]$.

on the connectivity of the underlying graph model must be imposed in statistical terms, depending on the form of convergence desired.

Consider the agreement space $\mathcal{A}$ defined in (3.9) and let $\mathbf{x}_a \in \mathcal{A}$. We may have the following forms of convergence of the consensus algorithm:

- **Sure convergence -** We say that $\mathbf{x}(k)$ converges to a vector $\mathbf{x}_a$ if
$$\lim_{k\to\infty} \mathbf{x}(k) = \mathbf{x}_a$$
or equivalently, $\mathbf{x}(k) \to \mathbf{x}_a$.
- **Almost sure convergence -** We say that $\mathbf{x}(k)$ converges almost surely -or with probability one- to $\mathbf{x}_a$ if
$$\Pr\left\{\lim_{k\to\infty} \mathbf{x}(k) = \mathbf{x}_a\right\} = 1$$
or equivalently, $\mathbf{x}(k) \xrightarrow{a.s.} \mathbf{x}_a$. Almost sure convergence is also denoted as *strong convergence*.
- **Convergence in expectation -** We say that $\mathbf{x}(k)$ converges to $\mathbf{x}_a$ in expectation if
$$\lim_{k\to\infty} \mathbb{E}\left[\mathbf{x}(k)\right] = \mathbf{x}_a.$$
- **Mean square convergence -** We say that $\mathbf{x}(k)$ converges to $\mathbf{x}_a$ in the mean square sense if
$$\lim_{k\to\infty} \mathbb{E}\left[\left\|\mathbf{x}(k) - \mathbf{x}_a\right\|_2^2\right] = 0$$
where $\|\cdot\|_2$ denotes the 2-norm, or equivalently, $\mathbf{x}(k) \xrightarrow{m.s.} \mathbf{x}_a$.
- **Convergence in probability -** We say that $\mathbf{x}(k)$ converges to $\mathbf{x}_a$ in probability if for all $\delta > 0$
$$\lim_{k\to\infty} \Pr\left\{\left\|\mathbf{x}(k) - \mathbf{x}_a\right\|_2 > \delta\right\} = 0$$
or equivalently, $\mathbf{x}(k) \xrightarrow{P} \mathbf{x}_a$.

Almost sure convergence implies convergence in probability. Convergence in the mean square sense implies also convergence in probability.

Most contributions on probabilistic convergence of the consensus algorithm found in literature address the analysis considering statistically independent links existing with a given nonzero probability, and the most relevant ones are reviewed in Chapter 5. Instantaneous communication links may be however spatially correlated not only due to an intrinsic correlation of the channels between pairs of nodes, but also to the communication protocol. An example is the random gossip algorithm, which is an asynchronous version of the consensus algorithm with spatially correlated links. In random gossip algorithms, a node wakes up randomly with probability $1/N$ and either establishes a bidirectional communication link with another randomly chosen node as in pair-wise [Boy06] and geographic gossiping [Dim08], or broadcasts its state to all its neighbors within its connectivity range as in broadcast gossiping [Ays09]. Relevant contributions regarding probabilistic convergence of the consensus algorithm assuming spatially correlated communication links are reviewed in Chapter 6.

In the following section we review some approaches reported in literature to design the weight matrix such that a consensus -constrained or unconstrained- is achieved.

3.4 Design of the Weight Matrix

The design of the weight matrix is performed using information on the topology of the network and depends on the available information at every node. According to the application, the communication links may be required to be undirected or directed. Some weighting models assume that the nodes have global information available while other models can be implemented using local information gathered directly by the nodes or obtained through cooperation. The weight matrix model is also chosen depending on the connectivity of the network, i.e., whether it is time-varying or fixed. When the topology is fixed, the model is chosen to satisfy the convergence conditions reviewed in Section 3.2, either (3.15) or (3.18), whereas when the topology is time-varying the instantaneous weight matrices are forced to satisfy either one of them.

3.4.1 Review of Common Weight Matrix Designs

Max-degree weights An approach to design the matrix $\mathbf{W}$ in graphs with time-invariant topology consists in assigning a weight on each edge equal to the maximum out-degree of the network [Xia03], i.e.,

$$\mathbf{W}_{ij} = \begin{cases} 1/d^{out}_{max} & j \in \mathcal{N}_i \\ 1 - d^{out}_i/d^{out}_{max} & i = j \\ 0 & \text{otherwise} \end{cases}$$

where

$$d^{out}_{max} = \max_{i \in \mathcal{V}} d^{out}_i \leq N - 1.$$

Note that the value of d^{out}_{max} must be previously determined and broadcasted across the network before running the consensus algorithm. This is an example of a distributed scheme using global information.

Local-degree weights Another approach to design the matrix $\mathbf{W}$ consists in assigning a weight on each edge equal to the largest out-degree of its two incident nodes as follows [Xia03]

$$\mathbf{W}_{ij} = \begin{cases} 1/\max\{d^{out}_i, d^{out}_j\} & j \in \mathcal{N}_i \\ 0 & j \notin \mathcal{N}_i \\ 1 - \sum_{j \neq i} \mathbf{W}_{ij} & i = j \end{cases}.$$

For this model, each node requires knowledge of the out-degrees of all its neighboring nodes. Analogous to the previous example, this information must be known by the nodes before running the consensus algorithm and can be exchanged locally at the beginning of the algorithm. This is an example of a distributed scheme using local information.

Pair-wise/geographic gossip This algorithm assumes a transmitting node i chosen randomly with probability $1/N$. Then, the transmitting node choses another node j with a nonzero probability to establish a bidirectional communication link. The weight matrix for a transmitting node i and a receiving node j at time k in random gossiping is modeled as follows

$$\mathbf{W}(k) = \mathbf{I} - \frac{(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T}{2}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix and $\mathbf{e}_i$ denotes its i^{th} column vector [Boy06]. In pair-wise gossiping, $j \in \mathcal{N}_i$, i.e., each node communicates only with neighboring nodes, whereas in geographic gossiping, a connection is established with any $j \in \mathcal{V}$ through geographic routing. The state of the nodes in both algorithms converge asymptotically to the average consensus. The pair-wise gossip and the geographic gossip are examples of distributed consensus algorithms using global information.

Broadcast gossip In broadcast gossiping, a transmitting node q randomly chosen with probability $1/N$ broadcasts its state to all the nodes within its connectivity range. The weight matrix for the broadcast gossip algorithm has $\{ij\}^{th}$ entry given by

$$\mathbf{W}_{ij}(k) = \begin{cases} 1 & i \notin \mathcal{N}_q,\ j = i \\ \gamma & i \in \mathcal{N}_q,\ j = i \\ 1-\gamma & i \in \mathcal{N}_q,\ j = q \\ 0 & \text{otherwise} \end{cases}$$

where γ is denoted the *mixing parameter* [Ays09]. For a proper choice of the mixing parameter, the broadcast gossip algorithm reaches a consensus asymptotically but on a value different to the initial average. The broadcast gossip requires the use of global information.

Nearest-neighbor rule A well-known method for assigning weights in a graph with time-varying topology is the nearest-neighbor rule [Vic95, Jad03], where

$$\mathbf{W}_{ij}(k) = \begin{cases} a_{ij}(k) & j \in \mathcal{N}_i(k) \\ \frac{1}{1+d_i^{out}(k)} & i = j \\ 0 & \text{otherwise} \end{cases}$$

$a_{ij}(k)$ is the $\{ij\}^{th}$ entry of the instantaneous adjacency matrix, $d_i^{out}(k)$ is the time-varying out-degree of node i and $\mathcal{N}_i(k)$ is the time-varying set of neighbors. The nearest-neighbor rule does not preserve the average because $\mathbf{1}^T\mathbf{W}(k) \neq \mathbf{1}^T$, so the asymptotic agreement depends on the set of initial measurements $\mathbf{x}(0)$ and on the sequence of graph topologies. This approach is therefore not suitable to compute the average consensus.

Metropolis weights The metropolis weights for a graph with a time-varying topology are defined as follows [Xia06c]

$$\mathbf{W}_{ij}(k) = \begin{cases} \frac{1}{1+\max\{d_i^{out}(k), d_j^{out}(k)\}} & j \in \mathcal{N}_i(k) \\ 1 - \sum\limits_{l \in \mathcal{N}_i(k)} \mathbf{W}_{il}(k) & i = j \\ 0 & \text{otherwise} \end{cases}.$$

Analogous to the local-degree weights model, each node requires knowledge of the out-degrees of the neighboring nodes, but in this case the set of neighbors vary with time. An approach to share the out-degree is to broadcast it together with the state value at each iteration. Note that the diagonal elements are adjusted so that the sum of the states remains unchanged and the conditions for convergence to the average consensus are fulfilled.

Adaptive weights The adaptive weights model is designed for a graph whose links may fail at random [Den08]. First, an optimal weight matrix $\mathbf{W}^{\mathcal{G}}$ computing the average consensus is modeled for the fixed supergraph $\mathcal{G}^s$ assuming no link failures. Then, the weight for each link is decided at each time step depending on whether there is a communication failure or not as follows

$$\mathbf{W}_{ij}(k) = \begin{cases} \mathbf{W}_{ij}^{\mathcal{G}} & j \in \mathcal{N}_i(k) \\ 1 - \sum\limits_{l \in \mathcal{N}_i(k)} \mathbf{W}_{il}(k) & i = j \\ 0 & \text{otherwise} \end{cases}.$$

The diagonal elements of $\mathbf{W}(k)$ are adjusted as well so the conditions for convergence to the average consensus are fulfilled.

3.4.2 Consensus Algorithms with Uniform Weights

A widely used model for the weight matrix in both time-invariant and time-varying topologies and the one assumed throughout the analysis included in this PhD thesis, consists in weighting the difference with the states of the neighboring nodes at each iteration with a positive constant ϵ, as proposed by [OS04]. The consensus algorithm with uniform weights

ϵ has an update equation given by

$$x_i(k) = x_i(k-1) + \epsilon \sum_{j \in \mathcal{N}_i} (x_j(k-1) - x_i(k-1)).$$

The weight matrix has in this case entries given by

$$\mathbf{W}_{ij} = \begin{cases} \epsilon, & j \in \mathcal{N}_i \\ 1 - \epsilon|\mathcal{N}_i| & i = j \\ 0 & \text{otherwise} \end{cases}$$

where $|\cdot|$ denotes cardinality, or expressed in matrix form

$$\mathbf{W} = \mathbf{I} - \epsilon\mathbf{L} \tag{3.22}$$

where $\mathbf{L}$ is the Laplacian matrix of the associated underlying graph. The choice of ϵ must be properly made to guarantee that the weight matrix satisfies the convergence conditions. For the time-varying case, the Laplacian matrix is time-varying and so is the weight matrix, given by

$$\mathbf{W}(k) = \mathbf{I} - \epsilon\mathbf{L}(k). \tag{3.23}$$

In some cases, the value of the link weights ϵ is time-varying. For instance, in the presence of additive noise [Hat05] proposes a decreasing link weight in order to reduce the variance of the consensus value, and a similar approach is proposed in [Kar09]. Moreover, [Por07] studies the consensus with link weights that are not necessarily positive. In our analysis however, ϵ is assumed equal for all the links, constant and positive, although a positive ϵ not necessarily results in a positive weight matrix. The optimum choice of ϵ is described separately in the following section.

3.4.2.1 Conditions to Reach a Consensus with Uniform Weights

Consider a WSN with time-invariant topology and implementing the consensus algorithm in (3.6) with weight matrix defined in (3.22). According to the results from Section 3.2.1, the weight matrix must satisfy the convergence conditions in (3.15) to reach a consensus. The uniform weights model in (3.22) satisfies the first condition since by construction, $\mathbf{L1} = \mathbf{0}$ and therefore $\mathbf{W1} = \mathbf{1}$. Further, in order to have $\rho(\mathbf{W} - \mathbf{1}\mathbf{v}_1^T) < 1$, observe that the eigenvalues of $\mathbf{W}$ are given by

$$\lambda_i(\mathbf{W}) = 1 - \epsilon\lambda_{N-i+1}(\mathbf{L}), \quad i = 1, \cdots, N. \tag{3.24}$$

As described in Chapter 2 -Section 2.5.1- all the eigenvalues of $\mathbf{L}$ are located inside a disc of radius d_{max}^{out} centered on the real axis and tangent to the imaginary axis [Hor06], and have therefore a nonnegative real part. Therefore, in order to have $\rho(\mathbf{W}) = \lambda_1(\mathbf{W}) = 1$, the interval of possible values of ϵ must be delimited. Using (3.24), we have that the dynamic range of ϵ is given by

$$\epsilon \in (0, 2\beta)$$

with

$$\beta = \min_{\substack{i=1,\cdots,N-1 \\ \lambda_i(\mathbf{L}) \neq 0}} \frac{\mathrm{Re}[\lambda_i(\mathbf{L})]}{|\lambda_i(\mathbf{L})|^2}$$

where Re[·] denotes the real part -see Fig. 3.3-. Then, with an ϵ belonging to its dynamic range, we ensure that $\lambda_i(\mathbf{W}) = 1$ whenever $\lambda_{N-i+1}(\mathbf{L}) = 0$, i.e., the eigenvalues equal to 1 of $\mathbf{W}$ are associated with the eigenvalues equal to 0 of $\mathbf{L}$, while the remaining ones are less than 1 in magnitude. If the matrix $\mathbf{L}$ is positive semidefinite, which is the case in undirected networks, its eigenvalues are real and can be arranged in non-increasing order as follows

$$\lambda_1(\mathbf{L}) \geq \cdots \geq \lambda_N(\mathbf{L}) = 0$$

and $\beta = 1/\lambda_1(\mathbf{L})$, such that the dynamic range of ϵ is $(0, 2/\lambda_1(\mathbf{L}))$. A practical but more restrictive solution is to choose ϵ in the interval $(0, 1/N-1]$, since from spectral graph theory we have that [God01]

$$2(N-1) \geq 2d_{max}^{out} \geq \lambda_1(\mathbf{L}).$$

Remark that substituting the value of ϵ for its upper bound $1/d_{max}^{out}$ we obtain the max-degree weights model presented in Section 3.4.

The next step consists in ensuring that the eigenvalue zero of $\mathbf{L}$ has algebraic multiplicity equal to one such that $\rho(\mathbf{W} - \mathbf{1}\mathbf{v}_1^T) < 1$. Recall that for digraphs, the eigenvalue zero of the Laplacian matrix has algebraic multiplicity one if the digraph has a spanning tree, which is a particular case of quasi-strongly connected topology where only one node can reach all the other nodes of the network. Then, the weight matrix in (3.22) will have an eigenvalue 1 with algebraic multiplicity one and a consensus will be asymptotically achieved for an ϵ belonging to its dynamic range. On the other hand, if the digraph has

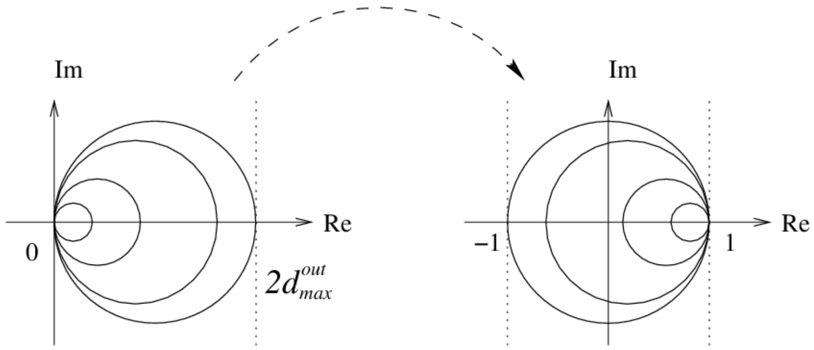


Figure 3.3: Visualization of Geršgorin circles in the complex plane for the eigenvalues of the Laplacian and the resulting ones for the weight matrix in (3.22) with $\epsilon \in (0, 2\beta)$.

no spanning tree, the algebraic multiplicity of $\lambda_N(\mathbf{L}) = 0$ is larger than one and the algebraic multiplicity of the eigenvalue $\lambda_1(\mathbf{W}) = 1$ is also larger than one for any choice of ϵ. Under these conditions a consensus of the form (3.8) can not be asymptotically achieved.

If the digraph is strongly connected, the Laplacian matrix is irreducible [Chu97] and so is the weight matrix $\mathbf{W}$. Further, if the digraph is strongly connected and balanced, $\mathbf{L}$ is irreducible and balanced with all column-sums equal to zero. The column-sums of $\mathbf{W}$ will be equal to one and the average consensus is asymptotically achieved for a proper choice of ϵ. If $\mathbf{W}$ is nonnegative, it is also a double-stochastic[4] matrix. If in addition the entries of the main diagonal of $\mathbf{W}$ are all positive, $\mathbf{W}$ is primitive and due to the Perron-Frobenius theorem $\rho(\mathbf{W}) = \lambda_1(\mathbf{W})$ with algebraic multiplicity one [Hor06]. From the results in [OS04], having a primitive double-stochastic weight matrix modeled as in (3.22) is a sufficient condition to achieve the average consensus in weighted directed networks. A non-symmetric and non-balanced $\mathbf{W}$ can reach a consensus on a value different from the initial average of $\mathbf{x}(0)$. However, Rabbat *et al.* [Rab05] shows that the average consensus can be asymptotically achieved in directed topologies provided that the value of the link weights tends to zero, but at the cost of significantly increasing the convergence time.

For undirected graphs $\mathbf{L}$ is symmetric and for undirected connected graphs $\mathbf{L}$ is in addition irreducible. The resulting $\mathbf{W}$ is also symmetric and irreducible for an ϵ belonging to the dynamic range $(0, 2/\lambda_1(\mathbf{L}))$, and under these connectivity conditions the average

[4] A nonnegative matrix is called double-stochastic if all its row-sums and column-sums are equal to 1.

consensus is asymptotically achieved. On the other hand, if the undirected graph is disconnected, the algebraic multiplicity of the eigenvalue zero of $\mathbf{L}$ is larger than one. The resulting $\mathbf{W}$ has several eigenvalues with modulus one and the state vector will converge in clusters to different consensus values corresponding to the algebraic multiplicity of the eigenvalue zero of $\mathbf{L}$. The multiplicity of the eigenvalue zero of $\mathbf{L}$ reflects the number of connected components of the graph.

In the following section, we review some strategies to achieve a faster convergence of the consensus algorithms, both for general weight matrix models and for the uniform weights model.

3.5 Minimizing the Convergence Time of Consensus Algorithms

Since iterations are power consuming, reducing the total number of iterations necessary to reach a consensus is of prime importance to reduce the energy consumption and possibly lengthen the network lifetime. The minimization of the convergence time of consensus algorithms is therefore an important topic of research. We describe some approaches assuming fixed topologies.

3.5.1 Minimizing the Second Largest Eigenvalue of the Weight Matrix

The convergence speed of $\mathbf{W}^k$ to a normalized all-ones matrix $\mathbf{J}_N$ is controlled by the second largest eigenvalue of $\mathbf{W}$, as it can be seen from (3.12), where a smaller $\rho(\mathbf{W}-\mathbf{J}_N)$ will speed up the convergence. In general, an approach to reduce the number of iterations needed to achieve a consensus consists therefore in decreasing the magnitude of $\rho(\mathbf{W}-\mathbf{J}_N)$.

[Xia03] defines the *asymptotic convergence factor* and the associated *convergence time* as follows

$$\begin{aligned} r_a &= \sup_{\mathbf{x}(0)\neq\mathbf{x}_{ave}} \lim_{k\to\infty} \left(\frac{\|\mathbf{x}(k)-\mathbf{x}_{ave}\|_2}{\|\mathbf{x}(0)-\mathbf{x}_{ave}\|_2}\right)^{1/k} \\ \tau_a &= \frac{1}{\log(1/r_a)} \end{aligned} \tag{3.25}$$

where $\mathbf{x}_{ave}$ is as defined in (3.17), whereas the *per-step convergence factor* and its associated convergence time are defined as

$$
\begin{aligned}
r_s &= \sup_{\mathbf{x}(k)\neq \mathbf{x}_{ave}} \frac{\|\mathbf{x}(k+1)-\mathbf{x}_{ave}\|_2}{\|\mathbf{x}(k)-\mathbf{x}_{ave}\|_2} \\
\tau_s &= \frac{1}{\log(1/r_s)}
\end{aligned}
\tag{3.26}
$$

Based on the per-step convergence factor in (3.25), [Xia03] proposes a method to obtain the optimum weight matrix that achieves the average consensus in time-invariant balanced topologies as the solution of a semi-definite convex programming. This weight matrix is allowed to have negative entries and is considered optimum in terms of the minimization of $\rho(\mathbf{W}-\mathbf{J}_N)$. A time-varying case is addressed in [Jak10], where the mean square convergence rate is considered as the optimization criterion to assign the optimum weights in random networks with correlated links, as we will see in Chapter 6.

3.5.2 Maximizing the Algebraic Connectivity of the Graph

A well-known result from graph theory states that the algebraic connectivity of an undirected graph, i.e., the second smallest eigenvalue of the Laplacian matrix $\lambda_{N-1}(\mathbf{L})$, reflects the degree of connectivity of the graph, meaning that a more connected graph has a larger $\lambda_{N-1}(\mathbf{L})$ [God01]. For the case of weighted digraphs, Olfati-Saber and Murray [OS04] shows that the value of $\lambda_{N-1}(\hat{\mathbf{L}})$ where $\hat{\mathbf{L}} = (\mathbf{L}+\mathbf{L}^T)/2$, is a measure of the speed of convergence for the consensus algorithm with uniform weights model, where again a larger $\lambda_{N-1}(\hat{\mathbf{L}})$ gives a faster convergence of the algorithm.

Some approaches found in literature concentrate on increasing the magnitude of the algebraic connectivity by designing the topology of the network. An example is proposed in [OS05] for undirected networks, where concepts from small-world networks are applied to design the best topology maximizing $\lambda_{N-1}(\mathbf{L})$. A similar concept is used also in [Ald05] to design both the topology and the weights to be assigned to each link, whereas [Kar06] focuses on minimizing the ratio $\lambda_1(\mathbf{L})/\lambda_{N-1}(\mathbf{L})$ to achieve a faster convergence to a consensus.

3.5.3 Choosing the Optimum Uniform Link Weights

The uniform weights model described in Section 3.4.2 assumes fixed topologies, i.e., the algebraic connectivity of the network is constant through time. The value of ϵ in (3.22) can be chosen such that the second largest eigenvalue of $\mathbf{W}$ is minimized, while still satisfying the convergence conditions. As previously stated, according to (3.24), minimizing $\rho(\mathbf{W}-\mathbf{J}_N)$ is equivalent to maximizing the second smallest eigenvalue of the graph Laplacian $\lambda_{N-1}(\mathbf{L})$.

In case all the eigenvalues of $\mathbf{W}$ are real, the second largest eigenvalue in magnitude will be the maximum between $\{\lambda_2(\mathbf{W}), -\lambda_N(\mathbf{W})\}$, or expressed in terms of the eigenvalues of $\mathbf{L}$,

$$\rho(\mathbf{W}-\mathbf{J}_N) = \max\{1-\epsilon\lambda_{N-1}(\mathbf{L}), \epsilon\lambda_1(\mathbf{L})-1\}, \quad \epsilon \in \left(0, \frac{2}{\lambda_1(\mathbf{L})}\right).$$

Therefore, the optimum link weight ϵ^* minimizing $\rho(\mathbf{W}-\mathbf{J}_N)$ will be given by

$$\epsilon^* = \frac{2}{\lambda_{N-1}(\mathbf{L}) + \lambda_1(\mathbf{L})}$$

which is attained at the intersection of two lines [Xia03,Sch04]. Remark however that any value of $\epsilon \in (0, 1/(N-1)]$ will satisfy the conditions for convergence. This is based on the fact that $d^{out}_{max} \leq (N-1)$ and[5] $\lambda_1(\mathbf{L}) \leq 2d^{out}_{max}$. Thus, only knowledge the total number of nodes would be sufficient to choose an ϵ guaranteeing convergence of $\mathbf{x}(k)$ to a consensus, although the convergence time would not be minimized.

For networks with random topologies, the dynamic range of ϵ and its optimum value will be determined for given conditions on the general parameters in the forthcoming chapters. First, in Chapter 4 we will consider the convergence of the time-invariant model in (3.6) assuming quantization noise. Then, in Chapter 5 we will consider the time-varying model in (3.19) for random topologies with instantaneous directed links and study the convergence in the mean square sense. Finally, in Chapter 6 we will study almost sure convergence of the time-varying model in (3.19) assuming random topologies and communication links which are allowed to exhibit spatial correlation.

[5] In practice, in a random geometric network where communication takes place only among neighboring nodes $d^{out}_{max} \ll N-1$.

3.6 Conclusions of the Chapter

This chapter has introduced the network model for the discrete-time consensus algorithm converging on the state. The time-invariant model for both directed and undirected topologies has been presented, where the conditions to reach a consensus and the conditions to reach the average consensus have been reviewed, not only in algebraic terms but also in terms of the connectivity of the underlying graph model. The time-varying consensus model has been presented, distinguishing between deterministic time-varying and random time-varying topologies, as well as different forms of probabilistic convergence for the state vector. A variety of designs for the weight matrix in time-varying and time-invariant topologies have been described, assuming both directed and undirected communication links, and using either global or local information. The uniform weights model has been introduced along with its conditions for convergence. Finally, some approaches to reduce the convergence time of consensus algorithms in fixed topologies have been discussed.

4

Consensus with Quantized Information Exchange

4.1 Introduction

The communications in WSNs are not only exposed to power constraints to reduce the energy consumption, but also to bandwidth constraints. Therefore, in order to reduce the requirements of bandwidth, the information is usually *quantized* before transmission.

Quantization is a nonlinear process applied to physical quantities to represent them numerically. A quantizer is typically a staircase function ψ_q that transforms a continuous set of values into a discrete set with a finite number of units by rounding them to the nearest unit or quantization level, such that

$$x^q = \psi_q[x] \tag{4.1}$$

represents the quantized version of an input signal x. The intervals between consecutive quantization levels determine the accuracy of the representation, where narrower intervals result in a better approximation to the original signal. If the quantization levels are uniformly spaced, the quantizer is denoted *uniform*. Conversely, if the quantization levels are not uniformly spaced, the quantizer is denoted *non-uniform*. A graphical representation of a uniform quantizer is depicted in Fig. 4.1, where the curve in solid line depicts a symmetric version of the quantizer, i.e., having the same number of levels at both sides of the ordinate axis, whereas the dotted line depicts the non-symmetric version of the same quantizer.

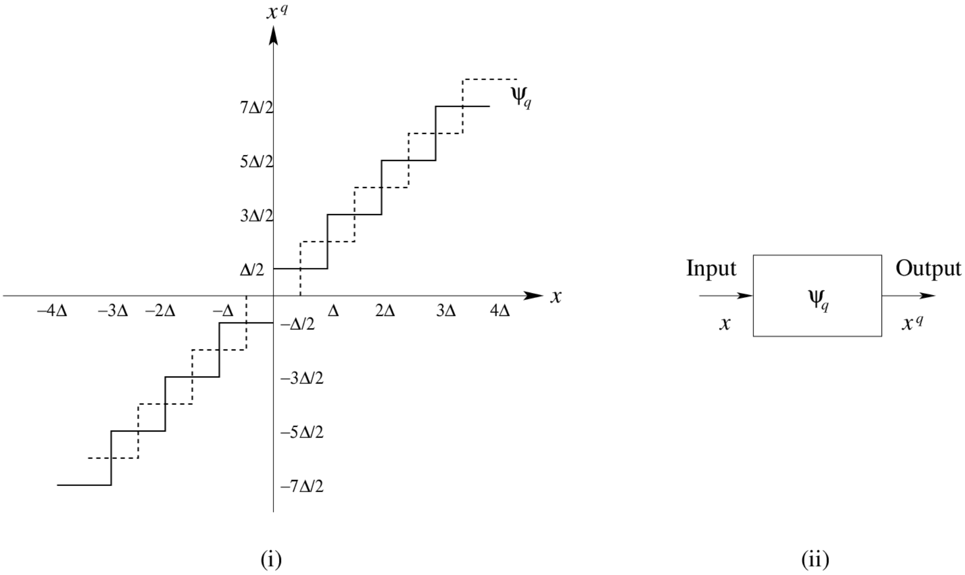


Figure 4.1: (i) Uniform symmetric quantizer (solid line) and uniform non-symmetric quantizer (dotted line) with step-size Δ and (ii) block-diagram of the quantizer ψ_q.

The use of a quantizer introduces an error determined by the difference between the input signal x and the output signal x^q, and called *quantization noise*. In the previous chapter, no assumptions on quantization are made for the consensus algorithm described. However, in a digital implementation of the algorithm, a more realistic assumption is that the computations are carried out using floating-point numbers with double precision -i.e., 64 bits precision-, while the information transmitted is quantized to reduce bandwidth and power consumption. Therefore, in this chapter each node of the network is assumed to update, encode and broadcast its state in quantized form into a packet at each iteration of the consensus algorithm.

Outline of the Chapter

This chapter is devoted to study the average consensus algorithm where at each iteration the information exchanged among neighboring nodes is quantized. A detailed overview of the most relevant contributions found in literature is provided in Section 4.2 and the quantization noise model assumed in the chapter is presented in Section 4.3. In Section 4.4 a simple transmission scheme combining data with floating-point precision and data with quantized precision is proposed, and which results from a modification of the well-known discrete-time consensus model by Olfati-Saber and Murray [OS04]. The effect of quantization noise on the performance of the algorithm is evaluated in Section 4.5 by analyzing the mean square error (MSE) of the state computed with respect to the average of the initial measurements. First, in Section 4.5.1 we assume that the quantization noise is temporally uncorrelated, a reasonable assumption at the beginning of the consensus algorithm. Then, in Section 4.5.2 we consider temporally correlated quantization noise. The performance of the proposed model is compared with the performance of two different algorithms found in literature including quantization noise. We will see that, conversely to these examples, the MSE of the state for the proposed model converges as time evolves and its limit admits a closed-form expression which can be computed offline. An upper bound for the limit of the MSE which depends on general network parameters is also derived, providing an a priori quantitative measure of the effects of quantization on the consensus algorithm. The numerical results and the conclusions of the chapter are included in Section 4.6 and Section 4.7 respectively.

4.2 Relevant Contributions on Consensus with Quantized Data

The concept of *quantized consensus* is introduced by Kashyap *et al.* [Kas06], where a random gossip algorithm denoted the *quantized gossip algorithm* is proposed. The quantized gossip algorithm restricts the state values of the network to be integers at each time step, and is said to converge to the quantized consensus if the state vector belongs to the set

$$\mathcal{S} \triangleq \left\{ \mathbf{z} : \{z_i\}_{i=1}^{N} \in \{L, L+1\}, S = \sum_{i=1}^{N} z_i(0) \right\}$$

where L is an integer such that $S = NL + R$ with $0 \leq R < N$. In other words, $\mathbf{z} \in \mathcal{S}$ has $N - R$ entries equal L and R entries equal $L+1$. The quantized gossip algorithm has the following iterative state update

$$\mathbf{z}(k) = \psi_q[\mathbf{W}\mathbf{z}(k-1)]$$

where $\mathbf{z}(k)$ is the vector of integer states at time k, ψ_q is the quantization function and $\mathbf{W}$ denotes the weight matrix satisfying the convergence conditions. The theory of Markov chains is applied to show that the state vector $\mathbf{z}(k)$ converges in probability to the set $\mathcal{S}$ as $k \to \infty$[1]. A similar problem is addressed in [Fra08] which shows that, as long as the distance from the consensus is much larger than the quantization step, the speed of convergence is almost the same as for the consensus algorithm with floating-point precision.

A straightforward approach to evaluate the impact of quantization consists in modeling the quantization noise as an additive noise. Based on the average consensus model in [OS04], Xiao *et al.* [Xia06b] studies the convergence to a consensus when the received values at each iteration are assumed corrupted with an additive noise. Then, the update equation can be expressed as

$$x_i(k) = \mathbf{W}_{ii} x_i(k-1) + \sum_{j=1, j\neq i}^{N} \mathbf{W}_{ij} x_j(k-1) + v_i(k-1)$$

where v_i, $i = 1, \cdots, N$ are independent identically distributed (i.i.d) random variables (r.v.'s) with zero mean and unit variance. Expressing the update equation for the whole network in matrix form yields

$$\mathbf{x}(k) = \mathbf{W}\mathbf{x}(k-1) + \mathbf{v}(k-1).$$

[Xia06b] shows that the variance of $\mathbf{x}(k)$ with respect to the vector of initial averages given by

$$\mathbf{x}_{ave} = \mathbf{J}_N \mathbf{x}(0) \tag{4.2}$$

where $\mathbf{J}_N$ is the normalized all-ones matrix, diverges with time. According to these results, if quantization noise is modeled as an additive noise with i.i.d. components the algorithm may fail to converge. On the other hand, Yildiz and Scaglione [Yil07] modifies the model of [OS04] to include source coding strategies and quantization noise, and shows however that when exploiting the increasing temporal and spatial correlation of the state values, the variance of the state does not increase unbounded as time evolves. The main conclusion of the paper is that, even using highly suboptimal coding strategies, when the correlation of the data is taken into account, the MSE evaluated with respect to $\mathbf{x}_{ave}$ does not increase with the number of iterations.

[1] The quantized consensus as defined by [Kas06] is therefore not a strict consensus, and is different from the model of consensus with quantized information assumed in this chapter.

Aysal *et al.* [Ays07a] presents an algorithm where the nodes implement a probabilistic quantizer such that the quantized state at each iteration is equal to the analog state in expectation. The update equation for this model is given by

$$x_i = \mathbf{W}_{ii} x_i^q(k-1) + \sum_{j=1, j\neq i}^{N} \mathbf{W}_{ij} x_j^q(k-1) \tag{4.3}$$

where x_i^q is the quantized version of x_i, as defined in (4.1)[2]. Using the theory of Markov chains, almost sure convergence to a consensus in one of the quantized levels is shown, where the expected value of the consensus value is equal to the average of the analog measurements. The rates of convergence for the average consensus algorithm using probabilistic quantization is addressed in [Ays07b]. Furthermore, Aldosari and Moura [Ald06] proposes a model of consensus algorithms for distributed detection in sensor networks based on the time-invariant model by [OS04] and studies the impact of quantization on the performance. For this model, each node updates the state using its own value with infinite precision and the values of its neighbors with quantized precision as follows

$$x_i(k) = \mathbf{W}_{ii} x_i(k-1) + \sum_{j=1, j\neq i}^{N} \mathbf{W}_{ij} x_j^q(k-1). \tag{4.4}$$

The paper shows through simulations that the performance of the dynamical system in (4.4), evaluated in terms of probability of decision error, is degraded by the quantization noise but can be improved introducing a small degree of randomness in the topology of the network, affecting thus the structure of the matrix $\mathbf{W}$. A trade-off between the total number of communication links of the network and the amount of bits used for quantization is established, where a higher number of connections leads to a faster convergence, while a higher number of quantization bits leads to a smaller probability of decision error. Although the state of the nodes for the model in [Ald06] converge to a steady-state value, the MSE of the state does not admit a closed-form expression, as we will see in Section 4.5. In this chapter we propose a model including quantization noise which admits a closed-form expression for the residual MSE of the state, providing an a priori quantitative measure of the effects of quantization on the consensus algorithm.

Other contributions not related to the work carried out in this chapter which are

[2] Remark that substituting $x_i^q(k-1)$ with $x_i(k-1)$ we obtain the time-invariant consensus model presented in Chapter 3.

also worth mentioning, include for instance Kar and Moura [Kar08a], which applies the theory of controlled Markov processes to show almost sure convergence to the average consensus when the data is randomized by adding an amount of statistical dither before quantization and assuming a sequence of link weights decreasing in time. As in [Ald06], the MSE is characterized with respect to $\mathbf{x}_{ave}$ and the trade-off between convergence rate and accuracy of the estimates is discussed. [Kar10] extends the results of [Kar08a] focusing on quantized consensus with random link failures, shows almost sure and mean square converge assuming an unbounded quantizer range, and derives probability bounds assuming a bounded quantizer range. Finally, Schizas *et al.* [Sch08] proposes distributed MLE and BLUE estimators for the estimation of deterministic signals in *ad hoc* WSNs. These estimators, formulated as the solution of convex minimization subproblems, exhibit resilience against quantization noise.

4.3 Quantization Noise Model

In this section we present the quantization noise model assumed throughout the analysis carried out in the chapter. Let's consider a uniform quantizer with b bits, that is, with a total of 2^b quantization levels. Assuming a range $[-X_{max}, X_{max}]$, the step-size of the quantizer is equal to

$$\Delta = \frac{2X_{max}}{2^b}.$$

Recall that using a higher number of bits for a given range, we obtain a smaller step-size Δ and a better approximation to the analog input signal and a smaller quantization error, but also the requirements of storage and data processing are more demanding. The trade-off between accuracy of the representation and bandwidth requirement is therefore established by the size of the step Δ. Since roundoff errors have values between plus and minus $\Delta/2$, we will consider that the quantization noise is modeled as a r.v. uniformly distributed within the interval $(-\Delta/2, \Delta/2]$. The variance of the quantization noise for this model is thus given by

$$\sigma_q^2 = \frac{X_{max}^2}{3} 2^{-2b}.$$

The quantization noise is modeled as an additive noise such that the vector of quantized values can be expressed as follows

$$\mathbf{x}^q(k) = \mathbf{x}(k) + \mathbf{e}_q(k) \tag{4.5}$$

where $\mathbf{e}_q(k) \in \mathbb{R}^{N\times 1}$ is the error vector at time k, and $\{\mathbf{e}_q(k), \forall k \geq 0\}$ conforms a set of independent uniformly distributed zero mean random vectors.

According to the mathematical formulation in (4.5), the state values should diverge for some models of the consensus algorithm, as we will see in Section 4.5, where the MSE of the state is studied. This is not the case in practice and we will corroborate that the states converge in the simulations section. The reason is that the error vectors for the consensus algorithm are neither uniformly distributed nor spatially independent, although both assumptions are made for mathematical simplicity. We propose however a model of consensus including quantization noise that admits a closed-form expression for the MSE of the state which converges as time evolves, presented in the following section.

4.4 Consensus Algorithms with Quantized Data

In this section we present a consensus algorithm based on the discrete-time model of [OS04], given by

$$\mathbf{x}(k) = \mathbf{W}\mathbf{x}(k-1), \quad \forall k > 0 \tag{4.6}$$

where the weight matrix is modeled using the uniform weights model

$$\mathbf{W} = \mathbf{I} - \epsilon\mathbf{L} \tag{4.7}$$

with constant $\epsilon > 0$ and Laplacian matrix $\mathbf{L}$ associated with the underlying graph. The eigenvalues of $\mathbf{W}$ and $\mathbf{L}$ are related as follows

$$\lambda_i(\mathbf{W}) = 1 - \epsilon\lambda_{N-i+1}(\mathbf{L}), \quad \forall i = 1, \cdots, N. \tag{4.8}$$

The Laplacian matrix is assumed time-invariant, symmetric and irreducible, such that for $\epsilon \in (0, 2/\lambda_1(\mathbf{L}))$ the weight matrix satisfies the following convergence conditions

$$\begin{aligned} \mathbf{1}^T\mathbf{W} &= \mathbf{1}^T \\ \mathbf{W}\mathbf{1} &= \mathbf{1} \\ \rho(\mathbf{W} - \mathbf{J}_N) &< 1 \end{aligned} \tag{4.9}$$

i.e., it has largest eigenvalue $\lambda_1(\mathbf{W}) = 1$ with algebraic multiplicity one and associated left and right eigenvector $\mathbf{1} \in \mathbb{R}^{N\times 1}$, an all-ones vector of length N.

We assume that the nodes compute their state at each iteration using infinite -floating-point- precision values but are only allowed to broadcast a quantized version of the updated state. Then, at time k a node i updates its state using the quantized states received from its neighbors combined with its own previous state as follows

$$x_i(k) = x_i(k-1) + \sum_{j\in\mathcal{N}_i}^{N} \mathbf{W}_{ij} \left(x_j^q(k-1) - x_i^q(k-1)\right)$$

i.e., its own value is used both with floating-point precision and quantized. Rewriting the expression above for the whole network in matrix form yields

$$\mathbf{x}(k) = \mathbf{x}(k-1) - \epsilon\mathbf{L}\mathbf{x}^q(k-1). \tag{4.10}$$

Assuming the quantization noise model given in Section 4.3, we can substitute (4.5) in (4.10) to obtain

$$\mathbf{x}(k) = \mathbf{W}\mathbf{x}(k-1) - \epsilon\mathbf{L}\mathbf{e}_q(k-1). \tag{4.11}$$

The performance of the algorithm in (4.11) is compared with the performance of two different models including quantization. The first model assumes that each node updates the state using its own previous value and the values received from neighboring nodes in quantized form as in (4.3), with the quantization noise model presented in Section 4.3. The difference equation for this model is given by

$$\mathbf{x}(k) = \mathbf{W}\mathbf{x}^q(k-1). \tag{4.12}$$

After the state update at each node, the state is quantized and broadcasted. The second model used for comparison is the one proposed in [Ald06], which assumes that the i^{th} node updates the state using the quantized state values received from its neighbors and its own state value with floating-point precision as in (4.4). The resulting dynamical equation in matrix notation is given by

$$\mathbf{x}(k) = \mathbf{W}_D\mathbf{x}(k-1) + (\mathbf{W}-\mathbf{W}_D)\mathbf{x}^q(k-1) \tag{4.13}$$

where $\mathbf{W}_D$ is a diagonal matrix with $\{ii\}^{th}$ entry equal to $[\mathbf{W}_D]_{ii} = \mathbf{W}_{ii}$.

In the following section we analyze the MSE of the state vector in (4.10) using the expression in (4.11), where we show that, conversely to the case of (4.12) and (4.13), the formulation in (4.11) admits a closed-form expression for the MSE of the state which converges as time evolves.

4.5 MSE Analysis of Consensus Algorithms with Quantized Data

In this section we analyze the MSE of the state vector in (4.10) using the expression in (4.11). Due to the randomness of the quantization noise model in (4.5), the convergence of $\mathbf{x}(k)$ is studied in probabilistic terms. We analyze the MSE of the state, defined as

$$\text{MSE}(x(k)) = \mathbb{E}\left[\|\mathbf{x}(k) - \mathbf{x}_{ave}\|_2^2\right] \tag{4.14}$$

where $\mathbf{x}_{ave}$ is the average consensus vector defined in (4.2). The MSE of the state can be decomposed into the sum of the variance and the squared bias as follows

$$\begin{aligned} \text{MSE}(x(k)) &= \mathbb{E}\left[\|\mathbf{x}(k) - \mathbb{E}[\mathbf{x}(k)]\|_2^2\right] + \|\mathbb{E}[\mathbf{x}(k)] - \mathbf{x}_{ave}]\|_2^2 \\ &= \text{var}(\mathbf{x}(k)) + \text{bias}^2(\mathbf{x}(k)) \end{aligned} \tag{4.15}$$

so that the terms can be analyzed separately. The theoretical analysis will be carried out under two different assumptions. First, we consider the case of spatially and temporally uncorrelated quantization noise in Section 4.5.1. Then, we consider the case of spatially uncorrelated but temporally correlated quantization noise in Section 4.5.2.

4.5.1 MSE with Temporally Uncorrelated Quantization Noise

We analyze the bias and the variance of the state vector $\mathbf{x}(k)$ in (4.11) when the quantization error is assumed uncorrelated among nodes and uncorrelated from one time instant to the next one. Under these conditions, the covariance matrix of the error vector can be expressed as

$$\mathbb{E}[\mathbf{e}_q(l)\mathbf{e}_q^T(m)] = \sigma_q^2 \delta_{lm} \mathbf{I} \tag{4.16}$$

with δ_{lm} denoting the Kronecker delta function. Note that the assumption of temporally uncorrelated quantization noise is reasonable at the beginning of the consensus algorithm. The MSE in (4.14) converges as time evolves, as shown by the following proposition.

Proposition 4.1. *Consider the algorithm in (4.11) with weight matrix* $\mathbf{W}$ *defined in (4.7) and satisfying (4.9), and i.i.d. zero mean temporally uncorrelated random vectors* $\{\mathbf{e}_q(k), \forall k > 0\}$ *assumed independent of* $\mathbf{x}(0)$*. The limit of the MSE of the state in (4.14) as* k *increases is given by*

$$\lim_{k\to\infty} MSE(x(k)) = \sigma_q^2 \sum_{i=2}^{N} \frac{1-\lambda_i(\mathbf{W})}{1+\lambda_i(\mathbf{W})} \tag{4.17}$$

where $\{\lambda_i(\mathbf{W}), i = 1, \cdots, N\}$ *is the set of eigenvalues of* $\mathbf{W}$ *arranged in non-increasing order and* σ_q^2 *is the variance of the quantization error.*

Proof. Considering the quantized vector in (4.5), the dynamical system in (4.11) can be rewritten as

$$\mathbf{x}(k) = \mathbf{W}^k \mathbf{x}(0) - \epsilon \sum_{l=1}^{k} \mathbf{W}^{l-1} \mathbf{L} \mathbf{e}_q(k-l). \tag{4.18}$$

We start the proof analyzing the bias term defined in (4.15). Since the error vectors $\mathbf{e}_q(k)$ are zero mean for all k, the expected value of the expression above is

$$\mathbb{E}[\mathbf{x}(k)] = \mathbf{W}^k \mathbf{x}(0). \tag{4.19}$$

Further, since $\mathbf{W}$ satisfies the convergence conditions in (4.9) we have

$$\lim_{k\to\infty} \mathbf{W}^k = \mathbf{J}_N$$

such that

$$\lim_{k\to\infty} \mathbb{E}[\mathbf{x}(k)] = \mathbf{J}_N \mathbf{x}(0) = \mathbf{x}_{ave}$$

and substituting for $\mathbb{E}[\mathbf{x}(k)]$ in the bias term in (4.15) we obtain

$$\lim_{k\to\infty} \left\| \mathbb{E}\left[\mathbf{x}(k)\right] - \mathbf{x}_{ave} \right\|_2^2 = 0 \tag{4.20}$$

showing that the bias is asymptotically zero.

The next step consists in analyzing the variance term defined in (4.15), which can be expressed as follows

$$\mathrm{var}(\mathbf{x}(k)) = \mathbb{E}\left[\left\| \epsilon \sum_{l=1}^{k} \mathbf{W}^{l-1} \mathbf{L} \mathbf{e}_q(k-l) \right\|_2^2 \right]$$

where we have substituted for (4.19). Then, considering that **L** and **W** are symmetric we have

$$\begin{aligned}\mathrm{var}(\mathbf{x}(k)) &= \epsilon^2 \mathbb{E}\left[\sum_{l=1}^{k} \mathbf{e}_q(k-l)^T \mathbf{L}\mathbf{W}^{2(l-1)}\mathbf{L}\mathbf{e}_q(k-l)\right] \\ &= \epsilon^2 \mathbb{E}\left[\mathrm{tr}\left(\sum_{l=1}^{k} \mathbf{W}^{2(l-1)}\mathbf{L}\mathbf{e}_q(k-l)\mathbf{e}_q(k-l)^T\mathbf{L}\right)\right]\end{aligned}$$

where tr(.) denotes the trace function and in the last equality we have applied the linearity of the trace and the expected value operators. Using the covariance matrix of the error in (4.16) yields

$$\mathrm{var}(\mathbf{x}(k)) = \sigma_q^2 \cdot \mathrm{tr}\left(\mathbf{W}^{-2}\sum_{l=1}^{k}\mathbf{W}^{2l}\epsilon^2\mathbf{L}^2\right). \tag{4.21}$$

Although at first sight it could seem that the variance in (4.21) does not converge due to the summation in the trace, we will show that indeed it does. Considering that **W** and **L** are diagonalized by the same set of orthonormal eigenvectors we have $\mathbf{W} = \mathbf{Q}\mathbf{\Lambda}_{\mathbf{W}}\mathbf{Q}^T$ and $\mathbf{L} = \mathbf{Q}\mathbf{\Lambda}_{\mathbf{L}}\mathbf{Q}^T$, where **Q** is an orthogonal matrix and $\mathbf{\Lambda}_{\mathbf{W}}, \mathbf{\Lambda}_{\mathbf{L}}$ are diagonal matrices with the eigenvalues of **W** and **L** respectively, denoted by $\{\lambda_i(\mathbf{W}), \lambda_i(\mathbf{L}), \forall i = 1, \cdots, N\}$. Then, remark that

$$\mathbf{W}^{-2}\sum_{l=1}^{k}\mathbf{W}^{2l}\epsilon^2\mathbf{L}^2 = \mathbf{Q}\mathbf{\Lambda}_{\mathbf{W}}^{-2}\mathbf{M}_k\epsilon^2\mathbf{\Lambda}_{\mathbf{L}}^2\mathbf{Q}^T \tag{4.22}$$

where $\mathbf{M}_k$ is an $N \times N$ diagonal matrix with entries given by

$$[\mathbf{M}_k]_{11} = k \tag{4.23a}$$

$$[\mathbf{M}_k]_{ii} = \frac{\lambda_i(\mathbf{W})^2 - \lambda_i(\mathbf{W})^{2k+2}}{1-\lambda_i(\mathbf{W})^2} \quad i = 2, \cdots, N \tag{4.23b}$$

where we have considered that $\lambda_1(\mathbf{W}) = 1$ and used the equality

$$\sum_{l=0}^{k}\lambda_i(\mathbf{W})^l = \frac{1-\lambda_i(\mathbf{W})^{k+1}}{1-\lambda_i(\mathbf{W})}. \tag{4.24}$$

Recalling that **Q** is orthogonal and using that $\lambda_N(\mathbf{L}) = 0$, the trace of the matrix in (4.22) is given by

$$\begin{aligned}\mathrm{tr}(\mathbf{W}^{-2}\sum_{l=1}^{k}\mathbf{W}^{2l}\epsilon^2\mathbf{L}^2) &= \sum_{i=2}^{N}\lambda_i(\mathbf{W})^{-2}[\mathbf{M}_k]_{ii}\epsilon^2\lambda_{N-i+1}(\mathbf{L})^2 \\ &= \sum_{i=2}^{N}\frac{1-\lambda_i(\mathbf{W})}{1+\lambda_i(\mathbf{W})}(1-\lambda_i(\mathbf{W})^{2k})\end{aligned} \tag{4.25}$$

where in the last equality we have substituted for (4.23) and used the eigenvalue relationship in (4.8). Substituting (4.25) in (4.21) and computing the limit as k increases yields

$$\lim_{k\to\infty} \mathrm{var}(\mathbf{x}(k)) = \sigma_q^2 \sum_{i=2}^{N} \frac{1-\lambda_i(\mathbf{W})}{1+\lambda_i(\mathbf{W})} \tag{4.26}$$

where we have considered that $\lambda_i(\mathbf{W}) < 1$, for all $i = 2, \cdots, N$. Finally, combining the results in (4.20) and (4.26) we obtain the expression for the limit of MSE of the state in (4.17), completing the proof.

□

Furthermore, considering the eigenvalue relationship in (4.8) and recalling that $\lambda_1(\mathbf{L}) \leq N$ [God01], we can obtain an upper bound for the limit of the MSE($\mathbf{x}(k)$) in (4.17), as stated in the following corollary.

Corollary 4.2. *The limit of the MSE of the state in (4.17) is upper bounded by*

$$\lim_{k\to\infty} \mathit{MSE}(x(k)) \leq (N-1)\sigma_q^2 \tag{4.27}$$

where σ_q^2 is the variance of the quantization error and N is the total number of nodes.

Proposition 4.1 shows that the limit MSE($\mathbf{x}(k)$) as k tends to infinity exists and it can be computed offline, whereas Corollary 4.2 provides an upper bound for the limit of the MSE($\mathbf{x}(k)$) that depends only on general parameters of the network. These parameters are the number of nodes of the network and the variance of the quantization noise.

For the sake of comparison, we will now analyze the MSE($\mathbf{x}(k)$) for the model in (4.12). Considering the model for the quantized state vector in (4.5), the dynamical system in (4.12) is equivalent to

$$\mathbf{x}(k) = \mathbf{W}\mathbf{x}(k-1) + \mathbf{W}\mathbf{e}_q(k-1) \tag{4.28}$$

$$= \mathbf{W}^k\mathbf{x}(0) + \sum_{l=1}^{k} \mathbf{W}^l\mathbf{e}_q(k-l). \tag{4.29}$$

Analogously to the previous case, we analyze the bias an the variance terms separately. The computation of the bias is similar to the one in (4.20), since the expected value of

$\mathbf{x}(k)$ in (4.28) is given by (4.19). We conclude therefore that the bias of $\mathbf{x}(k)$ for this model is asymptotically equal to zero. For the computation of the variance, using (4.19) and (4.29) and recalling that $\mathbf{e}_q(k)$ is zero mean and independent of $\mathbf{x}(0)$ for all k we obtain

$$\text{var}(\mathbf{x}(k)) = \mathbb{E}\left[\|\sum_{l=1}^{k}\mathbf{W}^l\mathbf{e}_q(k-l)\|_2^2\right].$$

Then, we have

$$\begin{aligned}\text{var}(\mathbf{x}(k)) &= \mathbb{E}\left[\text{tr}\left(\sum_{l=1}^{k}\mathbf{W}^{2l}\mathbf{e}_q(k-l)\mathbf{e}_q(k-l)^T\right)\right]\\ &= \sigma_q^2\cdot\text{tr}\left(\sum_{l=1}^{k}\mathbf{W}^{2l}\right) \qquad (4.30)\end{aligned}$$

where we have used (4.16). Note that the trace in (4.30) diverges for $k\to\infty$, and this can be seen observing that the sum can be expressed as a geometric series of matrices as follows

$$\sum_{m=0}^{k}\mathbf{W}_s^m - \mathbf{I}, \quad \text{where } \mathbf{W}_s = \mathbf{W}^2.$$

The geometric series above converges if and only if the eigenvalues of the matrix $\mathbf{W}_s$ are less than one in magnitude. Since we know that $\lambda_1(\mathbf{W}) = 1$, $\mathbf{W}_s$ has largest eigenvalue $\lambda_1(\mathbf{W}_s) = 1$ and the geometric series does not converge. An alternative way of showing that the trace in (4.30) diverges is using the eigenvalue decomposition of $\mathbf{W}$ as in the proof of Proposition 4.1, where

$$\begin{aligned}\text{tr}\left(\sum_{l=1}^{k}\mathbf{W}^{2l}\right) &= \text{tr}\left(\sum_{l=1}^{k}\mathbf{Q}\mathbf{\Lambda}_{\mathbf{W}}^{2l}\mathbf{Q}^T\right)\\ &= \text{tr}\left(\sum_{l=1}^{k}\mathbf{\Lambda}_{\mathbf{W}}^{2l}\right)\\ &= \sum_{i=1}^{N}[\mathbf{M}_k]_{ii} \qquad (4.31)\end{aligned}$$

with $\mathbf{M}_k$ defined in (4.23). Substituting (4.23) in (4.31) and replacing further in (4.30), yields

$$\text{var}(\mathbf{x}(k)) = \sigma_q^2\left(k + \sum_{i=2}^{N}\frac{\lambda_i(\mathbf{W})^2 - \lambda_i(\mathbf{W})^{2k+2}}{1-\lambda_i(\mathbf{W})^2}\right).$$

Finally, taking the limit of the expression above yields

$$\lim_{k\to\infty} \mathrm{var}(\mathbf{x}(k)) = \sigma_q^2 \lim_{k\to\infty} \left(k + \sum_{i=2}^{N} \frac{\lambda_i(\mathbf{W})^2 - \lambda_i(\mathbf{W})^{2k+2}}{1-\lambda_i(\mathbf{W})^2} \right) = \infty \tag{4.32}$$

which shows that the variance diverges as $k \to \infty$. Clearly, the variance of the quantization error σ_q^2 forces the variance of the state to increase with time because the model in (4.29) accumulates the error as time evolves. This behavior is already well-known and would imply to stop the dynamical system run by the nodes once a consensus is achieved since, otherwise, the states would start to diverge. In practice however, the consensus algorithm with quantized state values converges, as we will see in the simulations section.

We will now analyze the MSE($\mathbf{x}(k)$) for the model in (4.13). Considering again the quantized state vector in (4.5), the dynamical system in (4.13) is equivalent to

$$\begin{aligned} \mathbf{x}(k) &= \mathbf{W}\mathbf{x}(k-1) + (\mathbf{W}-\mathbf{W}_D)\mathbf{e}_q(k-1) \\ &= \mathbf{W}^k\mathbf{x}(0) + \sum_{l=1}^{k} \mathbf{W}^{l-1}(\mathbf{W}-\mathbf{W}_D)\mathbf{e}_q(k-l). \end{aligned} \tag{4.33}$$

Analogously to the previous case, the bias for this model converges to zero as $k \to \infty$, since the expected value of $\mathbf{x}(k)$ is given by (4.19). Regarding the variance term we have

$$\begin{aligned} \mathrm{var}(\mathbf{x}(k)) &= \mathbb{E}\left[\Big\| \sum_{l=1}^{k} \mathbf{W}^l(\mathbf{W}-\mathbf{W}_D)\mathbf{e}_q(k-l) \Big\|_2^2 \right] \\ &= \mathbb{E}\left[\mathrm{tr}\left(\sum_{l=1}^{k} \mathbf{W}^{2(l-1)}(\mathbf{W}-\mathbf{W}_D)\mathbf{e}_q(k-l)\mathbf{e}_q(k-l)^T(\mathbf{W}-\mathbf{W}_D) \right) \right]. \end{aligned}$$

Using (4.16) and taking the limit of the equation above yields

$$\lim_{k\to\infty} \mathrm{var}(\mathbf{x}(k)) = \sigma_q^2 \lim_{k\to\infty} \mathrm{tr} \left(\mathbf{W}^{-2} \sum_{l=1}^{k} \mathbf{W}^{2l}(\mathbf{W}-\mathbf{W}_D)^2 \right). \tag{4.34}$$

Although we have not found a closed-form expression for this limit, with computer simulations we have observed that this expression diverges, which similarly to the previous case, is due to the summation term in the trace.

4.5.2 MSE with Temporally Correlated Quantization Noise

In this section we analyze the bias and the variance of $\mathbf{x}(k)$ in (4.11) when the quantization error is assumed uncorrelated among nodes but is exactly the same from one time instant to the next one, once the state is stabilized. This is a reasonable assumption after the consensus algorithm has reached a steady-state value. The covariance matrix of the error vector in this case is given by

$$\mathbb{E}[\mathbf{e}_q(l)\mathbf{e}_q^T(m)] = \sigma_q^2\mathbf{I}. \tag{4.35}$$

The MSE defined in (4.14) converges as time evolves in this case as well, as shown by the following proposition.

Proposition 4.3. *Consider the algorithm in (4.11) with weight matrix defined in (4.7) and satisfying (4.9), and i.i.d. zero mean temporally correlated random vectors* $\{\mathbf{e}_q(k), \forall k > 0\}$ *assumed independent of* $\mathbf{x}(0)$. *The limit of the MSE of the state in (4.14) as k increases is given by*

$$\lim_{k\to\infty} MSE(x(k)) = (N-1)\sigma_q^2 \tag{4.36}$$

Proof. Following a similar procedure as in the proof of Proposition 4.1, we find that the bias term tends to zero whereas the variance is equal to

$$\text{var}(\mathbf{x}(k)) = \sigma_q^2 \cdot \text{tr}\left(\sum_{l=1}^{k}\sum_{m=1}^{k} \mathbf{W}^{m-1}\mathbf{W}^{l-1}\epsilon^2\mathbf{L}^2\right) \tag{4.37}$$

which is slightly different from (4.21). Using the spectral decomposition of $\mathbf{L}$ and $\mathbf{W}$ as in (4.22), the trace in (4.37) is equal to

$$\begin{aligned}\text{tr}\Big(\sum_{l=1}^{k}\sum_{m=1}^{k} \mathbf{W}^{m+l-2}\epsilon^2\mathbf{L}^2\Big) &= \text{tr}\Big(\sum_{l=1}^{k}\sum_{m=1}^{k} \mathbf{\Lambda}_{\mathbf{W}}^{m+l-2}\epsilon^2\mathbf{\Lambda}_{\mathbf{L}}^2\Big)\\ &= \sum_{i=2}^{N}\epsilon^2\lambda_{N-i+1}(\mathbf{L})^2\lambda_i(\mathbf{W})^{-2}\left(\frac{\lambda_i(\mathbf{W})-\lambda_i(\mathbf{W})^{k+1}}{1-\lambda_i(\mathbf{W})}\right)^2\end{aligned} \tag{4.38}$$

where we have used (4.24) and substituted for $\lambda_N(\mathbf{L}) = 0$. Since the eigenvalues of $\mathbf{W}$ are less than 1 in magnitude for $i = 2, \cdots N$, the term $\lambda_i(\mathbf{W})^{k+1}$ tends to zero as $k \to \infty$.

Taking the limit of the trace in (4.38) yields

$$\lim_{k\to\infty}\sum_{i=2}^{N}\epsilon^2\left(\frac{\lambda_{N-i+1}(\mathbf{L})}{\lambda_i(\mathbf{W})}\right)^2\left(\frac{\lambda_i(\mathbf{W})-\lambda_i(\mathbf{W})^{k+1}}{1-\lambda_i(\mathbf{W})}\right)^2=\sum_{i=2}^{N}\frac{\epsilon^2\lambda_{N-i+1}(\mathbf{L})^2}{(1-\lambda_i(\mathbf{W}))^2}$$
$$=(N-1).$$

where in the last equality we have used (4.8). Substituting for the bias and the variance in (4.15) we obtain the limit in (4.36), which completes the proof.

□

Remark 4.1. *The limit of the MSE in (4.36) obtained assuming a temporally correlated quantization error coincides with the upper bound for the MSE derived in (4.27) for the case of temporally uncorrelated quantization error. This behavior might be the result of a quantization error becoming temporally correlated as time evolves.*

Consider again the consensus model with quantization noise in (4.28). We analyze the MSE($\mathbf{x}(k)$) when the quantization error is assumed temporally correlated with covariance matrix given by (4.35). The bias term in this case tends to zero whereas the variance is given by

$$\begin{aligned}\mathrm{var}(\mathbf{x}(k)) &= \sigma_q^2\cdot\mathrm{tr}\left(\sum_{l=1}^{k}\sum_{m=1}^{k}\mathbf{W}^{m+l}\right)\\ &= \sigma_q^2\left(k+\sum_{i=2}^{N}\left(\frac{\lambda_i(\mathbf{W})-\lambda_i(\mathbf{W})^{k+1}}{1-\lambda_i(\mathbf{W})}\right)^2\right)\end{aligned}$$

where we have used (4.24). Again, as $k\to\infty$ the term $\lambda_i(\mathbf{W})^{k+1}$ tends to zero and the limit of the variance is

$$\lim_{k\to\infty}\mathrm{var}(\mathbf{x}(k))=\sigma_q^2\lim_{k\to\infty}\left(k+\sum_{i=2}^{N}\frac{\lambda_i(\mathbf{W})^2}{(1-\lambda_i(\mathbf{W}))^2}\right)=\infty \tag{4.39}$$

Similarly to the case of the limit computed in (4.32), assuming temporally correlated error vectors the variance of $\mathbf{x}(k)$ in (4.12) diverges as time evolves.

Consider now the consensus model with quantization noise in (4.33). We analyze the MSE($\mathbf{x}(k)$) when the quantization error is assumed temporally correlated with covariance matrix given by (4.35). The bias term in this case tends also to zero whereas the variance is given by

$$\mathrm{var}(\mathbf{x}(k)) = \sigma_q^2 \cdot \mathrm{tr}\left(\sum_{l=1}^{k}\sum_{m=1}^{k} \mathbf{W}^{m+l}(\mathbf{W}-\mathbf{W}_D)^2\right)$$

and taking the limit yields

$$\lim_{k\to\infty} \mathrm{var}(\mathbf{x}(k)) = \sigma_q^2 \lim_{k\to\infty} \mathrm{tr}\left(\sum_{l=1}^{k}\sum_{m=1}^{k} \mathbf{W}^{m+l}(\mathbf{W}-\mathbf{W}_D)^2\right) \tag{4.40}$$

which, similarly to the case in (4.34) diverges as time evolves.

According to the results obtained in (4.32) and (4.34), as well as the results in (4.39) and (4.40), we would have expected the state values in (4.12) and (4.13) to diverge. However, in the simulations section we will observe that the MSE($\mathbf{x}(k)$) in both cases indeed does not diverge. This mismatching between theory and practical results was already pointed out in [Ald06] where the authors concluded that the actual $\mathbf{e}_q(k)$ was not i.i.d. for different time instants when a quantizer is implemented. In fact, the authors observed that the variance of the quantization error decays with time. On the other hand, the expression in (4.17) shows that the MSE($\mathbf{x}(k)$) for the proposed model in (4.10) does not diverge. This theoretical approximation is useful to obtain closed-form expressions for the bounds of the MSE($\mathbf{x}(k)$) of the estimation reached by the network.

4.6 Numerical results

In this section we present the results of the computer simulations evaluating the performance of the model proposed in (4.10) implementing a uniform symmetric quantizer. The simulated models are summarized in Table 4.1, where the model in (4.6) in floating-point precision and denoted “No quant” is also included as a benchmark.

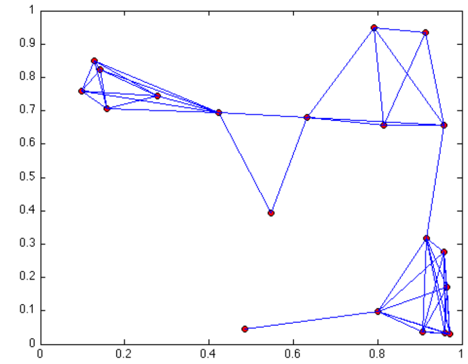
We consider a WSN composed of a set of nodes randomly deployed in a unit square where a pair of nodes are connected whenever the euclidean distance between them is smaller than a given radius a, centered at the transmitting node. The topology is assumed connected and the Laplacian matrix is symmetric. The initial measurements are modeled as Gaussian r.v.'s with mean x_0 and variance $\sigma_0^2 = 5$, uncorrelated among nodes. Since the performances of the three models with quantization depend on how close the mean x_0 is to a quantized value, x_0 is selected at random within the range $[-15, +15]$ and different at every realization. We implement a uniform symmetric quantizer operating in the range of $[-20, +20]$ with $b=3$ and $b=5$ quantization bits. The quantization noise variance for the case $b = 3$ is $\sigma_q^2 = 2.0833$, whereas for the case $b = 5$ is $\sigma_q^2 = 0.1302$. The link weights are constant and set to $\epsilon = 2/N$.

In the first set of simulations, we consider $N = 20$ nodes and connectivity radius $a = 0.35$. Then, we consider the same deployment but with a more connected topology letting $a = 0.50$. The topologies, depicted respectively in Fig. 4.2 (i) and Fig. 4.2 (ii), have an average out-degree equal to 5 in the first case and equal to 7.2 in the second case. Subfigures 4.2 (iii)-(vi) show the evolution of the MSE of the state as a function of the iteration index k, averaged over all nodes for 10.000 independent realizations of the models in Table 4.1. Subfigures (iii) and (v) depict the results for the network with radius $a=0.35$ for $b=3$ and $b=5$ quantization bits respectively, whereas subfigures (iv) and (vi) depict the results for a radius $a=0.50$ and $b=3$ and $b=5$ quantization bits.

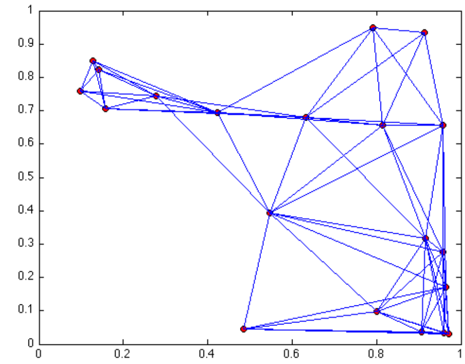
The simulations show that the MSE converges for all models and, as expected, the performance improves with a higher number of bits - which is clearly observed comparing subfigures (v) and (vi) -. The figures plot also the limits obtained with (4.17) for the temporally uncorrelated quantization noise - dashed-dotted line - and (4.36) for the temporally correlated quantization noise - dashed line -. For the case of $b = 5$, the distance between the limits obtained with (4.17) and with (4.36) becomes narrower while both values approach the theoretical benchmark. In all cases, the empirical MSE converges to a value confined between the two limits. The system with floating-point precision asymptotically reaches the average of the initial set of measurements, as expected.

No quant	$\mathbf{x}(k) = \mathbf{W}\mathbf{x}(k-1)$
Model 1	$\mathbf{x}(k) = \mathbf{W}\mathbf{x}^q(k-1)$
Model 2	$\mathbf{x}(k) = \mathbf{W}_D\mathbf{x}(k-1) + (\mathbf{W} - \mathbf{W}_D)\,\mathbf{x}^q(k-1)$
Proposed	$\mathbf{x}(k) = \mathbf{x}(k-1) - \epsilon\mathbf{L}\mathbf{x}^q(k-1)$

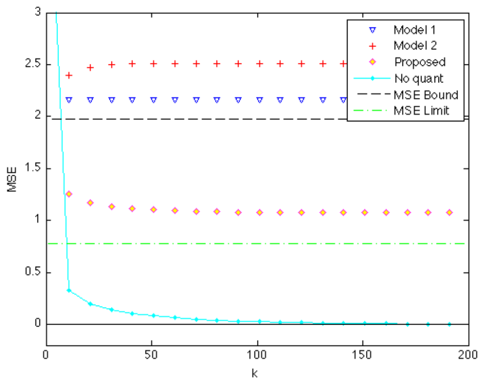
Table 4.1: Simulated consensus models including quantization.



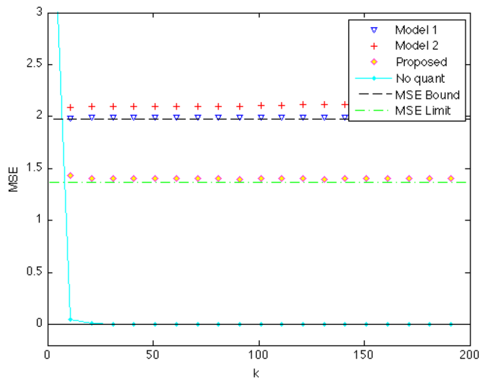
(i) Deployment of 20 nodes with average out-degree 5.



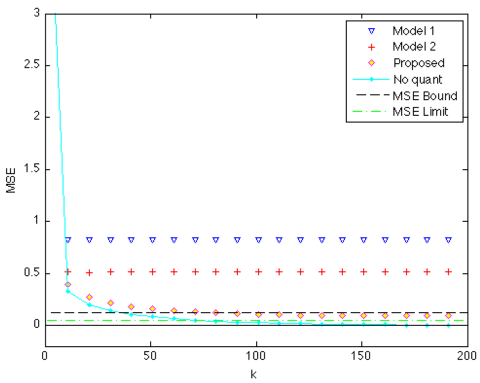
(ii) Deployment of 20 nodes with average out-degree 7.2.



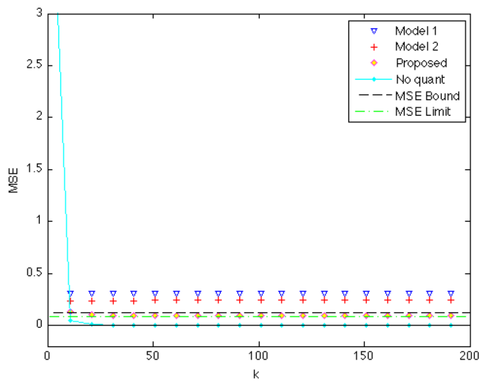
(iii) Averaged MSE of the state for $a = 0.35$ and $b = 3$.



(iv) Averaged MSE of the state for $a = 0.5$ and $b = 3$.



(v) Averaged MSE of the state for $a = 0.35$ and $b = 5$.



(vi) Averaged MSE of the state for $a = 0.5$ and $b = 5$.

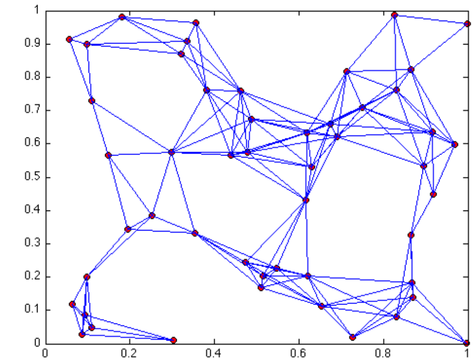
Figure 4.2: WSN deployments with $N = 20$ nodes and connectivity radii $a = 0.35$ (i), (iii), (v), and $a = 0.5$ (ii), (iv), (vi). The MSE of the state is depicted for the cases of $b = 3$ and $b = 5$ quantization bits respectively.

Clearly, as the number of quantization bits increases, the performance of all models improves and the curves approach the one for the infinite precision model. The results for Model 1 show the worst performance even with a relatively high number of quantization bits. Comparing the performance of the proposed model and Model 2, the former attains a smaller MSE in all cases. For a higher average out-degree, the performance of the MSE for the proposed model approaches the theoretical limit in (4.17). In other words, the theoretical model in (4.11) is better suited for a more connected network. We observe also that an increase in the average out-degree leads to a narrower distance between the theoretical limit in (4.17) and the theoretical limit in (4.36), which coincides with the upper bound derived in (4.27). This may be caused by the fact that, as the connectivity increases, the magnitude of the eigenvalues of $\mathbf{L}$ increase and the limit in (4.17) approaches the upper bound in (4.27).

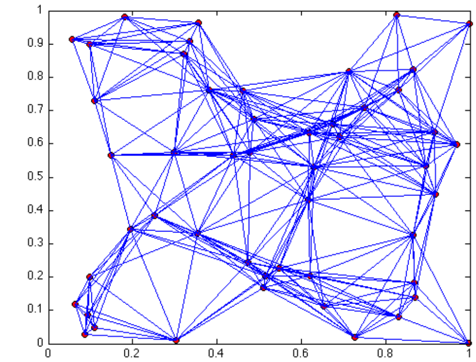
In the second set of simulations, we consider $N=50$ and connectivity radius $a=0.25$. We consider then the same deployment with a higher number of connection links letting $a=0.35$, depicted in Fig. 4.3 (i) and 4.3 (ii) respectively. The topology has an average out-degree equal to 6.6 in the first case and equal to 12 in the second case. Again, subfigures (iii) and (v) depict the results for the network with radius $a = 0.25$ for $b = 3$ and $b = 5$ quantization bits respectively, whereas subfigures (iv) and (vi) show the results for a radius $a=0.35$ with $b=3$ and $b=5$ quantization bits.

Analogously to the first set of simulations, we observe that an increase in the number of bits used for quantification leads to a smaller MSE for all models and a narrower distance between the theoretical limits. An increase in the connectivity of the network leads also to a better fit of the curve for the proposed model with the theoretical limit in (4.17).

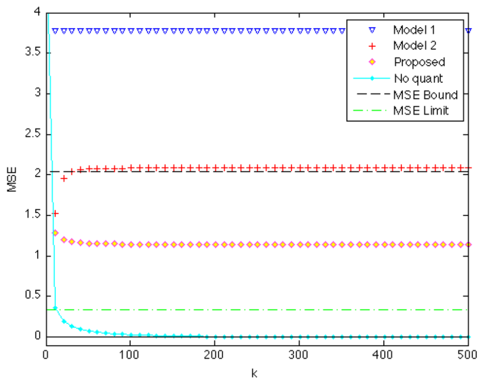
In all topologies, the proposed model in (4.10) outperforms Models 1 and Model 2 and results in a smaller residual MSE. Moreover, for a higher average out-degree, the performance of the proposed model approaches the theoretical limit in (4.17). For a number of quantization bits greater than 5, all models in Table 4.1 behave quite similar and the curves approach the average of the initial set of measurements. Summing up, increasing the number of bits used for quantization we obtain a smaller residual MSE with respect to the infinite precision model.



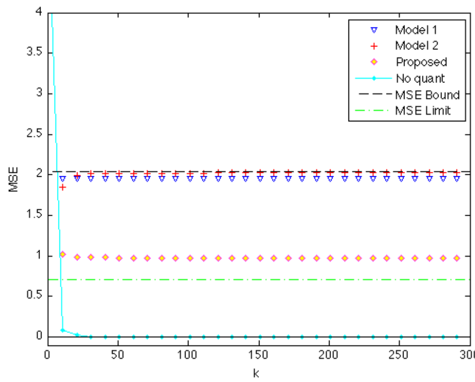
(i) Deployment of 50 nodes with average out-degree 6.6.



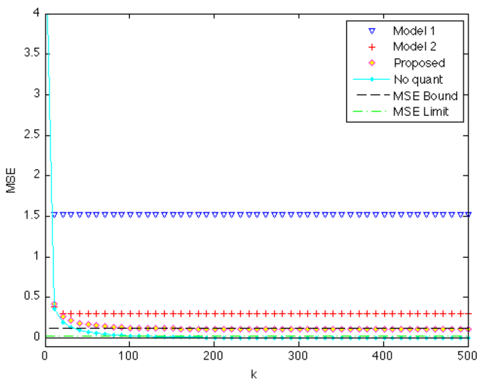
(ii) Deployment of 20 nodes with average out-degree 12.



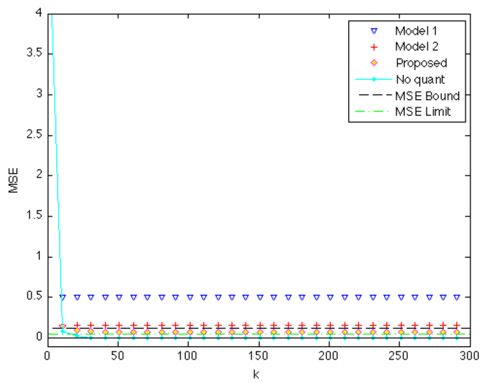
(iii) Averaged MSE of the state for $a = 0.25$ and $b = 3$.



(iv) Averaged MSE of the state for $a = 0.35$ and $b = 3$.



(v) Averaged MSE of the state for $a = 0.25$ and $b = 5$.



(vi) Averaged MSE of the state for $a = 0.35$ and $b = 5$.

Figure 4.3: WSN deployments with $N = 50$ nodes and connectivity radii $a = 0.25$ (i), (iii), (v), and $a = 0.35$ (ii), (iv), (vi). The MSE of the state is depicted for the cases of $b = 3$ and $b = 5$ quantization bits respectively.

4.7 Conclusions of the Chapter

A model to achieve the average consensus in a WSN where the information exchanged among nodes is quantized has been presented. The analysis of the MSE of the state has been carried out considering spatially independent and both temporally uncorrelated and temporally correlated quantization error vectors.

Conversely to the cases for two other models studied in the chapter, the MSE of the state for the proposed model converges and its limit admits a closed-form expression in both cases of temporally uncorrelated quantization noise and temporally correlated quantization noise. Therefore, the limit of the MSE of the state can be computed analytically offline.

The simulations show that, when a uniform symmetric quantizer is used, the proposed model outperforms similar existing consensus models that also include quantization in terms of the MSE. Moreover, the simulations show an increased agreement between the practical model including quantization and the theoretical approximation using the additive noise model as the average out-degree of the network increases.

An upper bound for the limit of the MSE that depends only on general network parameters is also derived. This upper bound might be useful in the design of the quantizer implemented by the nodes.

5

Mean Square Consensus in Random Networks

5.1 Introduction

The communications in a WSN are usually exposed to node and channel failures, changes in the environment, mobility of the nodes or asynchronous sleeping periods which make the topology vary randomly with time. The randomness of the topology can be owing also to random communication protocols. When the existence of a link between a given pair of nodes is random, the convergence of consensus algorithms is studied in probabilistic terms [Kus71,Mao94] and the information flow among the nodes of the network is modeled by a random graph.

The instantaneous links of the underlying random graph model of a WSN are usually assumed to exist with a given probability, which can be assumed equal for all nodes as in the case of Erdős-Rényi [Erd60] topologies or different for all pairs of nodes. If a random geometric topology is considered[1], the probability of connection for a link is nonzero only for neighboring nodes within a connectivity radius and zero otherwise [Bol01, Pen03].

The convergence analysis of the consensus algorithm in networks with random topologies can be carried out considering either instantaneous undirected topologies, i.e., the links are bidirectional at any time instant, or considering instantaneous directed topolo-

[1] The random geometric topology refers to a random geometric graph as defined in Chapter 2.

gies, i.e., the links are unidirectional at any time instant. When the instantaneous topology is assumed undirected or balanced directed, the states of the nodes converge to the average of the initial values. Particularly, when the instantaneous topology is undirected the Laplacian matrix of the underlying graph model is positive semidefinite and all its eigenvalues are real. In a practical implementation however, a symmetric instantaneous topology may be a strong requirement since it compels the communication system to implement for instance an acknowledgement protocol to ensure reciprocal exchange of information at each iteration. On the other side, when the instantaneous topology is allowed to be directed, the Laplacian matrix is non-symmetric at every iteration and, if it is in addition non-balanced, the states of the nodes converge to a value different from the initial average. This is a less restrictive assumption than the requirement of symmetry at every time instant.

Furthermore, the links can be spatially independent or spatially correlated, as for the case of random gossip algorithms. In this chapter we study the convergence in the mean square sense of the algorithm in [OS04] for WSNs with random directed topologies and statistically independent links, where the MSE of the state is evaluated with respect to the statistical mean of the initial measurements. The constraint on instantaneous link symmetry is relaxed and only symmetric probability of connection for every pair of links is assumed. We consider a design of the link weights based on the MSE analysis to minimize the convergence time of the algorithm. We start reviewing important contributions found in literature regarding statistically independent random links in undirected and directed topologies separately.

Contributions on Undirected Random Topologies

For Erdős-Rényi topologies, Hatano and Mesbahi [Hat04] shows convergence to a consensus in probability using notions of stochastic stability, whereas Patterson *et al.* [Pat07] characterizes the convergence rate of the consensus algorithm in terms of the eigenvalues of a Lyapunov-like matrix recursion.

For links with different probabilities, Kar and Moura [Kar07] relates mean square convergence of the consensus algorithm to the second smallest eigenvalue of the average

Laplacian and derives bounds on the convergence rate, whereas [Kar08b] uses semi-definite convex programming to propose a probabilistic topology that maximizes the convergence rate when an overall budget for the consumption of energy is considered. Furthermore, the works in [Kar09] and [Kar10] include the assumptions of channel and quantization noise respectively and propose two different models to reduce either the bias or the variance produced by the noise in the consensus value.

Contributions on Directed Random Topologies

For Erdős-Rényi topologies, Wu [Wu06] uses results from inhomogeneous Markov chains to derive sufficient conditions for convergence in probability in graphs containing spanning trees, whereas Preciado *et al.* [Pre10] derives closed-form expressions for the mean and the variance of the asymptotic consensus value.

Rabbat *et al.* [Rab05] shows that the average consensus is asymptotically achieved in directed topologies provided that the value of the link weights tends to zero, but at the cost of significantly increasing the convergence time.

Assuming stochastic weight matrices with positive diagonals, Tahbaz-Salehi and Jadbabaie [TS08] uses ergodicity properties to show almost sure convergence to a common value, and relates the convergence to the second largest eigenvalue of the average weight matrix.

Porfiri and Stilwell [Por07] shows that a sufficient condition for almost sure convergence in continuous systems is that the eigenvalues of the average Laplacian matrix have positive real parts and that the topology varies sufficiently fast, assuming either positive weights or arbitrary weights. The paper shows that consensus is asymptotically achieved in random directed graphs almost surely if the probability that the network is strongly connected is nonzero, and this happens whenever the expected network is strongly connected. A similar approach is presented in [Por08], where almost sure local synchronization of oscillators in a random weighted directed graph is shown.

Finally, in Zhou and Wang [Zho09] the asymptotic and per-step convergence factors from [Xia03] are redefined to characterize the convergence speed of the consensus algorithm using stochastic stability notions.

Outline of the Chapter

The chapter is organized as follows. Section 5.2 presents the consensus algorithm for networks with random topologies and in Section 5.3 convergence in expectation to the statistical mean of the initial values is shown. In Section 5.4 the convergence of the MSE of the state values is analyzed and a general expression for the MSE which can be computed for each iteration is derived. The general expression for the MSE is analyzed distinguishing between random links with equal probability and random links with different probabilities of connection. Section 5.5 addresses the case of an Erdős-Rényi random graph model [Erd60], where a closed-form expression for the MSE of the state and for the asymptotic MSE are derived along with the dynamic range and the optimum link weights minimizing the convergence time. In Section 5.6 the case of different probabilities of connection among the nodes is considered and an upper bound for the MSE of the state is derived. Furthermore, the asymptotic upper bound is studied and the optimum link weight minimizing the convergence time of the upper bound is defined. Section 5.7 presents a practical implementation of the algorithm with a scheme of randomized transmission power where the results from Section 5.6 are used to reduce the convergence time. Finally, the conclusions for the chapter are presented in Section 5.8.

5.2 Consensus in Random Directed Topologies

Consider a WSN with N nodes characterized by a directed random graph $\mathcal{G}(k) = \{\mathcal{V}, \mathcal{E}(k)\}$ where the communication links exist with a given probability, such that $e_{ij} \in \mathcal{E}(k)$ with probability $0 \leq p_{ij} \leq 1$. We assume that the probabilities are symmetric, i.e., $p_{ij} = p_{ji}$. Defining the connection probability matrix with entries $\mathbf{P}_{ij} = p_{ij}$ and $\mathbf{P}_{ii} = 0$ for all $\{i, j\} \in \mathcal{V}$, the instantaneous adjacency matrix $\mathbf{A}(k) \in \mathbb{R}^{N\times N}$ of $\mathcal{G}(k)$ is random with temporally independent entries given by

$$[\mathbf{A}(k)]_{ij} = \begin{cases} 1 & \text{with probability } p_{ij} \\ 0 & \text{with probability } 1 - p_{ij} \end{cases}$$

and mean $\bar{\mathbf{A}} = \mathbf{P}$. The instantaneous Laplacian matrix given by $\mathbf{L}(k) = \mathbf{D}(k) - \mathbf{A}(k)$ is random with mean $\bar{\mathbf{L}} = \bar{\mathbf{D}} - \mathbf{P}$, where $\bar{\mathbf{D}} = diag(\mathbf{P} \cdot \mathbf{1})$.

The consensus algorithm under these topology conditions has a difference equation given by

$$\mathbf{x}(k) = \mathbf{W}(k-1)\mathbf{x}(k-1), \quad \forall k > 0 \tag{5.1}$$

where $\mathbf{x}(0)$ is the vector of initial measurements, considered random, and the weight matrix at time k is modeled as

$$\mathbf{W}(k) = \mathbf{I} - \epsilon\mathbf{L}(k), \quad \forall k \geq 0 \tag{5.2}$$

with random non-symmetric Laplacian $\mathbf{L}(k)$ and constant $\epsilon > 0$, equal for all the links. The matrices $\{\mathbf{W}(k), \forall k \geq 0\}$ in (5.1) are random, assumed independent of each other and identically distributed with one eigenvalue equal to 1 and associated right eigenvector $\mathbf{1}$. The associated left eigenvector varies randomly from realization to realization because $\mathbf{W}(k)$ is, in general, non-symmetric. The evolution of the state vector $\mathbf{x}(k)$ in (5.1) can be then rewritten as follows

$$\mathbf{x}(k) = \mathbf{M}_{\mathbf{w}}(k)\mathbf{x}(0), \quad \forall k > 0 \tag{5.3}$$

where the product matrix

$$\mathbf{M}_{\mathbf{w}}(k) = \prod_{l=1}^{k} \mathbf{W}(k-l), \quad \forall k > 0 \tag{5.4}$$

is assumed independent of $\mathbf{x}(0)$ for all k. The product matrix $\mathbf{M}_{\mathbf{w}}(k)$ has row-sums equal to 1, and this is easily shown noticing that the matrices $\{\mathbf{W}(k), \forall k \geq 0\}$ have at least one eigenvalue equal to 1 with associated right eigenvector $\mathbf{1}$. For those realizations with $k \to \infty$ for which the eigenvalue $\lambda_1(\mathbf{M}_{\mathbf{w}}(k)) = 1$ has algebraic multiplicity one, we have that

$$\lim_{k\to\infty} \mathbf{x}(k) = \frac{1}{N}\mathbf{1}\boldsymbol{\gamma}_{\mathbf{M}}^T\mathbf{x}(0)$$

where $\boldsymbol{\gamma}_{\mathbf{M}}$ is the left eigenvector associated with $\lambda_1(\mathbf{M}_{\mathbf{w}}(k))$ for $k \to \infty$.

From the results in [Rab05], we know that as the value of ϵ approaches zero, the states of the nodes asymptotically reach the average consensus $\mathbf{x}_{ave} = \mathbf{J}_N\mathbf{x}(0)$ where $\mathbf{J}_N = \mathbf{1}\mathbf{1}^T/N$, coinciding in this case with the ML estimator. In other words, $\boldsymbol{\gamma}_{\mathbf{M}}$ tends to a vector of all-ones as k increases. In general, as the value of ϵ decreases, the convergence time of the consensus algorithm in (5.1) increases. For that reason, we consider a value of ϵ not tending to zero, although at the cost of deviating from the average of the initial measurements.

5.3 Convergence in Expectation to the Mean Average Consensus

Assume a set of initial measurements modeled as i.i.d. Gaussian r.v.'s with mean x_m and variance σ_0^2, such that

$$\begin{aligned} \mathbb{E}[\mathbf{x}(0)] &= x_m \mathbf{1} \\ \mathbb{E}[\mathbf{x}(0)\mathbf{x}(0)^T] &= \sigma_0^2 \mathbf{I} + x_m^2 \mathbf{J} \end{aligned} \tag{5.5}$$

where $\mathbf{J} = \mathbf{1}\mathbf{1}^T$. Recalling the independence assumption of $\mathbf{M_w}(k)$ and $\mathbf{x}(0)$ in (5.3), we have

$$\mathbb{E}\left[\mathbf{x}(k)\right] = \mathbb{E}\left[\mathbf{M_w}(k)\right]\mathbb{E}\left[\mathbf{x}(0)\right]$$

and using the i.i.d assumption of the weight matrices $\{\mathbf{W}(k), \forall k \geq 0\}$

$$\mathbb{E}\left[\mathbf{x}(k)\right] = \bar{\mathbf{W}}^k \mathbb{E}\left[\mathbf{x}(0)\right].$$

In order for the matrix power to converge, the expected weight matrix should satisfy the following conditions

$$\begin{aligned} \bar{\mathbf{W}}\mathbf{1} &= \mathbf{1} \\ \mathbf{1}^T\bar{\mathbf{W}} &= \mathbf{1}^T \;. \\ \rho(\bar{\mathbf{W}} - \mathbf{J}_N) &< 1 \end{aligned} \tag{5.6}$$

Remark that using (5.2), the expected weight matrix is given by $\bar{\mathbf{W}} = \mathbf{I} - \epsilon\bar{\mathbf{L}}$, where $\bar{\mathbf{L}}$ is the expected Laplacian. $\bar{\mathbf{L}}$ can be seen as the Laplacian of a fixed undirected graph, defined as the expected graph $\mathbb{E}[\mathcal{G}(k)] = \bar{\mathcal{G}}$. Due to the assumption of a symmetric connection probability matrix $\mathbf{P}$, $\bar{\mathbf{L}}$ is symmetric, meaning that $\bar{\mathcal{G}}$ is undirected. In addition, $\bar{\mathcal{G}}$ is assumed connected such that the smallest eigenvalue of the Laplacian $\lambda_N(\bar{\mathbf{L}}) = 0$ with algebraic multiplicity one. Summing up, $\bar{\mathbf{W}}$ is symmetric by construction and double-stochastic for $\epsilon \in (0, 2/\lambda_1(\bar{\mathbf{L}}))$ with eigenvalues $\lambda_i(\bar{\mathbf{W}}) = 1 - \epsilon\lambda_{N-i+1}(\bar{\mathbf{L}})$ for all $i \in \{1, \cdots, N\}$. Using (5.5), we can conclude that $\bar{\mathbf{W}}$ satisfies the convergence conditions in (5.6) since

$$\begin{aligned} \lim_{k\to\infty} \mathbb{E}\left[\mathbf{x}(k)\right] &= \lim_{k\to\infty} \bar{\mathbf{W}}^k \mathbb{E}\left[\mathbf{x}(0)\right] \\ &= \mathbf{J}_N x_m \mathbf{1} \\ &\triangleq \bar{\mathbf{x}}_{ave} \end{aligned}$$

where $\bar{\mathbf{x}}_{ave}$ is denoted the *mean average consensus* vector, showing that the estimation is asymptotically unbiased. Remark that if the set of initial values $\mathbf{x}(0)$ is assumed deterministic, the average consensus $\mathbf{x}_{ave}$ is reached in expectation. Since the state vector $\mathbf{x}(0)$ is random and since the underlying random graph has an instantaneous directed topology, we consider the convergence of the state vector $\mathbf{x}(k)$ in the mean square sense to $\bar{\mathbf{x}}_{ave}$. The following section addresses the analysis of the MSE of the state.

5.4 Mean Square Convergence to a Consensus

We evaluate the deviation of $\mathbf{x}(k)$ from the vector of statistical means $\bar{\mathbf{x}}_{ave}$ and study the impact of the value of ϵ on this deviation. Therefore, we analyze the MSE of the state defined as

$$\mathrm{MSE}(x(k)) = \frac{1}{N}\mathbb{E}\left[\left\|\mathbf{x}(k) - \bar{\mathbf{x}}_{ave}\right\|_2^2\right]. \tag{5.7}$$

Substituting (5.3) in (5.7) and expanding the expression yields

$$\begin{aligned}\mathrm{MSE}(x(k)) = \frac{1}{N}\mathbb{E}\big[&\mathbf{x}(0)^T\mathbf{M}_{\mathbf{w}}^T(k)\mathbf{M}_{\mathbf{w}}(k)\mathbf{x}(0)\\ &-\mathbf{x}(0)^T\mathbf{M}_{\mathbf{w}}^T(k)\bar{\mathbf{x}}_{ave} - \bar{\mathbf{x}}_{ave}^T\mathbf{M}_{\mathbf{w}}(k)\mathbf{x}(0) + \bar{\mathbf{x}}_{ave}^T\bar{\mathbf{x}}_{ave}\big].\end{aligned} \tag{5.8}$$

For convenience, consider the matrix

$$\mathbf{R}_{\mathbf{w}}(k) = \mathbb{E}\left[\mathbf{M}_{\mathbf{w}}^T(k)\mathbf{M}_{\mathbf{w}}(k)\right], \quad k > 0 \tag{5.9}$$

which has row-sums and column-sums equal to 1. This is easily observed noting that $\mathbf{R}_{\mathbf{w}}(k)$ is symmetric by definition and has row-sums equal to 1, i.e.,

$$\begin{aligned}\mathbf{R}_{\mathbf{w}}(k)\mathbf{1} &= \mathbb{E}\left[\mathbf{M}_{\mathbf{w}}^T(k)\mathbf{M}_{\mathbf{w}}(k)\mathbf{1}\right]\\ &= \mathbb{E}\left[\mathbf{M}_{\mathbf{w}}^T(k)\mathbf{1}\right] = (\bar{\mathbf{W}}^k)^T\mathbf{1} = \mathbf{1}\end{aligned}$$

where the second equality holds because $\mathbf{M}_{\mathbf{w}}(k)\mathbf{1} = \mathbf{1}$ and the last equality holds because $\bar{\mathbf{W}}$ satisfies the convergence conditions in (5.6). However, $\mathbf{R}_{\mathbf{w}}(k)$ is not necessarily a nonnegative matrix for all k. Furthermore, using (5.5) and (5.9) and considering the independence of $\mathbf{M}_{\mathbf{w}}(k)$ and $\mathbf{x}(0)$, the expression in (5.8) can be rewritten as

$$\begin{aligned}\mathrm{MSE}(x(k)) &= \frac{1}{N}\Big(\mathrm{tr}\big((\sigma_0^2\mathbf{I} + x_m^2\mathbf{J})\cdot\mathbf{R}_{\mathbf{w}}(k)\big) - x_m\mathbf{1}^T(\bar{\mathbf{W}}^k)^T\bar{\mathbf{x}}_{ave}\\ &\quad - x_m\bar{\mathbf{x}}_{ave}^T\bar{\mathbf{W}}^k\mathbf{1} + \bar{\mathbf{x}}_{ave}^T\bar{\mathbf{x}}_{ave}\Big)\\ &= \frac{\sigma_0^2}{N}\cdot\mathrm{tr}\Big(\mathbf{R}_{\mathbf{w}}(k)\Big).\end{aligned} \tag{5.10}$$

The expression in (5.10) tells us that the MSE$(x(k))$ will be deviated from the variance of the ML estimator, i.e., σ_0^2/N, by a factor $\mathrm{tr}(\mathbf{R_w}(k))$. In order to characterize this deviation we start expanding the expression for $\mathbf{R_w}(k)$ in (5.9). Applying the linearity of the trace and the expected value operators and using the fact that the random matrices $\{\mathbf{W}(k), \forall k \geq 0\}$ are i.i.d., (5.10) can be expressed as follows

$$\begin{aligned} \mathrm{MSE}(x(k)) &= \frac{\sigma_0^2}{N} \cdot \mathrm{tr}\Big(\mathbb{E}\big[\mathbf{W}^T(1)\dots\mathbf{W}^T(k-1)\mathbf{W}(k-1)\dots\mathbf{W}(1)\big] \cdot \mathbb{E}\big[\mathbf{W}(0)\mathbf{W}^T(0)\big]\Big) \\ &= \frac{\sigma_0^2}{N} \cdot \mathrm{tr}\Big(\mathbf{R_w}(k-1) \cdot \mathbf{C_w}\Big) \end{aligned} \tag{5.11}$$

where

$$\begin{aligned} \mathbf{C_w} &= \mathbb{E}\big[\mathbf{W}(k)\mathbf{W}^T(k)\big] \\ &= \mathbf{I} - 2\epsilon\bar{\mathbf{L}} + \epsilon^2\mathbb{E}\big[\mathbf{L}(k)\mathbf{L}(k)^T\big] \end{aligned}$$

and in the last equality we have substituted for (5.2). After some matrix manipulations, we observe that $\bar{\mathbf{L}}$ and $\mathbb{E}\big[\mathbf{L}(k)\mathbf{L}(k)^T\big]$ have entries given respectively by

$$\bar{\mathbf{L}}_{ij} = \begin{cases} \sum_{l=1}^{N} p_{il} & i = j \\ -p_{ij} & i \neq j \end{cases} \tag{5.12}$$

$$\mathbb{E}\big[\mathbf{L}(k)\mathbf{L}(k)^T\big]_{ij} = \begin{cases} 2\sum_{l=1}^{N} p_{il} + \sum_{l=1}^{N}\sum_{\substack{m=1\\ m\neq l}}^{N} p_{il}p_{im} & i = j \\ -p_{ij}\left(\sum_{l=1}^{N} p_{il} + \sum_{l=1}^{N} p_{jl}\right) + \sum_{\substack{l=1\\ l\neq j}}^{N} p_{il}p_{jl} & i \neq j \end{cases}. \tag{5.13}$$

Therefore, $\mathbf{C_w}$ can be expressed analytically as

$$\mathbf{C_w} = \mathbf{I} - 2\epsilon\bar{\mathbf{L}} + \epsilon^2(\bar{\mathbf{L}}^2 + 2(\bar{\mathbf{D}} - \tilde{\mathbf{D}})) \tag{5.14}$$

where $\tilde{\mathbf{D}}$ is a diagonal matrix with $\{ii\}^{th}$ entry

$$\tilde{\mathbf{D}}_{ii} = [(\mathbf{P} \odot \mathbf{P})\mathbf{1}]_i, \quad \forall i = 1, \cdots, N \tag{5.15}$$

and $\odot$ denotes the Schur product. Remark that $\tilde{\mathbf{D}}$ is the degree matrix of a graph whose adjacency matrix has entries equal to the squared entries of $\mathbf{P}$. The next step consists in analyzing the MSE$(x(k))$ expression in (5.11) in two different scenarios: when all the links have the same probability of connection and when all the links have different probability of connection.

5.5 MSE Analysis for Links with Equal Probability

In this section we assume an Erdős-Rényi random graph, where for any pair of nodes $\{i, j\} \in \mathcal{V}$, a communication link between them exists with probability $p_{ij} = p_{ji} = p$. The following theorem resumes an important result:

Theorem 5.1. *Consider the consensus algorithm in (5.1) with N nodes and weight matrix defined in (5.2), probability of connection $0 \le p \le 1$ equal for all the links and i.i.d. initial values $\mathbf{x}(0)$ with mean x_m and variance σ_0^2. The MSE$(x(k))$ in (5.7) is equal to*

$$MSE(x(k)) = \sigma_0^2 \left(\frac{b}{1-a+b} - \frac{(a-1)}{1-a+b}(a-b)^k \right) \tag{5.16}$$

where

$$\begin{aligned} a &= 1 - 2(N-1)p\epsilon + (2(N-1)p + (N-1)(N-2)p^2)\epsilon^2 \\ b &= 2p\epsilon - Np^2\epsilon^2 \end{aligned} \tag{5.17}$$

Proof. Replacing for $\bar{\mathbf{L}} = p(N\mathbf{I} - \mathbf{J})$, $\bar{\mathbf{D}} = (N-1)p\mathbf{I}$ and $\tilde{\mathbf{D}} = (N-1)p^2\mathbf{I}$ in (5.15), we obtain

$$\mathbf{C}_\mathbf{w} = b \cdot \mathbf{J} + \mathbf{I}(a-b) \tag{5.18}$$

where a and b are as defined in (5.17). Then, replacing (5.18) in (5.11), the expression for the MSE becomes

$$\text{MSE}(x(k)) = \frac{\sigma_0^2}{N}\Big(Nb + (a-b)\cdot \text{tr}(\mathbf{R}_\mathbf{w}(k-1))\Big)$$

where we have used that $\mathbf{R}_\mathbf{w}(k-1)$ has row-sums equal to 1. Substituting the trace above recursively and noting that $\text{tr}(\mathbf{R}_\mathbf{w}(1)) = \text{tr}(\mathbf{C}_\mathbf{w}) = Na$, we obtain

$$\begin{aligned} \frac{\sigma_0^2}{N}\cdot \text{tr}(\mathbf{R}_\mathbf{w}(k)) &= \frac{\sigma_0^2}{N}\cdot N\left(\sum_{l=0}^{k-2} b(a-b)^l + a(a-b)^{k-1}\right) \\ &= \sigma_0^2\left(\frac{b}{1-a+b} - \frac{(a-1)}{1-a+b}(a-b)^k\right) \end{aligned}$$

and the proof is completed.

□

Remark 5.1. *The closed-form expression derived in Theorem 5.1 allows us to compute the MSE*$(x(k))$ *at* any *time instant off-line, as it only requires knowledge of the general parameters* N, p *and* ϵ.

In the following sections we study the convergence conditions for the MSE$(x(k))$ in (5.16) and the residual MSE after convergence.

5.5.1 Fast MSE Convergence and Optimum Link Weights

Since the convergence of the MSE$(x(k))$ is related to the value of the link weights ϵ, we aim at determining the dynamic range of ϵ, finding the value that maximizes the convergence rate of the MSE$(x(k))$ and evaluating the convergence time.

We observe that the convergence of (5.16) is related to the term $(a-b)$. For simplicity, for a given number of nodes $N > 1$ and probability of connection $0 \leq p \leq 1$ consider the following function

$$\begin{aligned} f(\epsilon) &= a - b \\ &= 1 - 2Np\epsilon + \left(2(N-1)p + ((N-1)^2 + 1)p^2\right)\epsilon^2. \end{aligned} \tag{5.19}$$

According to (5.16), the MSE$(x(k))$ converges whenever $f(\epsilon)^k \to 0$, and a necessary and sufficient condition for the power to approach zero as $k \to \infty$ is that $|f(\epsilon)| < 1$. It is not difficult to check that for $N > 1$ and $0 \leq p \leq 1$, $f(\epsilon)$ is a quadratic nonnegative function with $f(0) = 1$ and negative derivative in the proximity of $\epsilon = 0$. Therefore, the optimum ϵ can be easily determined, as stated in the following lemma:

Lemma 5.1. *For a given number of nodes* N *and a given probability of connection* $0 \leq p \leq 1$*, the value of* ϵ *that minimizes the convergence time of the MSE*$(x(k))$ *in (5.16) is given by*

$$\epsilon^* = \frac{N}{2(N-1) + (N-1)^2 p + p}. \tag{5.20}$$

Proof. The value of ϵ^* in (5.20) corresponds to the minimum of a convex nonnegative quadratic function, namely $f(\epsilon)$ in (5.19). The value of ϵ^* in (5.20) minimizes therefore $a-b$, resulting in a faster convergence of the MSE expression in (5.16).

□

The value of $f(\epsilon)$ at the point ϵ^* is given by

$$f(\epsilon^*) = \frac{2(N-1)(1-p)}{2(N-1)+(N-1)^2p+p}.$$

Note that since $f(0)=1$ and the derivative of $f(\epsilon)$ is negative in the proximity of $\epsilon=0$, we have $0 < f(\epsilon^*) < 1$. The convergence time of the MSE$(x(k))$ is proportional to the term

$$\tau_{mse} = -\frac{1}{\ln f(\epsilon)}.$$

Substituting for $f(\epsilon^*)$, the convergence rate of the algorithm is proportional to

$$\tau_{mse} = -\ln^{-1}\left(\frac{2(N-1)(1-p)}{2(N-1)+(N-1)^2p+p}\right). \tag{5.21}$$

It is interesting to note that, as the network size increases, the value of ϵ^* approaches zero and the convergence rate of the algorithm given by (5.21) is faster. The impact of the network size on the convergence time is more evident for low values of p because, as p approaches 1, the consensus value is reached in a single iteration.

The results from Lemma 5.1 allow us to bound the values of ϵ ensuring convergence in the mean square sense, as stated in the following theorem:

Theorem 5.2. *Consider the consensus algorithm in (5.1) with N nodes and weight matrix defined in (5.2), probability of connection $0 \le p \le 1$ equal for all the links and i.i.d. initial values $\mathbf{x}(0)$ with mean x_m and variance σ_0^2. Then, $\mathbf{x}(k)$ converges in the mean square sense if*

$$0 < \epsilon < \frac{2N}{2(N-1)+(N-1)^2p+p}. \tag{5.22}$$

Proof. The demonstration follows from Lemma 5.1 and the fact that $f(\epsilon)$ is a quadratic nonnegative function with $f(0) = 1$, negative derivative in the proximity of $\epsilon = 0$ and vertex at the point ϵ^*, whose value is given by (5.20). The dynamic range in (5.22) is therefore determined by the values of ϵ ensuring $f(\epsilon) < 1$.

□

Remark 5.2. *Theorem 5.2 resumes an important result. It states that if we choose ϵ belonging to the interval defined in (5.22), we can guarantee that as $k \to \infty$, the algorithm in (5.1) converges in the mean square sense to a consensus. In addition, choosing the optimum link weight ϵ^* in (5.20), the convergence time of the MSE$(\mathbf{x}(k))$ is minimized and proportional to (5.21).*

For a network with a Erdős-Rényi topology, the expected Laplacian $\bar{\mathbf{L}}$ has one eigenvalue equal to 0 with algebraic multiplicity one and one eigenvalue equal to Np with algebraic multiplicity $(N-1)$. As the number of nodes increases, the upper bound for ϵ in (5.22) approaches the value $2/Np = 2/\lambda_1(\bar{\mathbf{L}})$. Moreover, in that case

$$\epsilon^* = \frac{1}{Np} = \frac{2}{\lambda_2(\bar{\mathbf{L}}) + \lambda_N(\bar{\mathbf{L}})}$$

coinciding with the results derived in [Xia03] for a network with time-invariant topology.

According to the results in [TS08, Corollary 4], a necessary and sufficient condition for almost sure convergence of the consensus algorithm in (5.1) is that the second largest eigenvalue of the average weight matrix satisfies $|\lambda_2(\bar{\mathbf{W}})| < 1$, assuming that the matrices $\{\mathbf{W}(k), \forall k\}$ have positive diagonal entries. In our model however, we do not restrict the instantaneous weight matrices to have positive diagonal entries, since that condition would require $\epsilon < 1/(N-1)$ and $1/(N-1) < \epsilon^*$ for $p < 1 - 1/N$, resulting in a penalization of the convergence speed.

[Kar08b, Lemma 11] states that for a network connected in average over time and implementing the uniform weights model, the value of the link weight that minimizes the convergence time of (5.1) in the mean square sense belongs to the interval $(0, 2/\lambda_2(\bar{\mathbf{L}}))$, which in this case translates to $0 < \epsilon < 2/Np$. However, the value of the optimum link weight is not specified in [Kar08b], whereas in Lemma 5.1 we provide this value when the links have equal probability of connection. In fact, it can be checked that ϵ^* in (5.20) belongs also to the interval $(0, 2/Np)$ specified in [Kar08b].

5.5.2 Asymptotic Behavior of the MSE

In this section we aim at determining the impact of the number of nodes and the probability of connection on the asymptotic MSE$(x(k))$. In order to determine the asymptotic MSE for the case of equally probable links, we choose a value of ϵ in the interval defined in (5.22). Then, the limit of the MSE$(x(k))$ as $k \to \infty$ is given by

$$\begin{aligned}\lim_{k\to\infty} \mathrm{MSE}(x(k)) &= \sigma_0^2 \left(\frac{b}{1-a+b}\right)\\ &= \frac{\sigma_0^2}{N}\left(\frac{2N - N^2 p\epsilon}{2N - (2(N-1) + (N-1)^2 p + p)\,\epsilon}\right)\\ &= \frac{\sigma_0^2}{N}\cdot g(\epsilon) \qquad (5.23)\end{aligned}$$

where we have substituted for a and b defined in (5.17). Clearly, the function $g(\epsilon)$ in (5.23) approaches 1 as $\epsilon \to 0$, so the MSE at each node tends to σ_0^2/N as the value of $\epsilon \to 0$, and this result is compliant with [Rab05]. Therefore, $g(\epsilon)$ provides the deviation of the MSE$(x(k))$ with respect to the optimum σ_0^2/N. Particularly, the deviation with respect to σ_0^2/N for the optimum ϵ^* can be computed as follows

$$g(\epsilon^*) = \frac{4(N-1)(1-p) + N^2 p}{2(N-1) + (N-1)^2 p + p}.$$

Furthermore, it can be seen that this deviation increases monotonically for $\epsilon \in (0, 2\epsilon^*)$ and tends to infinity as $\epsilon \to 2\epsilon^*$. In order to gain intuitive insight into the impact of N and p on the limit in (5.23), we assume that ϵ is sufficiently small to approximate $g(\epsilon)$ using a first-order Taylor series expansion. Noting that $g(0) = 1$ and $g'(0) = ((N-1)/N)(1-p)$, in the vicinity of $\epsilon = 0$, the limit in (5.23) behaves as

$$\lim_{k\to\infty} \mathrm{MSE}(x(k)) \approx \frac{\sigma_0^2}{N}\left(1 + \frac{N-1}{N}(1-p)\epsilon\right).$$

This result shows that for small values of ϵ, the impact of N on the deviation of the asymptotic MSE with respect to σ_0^2/N becomes negligible after a relatively high number of nodes. On the other hand, the higher the probability of connection of the links is, the closer the asymptotic MSE is to the benchmark σ_0^2/N.

In summary, closed-forms expressions for the MSE($x(k)$) and for the asymptotic MSE($x(k)$) have been derived, as well as for the dynamic range and for the optimum link weights minimizing the convergence time in the mean square sense of the consensus algorithm in Erdős-Rényi random networks with instantaneous directed links. Finally, the impact of the number of nodes and the impact of the probability of connection on the asymptotic MSE($x(k)$) have been discussed. In the following section we present the simulations results for an Erdős-Rényi network, whereas in Section 5.6 we generalize the expression in (5.11) for the case of links with different probabilities of connection.

5.5.3 Numerical Results

The analytical results obtained in Section 5.5.1 and in Section 5.5.2 are supported with computer simulations of a WSN composed of $N = 20$ nodes randomly deployed in the unit square where the communication links are randomly generated with probability $p = 0.4$. The entries of the vector $\mathbf{x}(0)$ are modeled as Gaussian r.v.'s with mean $x_m = 20$ and variance $\sigma_0^2 = 5$. A total of 10.000 independent realizations were run to obtain the empirical MSE($x(k)$), where the position of the nodes and the connection probability matrix $\mathbf{P}$ are kept fixed for all the realizations, while a new Laplacian matrix is generated at each iteration.

Figure 5.1 shows the empirical MSE($x(k)$) defined in (5.7) in log-linear scale along with the theoretical closed-form expression in (5.16) (patterns) for three different cases:

(1) $\epsilon^* = 0.1094$ found using (5.20) (dotted line)
(2) $\epsilon = 1/(N-1) = 0.0526$ (dashed line)
(3) $\epsilon = 0.2 < 2\epsilon^*$ (dashed-dotted line)

The benchmark value $\sigma_0^2/N = 0.25$ is included in solid line. As expected, the empirical values obtained with (5.7) match the theoretical values obtained with (5.16). We observe

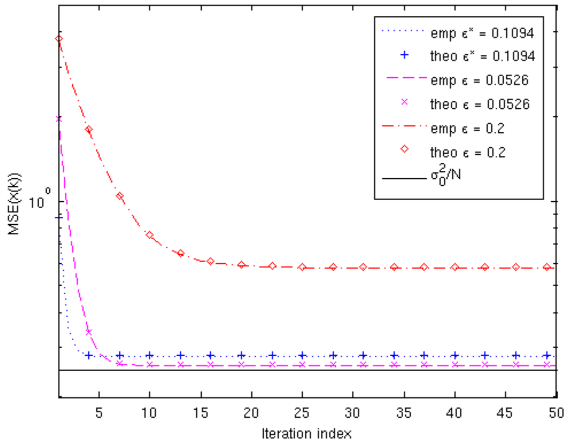


Figure 5.1: Empirical and theoretical MSE($x(k)$) as a function of k averaged over $N = 20$ nodes with $p = 0.4$ and different values of ϵ.

that the gap corresponding to the term $g(\epsilon)$ defined in (5.23) decreases with ϵ. However, choosing the optimum ϵ^* we achieve fastest convergence of the MSE($x(k)$) (dotted line) as stated by Lemma 5.1, whereas choosing the smallest ϵ the curve is closer to the benchmark (dashed line).

5.6 MSE Analysis for Links with Different Probabilities

In this section we generalize the results from Section 5.5 and analyze the MSE expression in (5.11) for the case of communication links having different probabilities of connection. At this point we make use of the results in [Fan94, Theorem 3], reproduced bellow:

Lemma 5.2. ***[Fan94] Inequality for the trace of matrix product -*** *For any matrix* $\mathbf{X} \in \mathbb{R}^{N\times N}$ *and any symmetric matrix* $\mathbf{Y} \in \mathbb{R}^{N\times N}$*, let* $\widehat{\mathbf{X}} = (\mathbf{X} + \mathbf{X}^T)/2$*. Then*

$$\mathrm{tr}(\mathbf{XY}) \leq \lambda_1(\widehat{\mathbf{X}})\mathrm{tr}(\mathbf{Y}) - \lambda_N(\mathbf{Y})\left(N\lambda_1(\widehat{\mathbf{X}}) - \mathrm{tr}(\mathbf{X})\right) \tag{5.24}$$

where $\lambda_1(.)$ *and* $\lambda_N(.)$ *denote largest and smallest eigenvalue respectively.*

Using Lemma 5.2 we can resume an important result in the following theorem:

Theorem 5.3. *Consider the consensus algorithm in (5.1) with* N *nodes and weight matrix defined in (5.2), symmetric nonnegative connection probability matrix* $\mathbf{P}$ *and i.i.d. initial values* $\mathbf{x}(0)$ *with mean* x_m *and variance* σ_0^2*. Assuming that the largest eigenvalue of* $\mathbf{R_w}(k)$ *in (5.9) is equal to* 1 $\forall k > 0$*, the MSE*$(x(k))$ *in (5.7) is upper bounded by*

$$MSE(x(k)) \leq \frac{\sigma_0^2}{N}\left(N + (\mathrm{tr}(\mathbf{C_w}) - N)\,\frac{1-\lambda_N^k(\mathbf{C_w})}{1-\lambda_N(\mathbf{C_w})}\right) \tag{5.25}$$

where $\mathbf{C_w}$ *is the matrix defined in (5.14) and* $\lambda_N(\mathbf{C_w})$ *denotes its smallest eigenvalue.*

Proof. Applying Lemma 5.2, the trace term in (5.11) is upper bounded by

$$\mathrm{tr}(\mathbf{R_w}(k)) \leq \mathrm{tr}(\mathbf{C_w}) - \lambda_N(\mathbf{C_w})N + \lambda_N(\mathbf{C_w})\mathrm{tr}(\mathbf{R_w}(k-1))$$

where we have substituted for $\mathbf{X} = \mathbf{R_w}(k-1)$ and $\mathbf{Y} = \mathbf{C_w}$ in (5.24), and assumed that $\lambda_1(\mathbf{R_w}(k-1)) = 1$. Replacing $\mathrm{tr}(\mathbf{R_w}(k-1))$ for $\mathrm{tr}(\mathbf{R_w}(k-2)\cdot\mathbf{C_w})$ above, and computing the upper bound recursively until reaching $\mathbf{R_w}(1)$ we obtain

$$\mathrm{tr}(\mathbf{R_w}(k)) \leq \mathrm{tr}(\mathbf{C_w}) + \mathrm{tr}(\mathbf{C_w})\sum_{i=1}^{k-2}\lambda_N^i(\mathbf{C_w}) + \lambda_N^{k-1}(\mathbf{C_w})\mathrm{tr}(\mathbf{R_w}(1)) - N\sum_{i=1}^{k-1}\lambda_N^i(\mathbf{C_w}).$$

This inequality can be further simplified to

$$\begin{aligned}\mathrm{tr}(\mathbf{R_w}(k)) &\leq \mathrm{tr}(\mathbf{C_w}) + (\mathrm{tr}(\mathbf{C_w}) - N)\sum_{i=1}^{k-1}\lambda_N^i(\mathbf{C_w}) \\ &= N + (\mathrm{tr}(\mathbf{C_w}) - N)\sum_{i=0}^{k-1}\lambda_N^i(\mathbf{C_w})\end{aligned}$$

and replacing for the trace in (5.11) the proof is completed.

□

Theorem 5.3 states that, for a known connection probability matrix $\mathbf{P}$, we can find $\mathbf{C_w}$ using (5.14), compute its eigenvalues and then compute the upper bound for the MSE$(x(k))$ in (5.25) for *any* time instant. It can be checked that when the link probabilities are all equal, the value of the upper bound in (5.25) coincides with the exact expression in (5.16). This is due to the particular structure of $\mathbf{C_w}$ when substituting $p_{ij}=p \; \forall\{i,j\}$.

Theorem 5.3 provides an upper bound for the estimation error of the consensus algorithm whenever the matrix $\mathbf{R_w}(k)$ has largest eigenvalue equal to 1. This is a strong condition and in the following subsection we give sufficient conditions on the value of ϵ to guarantee that $\mathbf{R_w}(k)$ has largest eigenvalue $\lambda_1(\mathbf{R_w}(k))=1 \; \forall k>0$ and that the MSE upper bound in (5.25) converges.

5.6.1 Asymptotic MSE Upper Bound and Optimum Link Weights

We observe that the term $\lambda_N^k(\mathbf{C_w})$ on the right-hand side of the inequality in (5.25) tends to zero as k increases and therefore, the upper bound for the MSE in (5.25) converges whenever $|\lambda_N(\mathbf{C_w})|<1$. In order to analyze the asymptotic behavior of the upper bound, we make use of the following Lemma:

Lemma 5.3. *Consider the matrix $\mathbf{C_w}$ defined in (5.14) with symmetric connection probability matrix $\mathbf{P}$. If $\epsilon \in (0, 1/(N-1)]$, the smallest eigenvalue of $\mathbf{C_w}$ satisfies*

$$0 \leq \lambda_N(\mathbf{C_w}) < 1, \quad \forall\, 0 \leq p_{ij} \leq 1\,. \tag{5.26}$$

Proof. The left inequality in (5.26) holds because $\mathbf{C_w}$ in (5.14) is a real, symmetric and positive semi-definite matrix, and therefore its eigenvalues are all real and nonnegative. To prove the right inequality in (5.26) we have that

$$\mathrm{tr}(\mathbf{C_w}) = \sum_{i=1}^{N} \lambda_i(\mathbf{C_w}) \geq N\lambda_N(\mathbf{C_w})$$

therefore, the smallest eigenvalue of $\mathbf{C_w}$ is upper bounded by

$$\begin{aligned} \lambda_N(\mathbf{C_w}) &\leq \frac{1}{N}\sum_{i=1}^{N}[\mathbf{C_w}]_{ii} \\ &= 1 - \frac{2\epsilon}{N}\sum_{i=1}^{N}\sum_{l=1}^{N} p_{il} + \frac{\epsilon^2}{N}\left(\sum_{i=1}^{N}\sum_{l=1}^{N} p_{il}\left(2 + \sum_{\substack{m=1\\ m\neq l}}^{N} p_{im}\right)\right) \end{aligned} \tag{5.27}$$

where we have replaced for the diagonal entries of $\mathbf{C_w}$. For simplicity, we denote the expression on the right-hand side of (5.27) as the function $h(\epsilon)$. This function satisfies $h(0) = 1$ and, after some algebraic manipulations we have

$$\begin{aligned} h\left(\frac{1}{N-1}\right) &= 1 + \frac{1}{(N-1)^2} \times \sum_{i=1}^{N}\sum_{l=1}^{N} p_{il}\left(-2(N-2) + \sum_{\substack{m=1\\m\neq l}}^{N} p_{im}\right) \\ &\leq 1 - \frac{N-2}{N(N-1)^2}\sum_{i=1}^{N}\sum_{l=1}^{N} p_{il} \\ &= 1 - \frac{N-2}{N-1} < 1 \end{aligned}$$

where the first inequality holds because $\sum_{\substack{m=1\\m\neq l}}^{N} p_{im} \leq (N-2)$ and the second inequality holds because we assume $p_{ij} > 0$ for any pair $\{i,j\} \in \mathcal{V}$. Since $h(\epsilon)$ above is a convex function on ϵ, we have that

$$h(\epsilon) < 1, \quad \forall \epsilon \in \left(0, \frac{1}{N-1}\right]. \tag{5.28}$$

Substituting (5.28) in (5.27) we prove the right inequality of (5.26).

□

Applying the results of Lemma 5.3, the asymptotic upper bound in (5.25) is

$$\lim_{k\to\infty} \mathrm{MSE}(\mathbf{x}(k)) \leq \sigma_0^2 \left(\frac{\mathrm{tr}(\mathbf{C_w}) - N\lambda_N(\mathbf{C_w})}{1-\lambda_N(\mathbf{C_w})}\right) \text{ for } \epsilon \in (0, 1/(N-1)].$$

The expression above gives an upper bound for the error after the algorithm has converged, and this upper bound can be computed off-line, since the matrix $\mathbf{C_w}$ depends on the probability matrix $\mathbf{P}$ and on the constant ϵ.

Lemma 5.3 shows that if we choose ϵ in the interval in (5.28), the smallest eigenvalue of $\mathbf{C_w}$ is less than one in magnitude, so the term $\lambda_N^k(\mathbf{C_w})$ tends to zero as $k \to \infty$. Note that this dynamic range also guarantees that the weight matrices $\{\mathbf{W}(k), \forall k \geq 0\}$ in (5.1) are nonnegative with positive diagonal entries. In that case, it can be shown that the matrix $\mathbf{R_w}(k)$ is nonnegative for all $k > 0$, and due to Corollary 8.1.30 in [Hor06], it has largest eigenvalue one with algebraic multiplicity one. Thus, the upper bound for the MSE$(x(k))$ in (5.25) applies and converges for $\epsilon \in (0, 1/(N-1)]$ and in addition, $\mathbf{x}(k)$ *converges almost surely* to a consensus by the results in [TS08].

In general, the optimum ϵ^* minimizing $\lambda_N(\mathbf{C_w})$ and therefore minimizing the convergence time of the MSE upper bound in (5.25) is greater than the upper limit $1/(N-1)$, which further guarantees that $\lambda_1(\mathbf{R_w}(k)) = 1$ with algebraic multiplicity one. However, we have observed that the upper bound still converges for values of ϵ exceeding the upper limit in (5.28). This might happen because the matrix $\mathbf{R_w}(k)$ still has largest eigenvalue equal to 1.

Since a smaller $\lambda_N(\mathbf{C_w})$ would lead to a faster convergence of the upper bound, and since $\lambda_N(\mathbf{C_w})$ is not convex on ϵ, we propose to select the value of ϵ that minimizes $\lambda_N(\mathbf{C_w})$ using an exhaustive search in a closed interval. As we will observe in the simulation results, the choice of ϵ minimizing $\lambda_N(\mathbf{C_w})$ provides convergence of the empirical MSE$(x(k))$ and in addition, fast convergence of the algorithm.

5.6.2 Numerical Results

The analytical results obtained in Section 5.6 are supported with computer simulations. Analogously to the previous set of simulations, we consider a WSN with N nodes randomly deployed in the unit square where the communication links are randomly generated. The entries of the vector $\mathbf{x}(0)$ are modeled as Gaussian r.v.'s with mean $x_m = 20$ and variance $\sigma_0^2 = 5$. A total of 10.000 independent realizations were run to obtain the empirical MSE$(x(k))$, where a new Laplacian matrix is generated at each iteration.

First, we simulate a small-world network [Wat98] with $N = 20$, 4 nearest neighbors and shortcut probability 0.4. The non-zero entries of the matrix $\mathbf{P}$ are set equal to $p = 0.4$. Three different values of ϵ were tested for the deployment, all of them satisfying $\lambda_N(\mathbf{C_w}) < 1$. For the small-world network, we obtained

(1) $\epsilon^* = 0.1781$ (solid line)
(2) $\epsilon = 0.0526$ (dashed line)
(3) $\epsilon = 0.36 \approx 2\epsilon^*$ (dashed-dotted line)

Figure 5.2 shows the empirical MSE$(x(k))$ in log-linear scale for the small-world network and the theoretical upper bound computed with (5.25) depicted with patterns. The benchmark value $\sigma_0^2/N = 0.25$ is included in solid line. We observe that although the upper

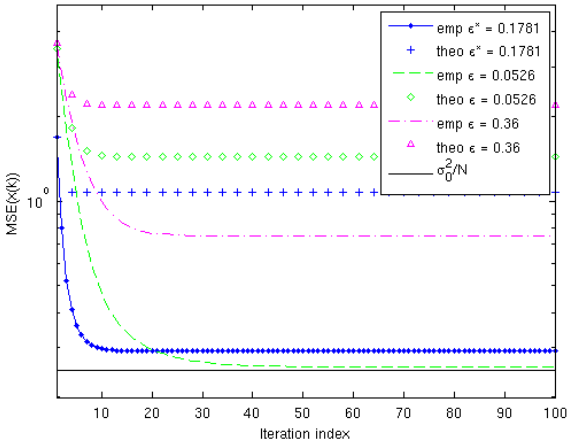


Figure 5.2: Empirical MSE$(x(k))$ and theoretical upper bound for a small-world network with $N = 20$, 4 nearest neighbors, connection probability 0.4 and different values of ϵ.

bound curves for the two smallest values of ϵ converge quite fast, the curve for ϵ^* converges even faster. We observe also that the curves for the empirical MSE$(x(k))$ behave rather similar to the upper bound curves in terms of the convergence time, since the empirical MSE$(x(k))$ shows a faster convergence also when the optimum ϵ^* is chosen.

Finally, we simulate two random geometric networks[2] with $N = 20$ and $N = 100$ nodes respectively. In this case, the non-zero entries of the matrix $\mathbf{P}$ are modeled as i.i.d. r.v.'s uniformly distributed between 0 and 1. The optimum link weights minimizing $\lambda_N(\mathbf{C_w})$ are $\epsilon^* = 0.0740$ and $\epsilon^* = 0.0170$ respectively (dotted lines), found using exhaustive searches over closed intervals of positive values of ϵ. The remaining choices are $\epsilon = 1/(N-1)$ (dashed-lines) and $\epsilon \approx 2\epsilon^*$ (dashed-dotted line) as for the previous case. Figures 5.3 and 5.4 show the empirical MSE$(x(k))$ for the networks composed of 20 and 100 nodes respectively, where the theoretical upper bound computed with (5.25) is depicted with patterns and the benchmark values $\sigma_0^2/N = 0.25$ and $\sigma_0^2/N = 0.05$ respectively are included in solid line.

Analogously to the small-world case, the upper bound curves for ϵ^* in Fig. 5.3 and Fig. 5.4 converge faster than the curves for the two remaining cases, although the difference

[2] We refer here to a random geometric graph as defined in Chapter 2.

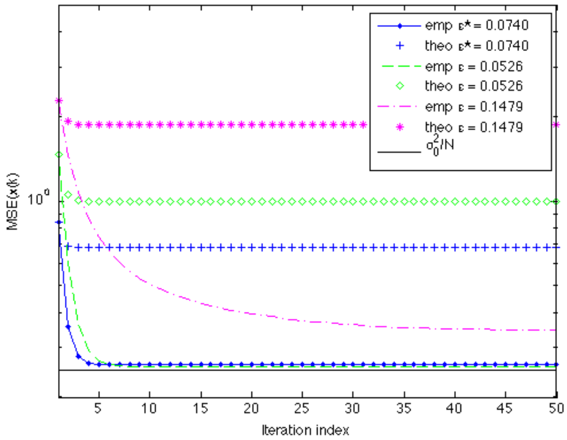


Figure 5.3: Empirical MSE($x(k)$) and theoretical upper bound as a function of k averaged over $N = 20$ nodes for different probabilities of connection and different values of ϵ.

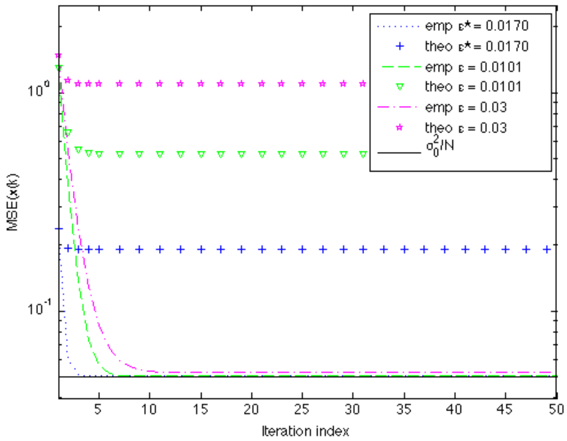


Figure 5.4: Empirical MSE($x(k)$) and theoretical upper bound as a function of k averaged over $N = 100$ nodes for different probabilities of connection and different values of ϵ.

is less remarkable. Again, the empirical MSE($x(k)$) curves behave similarly to the upper bound curves in terms of the convergence time, obtaining a faster convergence when the optimum ϵ* is chosen.

The simulations show that all the values of ϵ for which $\lambda_N(\mathbf{C_w}) < 1$ is satisfied, guarantee the convergence of the upper bound but not necessarily the convergence of the empirical MSE($x(k)$). For instance, with $N = 20$ and $\epsilon = 0.2$ -no figures included- the upper bound converges but the MSE($x(k)$) diverges. This might be due to the fact that the eigenvalue equal to 1 of the matrix $\mathbf{R_w}(k)$ is no longer the largest, and therefore Theorem 5.3 does not be apply in this case. This problem is further studied in Chapter 6 where we study almost sure convergence of the consensus algorithm in networks with random topologies where the communication links might exhibit spatial correlation. In the following section we propose a practical implementation using random transmission power at each iteration to reduce the overall power consumption of the network.

5.7 Application: Randomized Transmission Power Network

In this section, we present a heuristic scheme of randomized power transmission intended to reduce the energy consumption of the network until reaching a consensus. After convergence, we assume that the nodes switch to a save-energy mode in order to extend the network lifetime. In the proposed scheme, the nodes transmit at each time instant using different power levels selected at random from a predefined range of values and independently of the rest of the nodes, such that the total amount of energy consumed by the network is balanced among the nodes. In other words, the average power used is the same for all the nodes. The transmission power at each node varies therefore with time and is, in general, different among nodes at the same time instant.

The randomized transmission power scheme establishes random links between nodes with different probabilities of connection, which results in an instantaneous random directed topology. In order to minimize the convergence time of the algorithm, we use the results from the analysis of the MSE($x(k)$) in Section 5.6. As we will see with computer simulations, the overall energy consumption of the network until convergence of the consensus algorithm is strongly reduced with respect to a fixed symmetric topology that spends the same average transmission power.

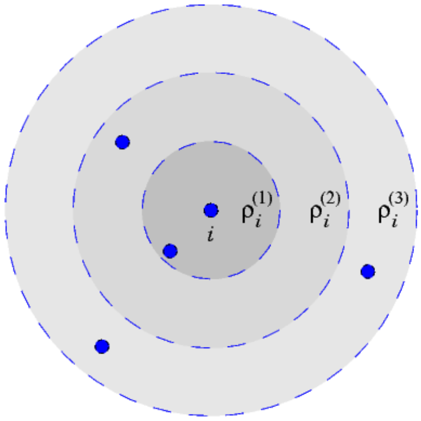


Figure 5.5: Example network with $N = 5$ nodes and $|R| = 3$ power levels. The connectivity radii for node i are $\rho_i^{(1)}$, $\rho_i^{(2)}$ and $\rho_i^{(3)}$.

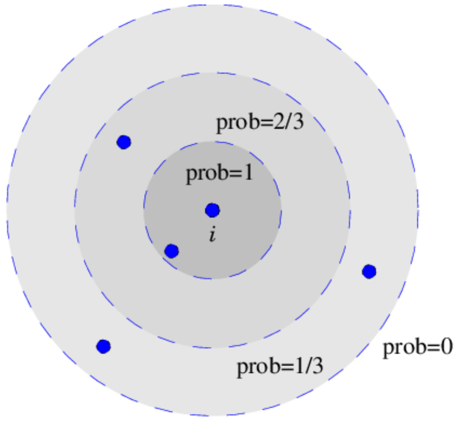


Figure 5.6: Map of probabilities of receiving a packet from node i depending on distance $d(i, j)$ and on the connectivity radii.

5.7.1 Network Model

We assume a set of nodes randomly deployed in a given area. At each iteration, each node transmits using a power level randomly selected from a predefined set of values, and independently from the rest of the nodes. These power levels describe different concentric circles of connectivity, centered at the transmitting node, and with a radius denoted $\rho_i(k)$ for node i at time k, proportional to the square root of the associated transmit power level.

Without loss of generality, we define the set containing all the possible radii, arranged in increasing order of magnitude as follows

$$\mathcal{R} = \{\rho^{(1)}, \rho^{(2)}, \cdots, \rho^{(R)}\} \tag{5.29}$$

where $\rho^{(1)}$ is the connectivity radius associated with the minimum power level, $\rho^{(R)}$ is the connectivity radius associated with the maximum power level and R is the total number of radii in $\mathcal{R}$. Consider the distance between two nodes $\{i, j\} \in \mathcal{V}$ defined as $d(i, j)$ and let the connectivity radius for node i at time k belong to the set of radii, i.e., $\rho_i(k) \in \mathcal{R}$. For instance, if $d(i, j) < \rho_i(k)$ is satisfied at iteration k, then node j receives the information from node i at time k. An example deployment of 5 nodes and 3 different transmission power levels, with transmitting node i is depicted in Fig. 5.5.

For simplicity, we assume that the power levels in (5.29) have all the same probability of begin chosen, equal to

$$\pi = \frac{1}{R}.$$

Then, the nodes located inside the first circle centered at node i will receive the information from node i with probability 1, while the nodes lying only in the last annulus receive the information from node i with probability π. The nodes outside the outer circle receive information with probability 0. The connection probability matrix has entries $\mathbf{P}_{ij} = p_{ij}$ with $p_{ij} = p_{ji}$, given by

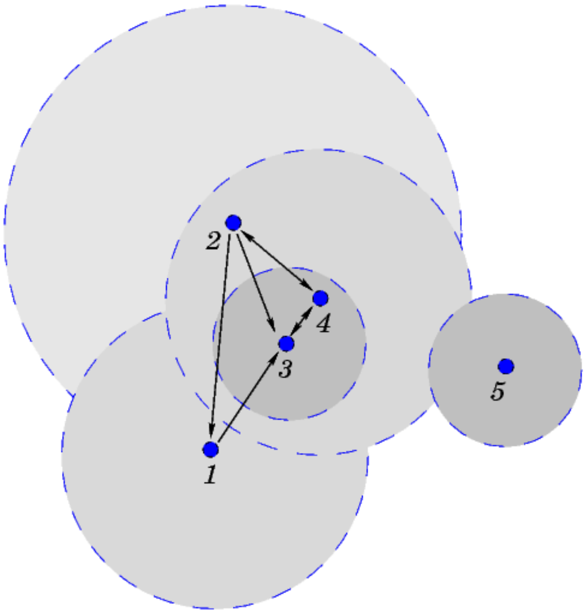
$$\mathbf{P}_{ij} = \begin{cases} (R-l+1)\pi & \text{if } \rho^{(l-1)} < d(i,j) \leq \rho^{(l)}, \quad l = 1, \cdots, R \\ 0 & \text{otherwise} \end{cases} \tag{5.30}$$

for all pairs $\{i, j\} \in \mathcal{V}$ with $\rho^{(0)} = 0$. The corresponding map of probabilities for the example network in Fig. 5.5 is depicted in Fig. 5.6.

The set of neighbors for every node and consequently the corresponding adjacency matrix $\mathbf{A}(k)$, varies randomly from iteration to iteration depending on the instantaneous choice of power levels. $\mathbf{A}(k)$ is clearly non-symmetric and random with symmetric mean $\bar{\mathbf{A}} = \mathbf{P}$. An example of a random realization along with the corresponding instantaneous adjacency matrix is depicted in Fig. 5.7.

Consider the consensus algorithm in (5.1) with weight matrix given in (5.2) and instantaneous connectivity radii $\rho_i(k) \in \mathcal{R}$ for each node. Due to the random nature of $\mathbf{A}(k)$, $\mathbf{L}(k)$ is random and in general non-symmetric[3]. Therefore, the weight matrices $\{\mathbf{W}(k), \forall k \geq 0\}$ are by construction random, temporally independent of each other, non-symmetric and satisfy $\mathbf{W}(k)\mathbf{1} = \mathbf{1}$. Assuming again that the network is connected in expectation, $\bar{\mathbf{L}}$ is irreducible and so is the expected weight matrix $\bar{\mathbf{W}}$. Since our aim is to minimize the overall energy consumption of the network, we aim at finding the ϵ that minimizes the convergence time of the state vector $\mathbf{x}(k)$ and the dynamic range of ϵ guaranteeing mean square convergence of the consensus algorithm implementing the randomized transmission power scheme. Because of the lack of a closed form expression

[3] Note that for this model the entries of $\mathbf{L}(k)$ are spatially correlated among each other.



$$\mathbf{A}(k) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 5.7: Example of a resulting instantaneous topology and its corresponding adjacency matrix.

for the MSE$(x(k))$, we use the theoretical results derived in Section 5.6 to characterize the convergence time of the algorithm, particularly we compute the upper bound derived in Theorem 5.3 for different probabilities of connection for the links.

5.7.2 Convergence of the MSE Upper Bound

Recall that the upper bound for the MSE in (5.25) converges if and only if $\lambda_N(\mathbf{C_w}) < 1$, and the convergence time of the upper bound decreases as $\lambda_N(\mathbf{C_w}) \to 0$. Using (5.12) and (5.13), the matrix $\mathbf{C_w}$ can be rewritten as follows

$$\mathbf{C_w} = \mathbf{I} - 2\epsilon\bar{\mathbf{L}} + \epsilon^2\boldsymbol{\Upsilon} \tag{5.31}$$

where $\boldsymbol{\Upsilon}$ has entries given by

$$\boldsymbol{\Upsilon}_{ii} = 2\sum_{l=1}^{N} p_{il} + \sum_{l=1}^{N}\sum_{\substack{m=1\\ m\neq l}}^{N} p_{il}p_{im}$$

$$\boldsymbol{\Upsilon}_{ij} = -p_{ij}\left(\sum_{l=1}^{N} p_{il} + \sum_{l=1}^{N} p_{jl}\right) + \sum_{\substack{l=1\\ l\neq j}}^{N} \Phi_l,\ i \neq j$$

with

$$\Phi_l = \begin{cases} p_{il} & \text{if } \rho^{(R)} > d(i,l) > d(j,l) \\ p_{jl} & \text{if } \rho^{(R)} > d(j,l) \geq d(i,l) \end{cases}, \forall l. \tag{5.32}$$

Since $\mathbf{C_w}$ in (5.31) depends on the matrix $\mathbf{P}$, it can be computed off-line whenever we have knowledge of the probabilities of connection. For the randomized transmission power model, the matrix $\mathbf{P}$ can be derived using (5.30) when the node locations -which we assume fixed- and the set of power levels $\mathcal{R}$ in (5.29) are known. Then, using an exhaustive search over all values of ϵ in a given interval we can choose the ϵ that minimizes the magnitude of $\lambda_N(\mathbf{C_w})$.

5.7.3 Performance Evaluation

In this section we evaluate the performance of the consensus algorithm in terms of convergence time and energy consumption in a network implementing the randomized transmission power scheme. We compare the performance of the proposed scheme with two networks with fixed symmetric topology where the nodes transmit using constant transmission power.

For that purpose, we simulate a WSN composed of $N = 100$ nodes uniformly deployed in a squared area of dimensions 100×100, where each node measurement is modeled as an independent Gaussian r.v. with mean $x_m = -4$ and variance $\sigma_0^2 = 25$. We evaluate the performance of the randomized power network with several power levels and applying uniform link weights ϵ^*, where ϵ^* is the link value that minimizes $\lambda_N(\mathbf{C_w})$. The instantaneous radius of connectivity for each node, denoted hereafter as ρ_{var}, takes one of the values in the set

$$\mathcal{R} = \{5, 10, 15, 30\}.$$

The fixed topology networks use constant transmission power and optimum link weights derived in [Xia03], given by

$$\epsilon_{opt} = \frac{2}{\lambda_2(\mathbf{L}) + \lambda_N(\mathbf{L})}$$

where $\mathbf{L}$ is the fixed Laplacian matrix. In the first model implemented for comparison, the transmission power is equal to the average power of the randomized network, such that the connectivity radius for each node is given by

$$\rho_{ave} = \sqrt{\frac{1}{R}\sum_{l=1}^{R}\left(\rho^{(l)}\right)^2} \tag{5.33}$$

and in this case equal to $\rho_{ave} \simeq 17.67$. Note that this network will spend the same transmission power as the randomized power network on average over time.

In the second model, the transmission power is equal to the maximum power of the randomized network, such that the connectivity radius for each node is equal to

$$\rho_{max} = \rho^{(R)} \tag{5.34}$$

and in this case equal to $\rho_{max} = 30$. Due to the dimensions of the simulated network and the power levels assumed, we are close to a fully connected network when all the nodes transmit using ρ_{max}. Note that this case is interesting because in general, a faster convergence of the consensus algorithm is attained when the network is almost fully connected. However, this scenario does not guarantee a minimum energy consumption until convergence when all the nodes transmit at the same power level, as we will observe from the simulation results.

A total of 10.000 realizations of $\mathbf{x}(0)$ were simulated to obtain the empirical MSE$(x(k))$. Fig. 5.8 shows the empirical MSE$(x(k))$ plotted in log-linear scale as a function of the iteration index for the three cases:

(1) the fixed network with radius ρ_{max} and $\epsilon_{opt} = 0.0576$ (line-dots)
(2) the fixed network with radius ρ_{ave} and $\epsilon_{opt} = 0.1221$(dashed line)
(3) the randomized power network with $\epsilon^* = 0.1$ (dashed-dotted line)

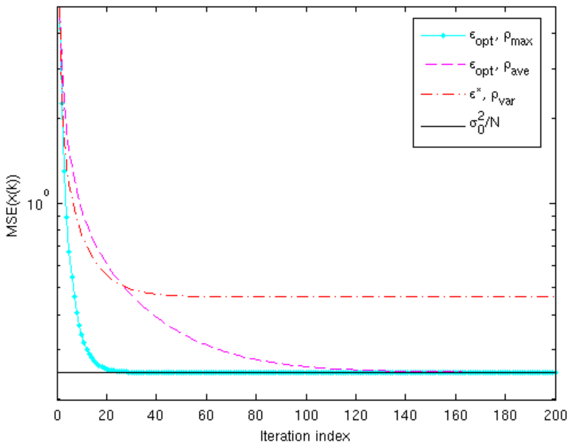


Figure 5.8: Empirical MSE($x(k)$) as a function of k for a network with randomized power and variable radius of connectivity ρ_{var}, and for fixed topology networks with fixed power and radii ρ_{ave} and ρ_{max}.

The benchmark value $\sigma_0^2/N = 0.25$ is included in solid line. Convergence has been considered reached when the difference between subsequent states is less than $2 \cdot 10^{-5}$. As expected, the curves for the fixed topology networks reach the benchmark, since $\mathbf{L}$ is in both cases time-invariant and symmetric. First, we observe that the curve for the randomized power network converges faster than the curve for the average power case (ca. 84 and 164 iterations respectively), although it stays over due to the non-symmetry of $\mathbf{L}(k)$. This is due to the fact that at some time instants, a given node i can communicate with a node j far away (with maximum distance $d(i, j) < \rho^{(R)}$), increasing the instantaneous connectivity of the network.

Comparing the performance of the network using maximum transmission power with the performance of the network with randomized transmission power, we observe that the former converges faster, i.e., 40 vs 84 iterations. However, the overall energy consumption in the fixed network is greater than in the randomized power one, since the energy spent by every node is proportional to the term $K \times \rho_{max}^2$, where K denotes the convergence time. The results for this and for other combinations of power levels are included in Table 5.1, showing the approximate number of iterations to reach consensus and the consumed power by each node until reaching a consensus in terms of the quantity $K \times \text{radius}^2$. Despite numerical differences, in all cases the network with randomized transmission power

Connectivity radii	Iterations (K) with:			Energy consumption $\propto$		
R & (ρ_{ave})	ρ_{max}	ρ_{var}	ρ_{ave}	$K \cdot \rho_{max}^2$	$K \cdot \rho_{var}^2$	$K \cdot \rho_{ave}^2$
$\{5, 10, 15, 30\}$ (17.7)	40	84	164	36000	26250	51250
$\{5, 10, 15, 50\}$ (26.7)	16	25	73	40000	17812	52012
$\{10, 20, 30, 60\}$ (35.3)	12	22	29	43200	27500	36250
$\{10, 20, 70\}$ (42.4)	9	13	21	44100	23400	37800

Table 5.1: Number of iterations for several combinations of power levels.

converges faster than the fixed network with average transmission power.

These results show that using the randomized transmission power scheme, the convergence time of the consensus algorithm can be improved at the same overall power consumption. This means that the energy consumption of the network to reach consensus is lowered, although at the cost of a reduction in the accuracy of the estimation. In addition, the value of ϵ^* that minimizes the convergence time of the upper bound seems to be a good choice to reduce the convergence time of the empirical MSE($x(k)$).

5.8 Conclusions of the Chapter

This chapter has shown that consensus can be reached in the mean square sense in a WSN with random topology and instantaneous asymmetric links, and that the mean average consensus can be reached in expectation. Moreover, the MSE of the state vector can be characterized analytically with knowledge of the probability of connection of the links and the statistics of the initial set of measurements.

For the case of links with equal probability of connection, closed-forms expressions for the MSE($x(k)$) and for the asymptotic MSE($x(k)$) have been derived, as well as for the dynamic range of the link weight that guarantees mean square convergence of the consensus algorithm. A closed-form expression for the optimum link weights providing maximum convergence rate has been derived and the impact of the number of nodes as well as the impact of the probability of connection on the asymptotic MSE($x(k)$) have been discussed. The MSE expression derived proves to be useful to characterize the convergence time of the consensus algorithm.

For the case of links with different probabilities of connection, an upper bound for the MSE of the state has been derived and its convergence and asymptotic behavior have been studied. Although the upper bound differs from the empirical MSE($x(k)$), it can be employed for the computation of a link weight that reduces the convergence time of the consensus algorithm under these connectivity conditions.

Additionally, a practical transmission scheme intended to extend the lifetime of a WSN running consensus algorithms has been proposed. With the proposed scheme, the nodes transmit using different power levels at every time instant, selected independently of other nodes and with equal probability. The computer simulations show that the convergence time is reduced with respect to a fixed topology where the nodes transmit using the same average power. Therefore, the total energy required to reach a consensus is reduced and the network lifetime can be lengthened.

The randomized transmission power scheme is also more energy efficient than a fully connected network. The price to be paid for reducing the energy consumption is a reduction in the accuracy of the estimation. The preference of improving either the accuracy of the estimation or the energy consumption can be determined by the final application.

6

Almost Sure Consensus in Random Networks

6.1 Introduction

Most contributions found in literature studying the convergence of consensus algorithms in random networks assume spatially independent communication links, that is, the existence of a link between a given pair of nodes is completely independent of the existence of a link between another pair of nodes. Communication links may be however spatially correlated not only due to an intrinsic correlation of the channels among the nodes, but also due to the communication protocol. An example is the pair-wise gossip algorithm, which is a particular case of consensus in correlated random networks with undirected links, where at each time instant a bidirectional communication link is established between two nodes selected at random. Another example is the broadcast gossip algorithm, a particular case of consensus in correlated random networks with directed links where at each iteration a single node selected at random broadcasts its value to all its neighbors, creating directional communication links. A further example is the randomized transmission power scheme presented in Chapter 5 -Section 5.7- where at each iteration the nodes use different power levels randomly chosen from a predefined set of values. Since the existence of a link depends on the instantaneous transmission power applied and the distance between two nodes, the resulting instantaneous topology is random directed and the links are spatially correlated.

Several results for gossip algorithms can be generalized for the consensus algorithm, e.g. [Boy06, Pic07, Fag08, Ays09]. Fagnani and Zampieri [Fag08] studies the asymptotic convergence rate of randomized consensus schemes based on either mean square convergence or on Lyapunov exponents. Based on [Boy06] and [Fag08], Aysal *et al.* [Ays09] derives a sufficient condition for almost sure convergence of the broadcast gossip algorithm. Furthermore, [Ays10] provides a formulation which includes sum-preserving, non sum-preserving, quantized and noisy gossip algorithms, and derives sufficient conditions for almost sure convergence as well as an asymptotic upper bound on the mean square performance. Regarding consensus algorithms in correlated topologies, Jakovetić *et al.* [Jak10] shows that the problem of assigning the optimum weights is a convex optimization problem. For undirected networks, the optimization criterion is the minimization of the MSE computed with respect to the average of the initial values, while for directed networks it is the minimization of the mean square deviation with respect to the instantaneous state average. Moreover, Abaid and Porfiri [Aba10] focuses on numerosity-constrained directed networks, i.e., n-regular topologies where all the nodes have out-degree n, and derives closed-form expressions for the asymptotic convergence factor as a function of n, denoted as the numerosity factor.

In this chapter, the convergence of the consensus algorithm in WSNs assuming instantaneous directed and instantaneous undirected topologies is studied, where the links are allowed to be spatially correlated. Whereas [Fag08] and [Ays09] focus specifically on gossip and [Aba10] restricts the nodes to have a fixed out-degree, we consider the model of consensus with uniform weights as presented in Chapter 3 -Section 3.4.2- not necessarily restricted to symmetric and nonnegative instantaneous weight matrices, where we use the results in [Zho09] to show almost sure convergence to the agreement space. As we will see, almost sure convergence to a consensus can be related to the spectral radius of a positive semidefinite matrix for which we derive closed-form expressions. We consider the minimization of this spectral radius as the optimization criterion to reduce the convergence time of the algorithm, and show that the value of the optimum link weights can be obtained as the solution of a convex optimization problem. In addition to show convexity of the optimization problem, we show that the closed-form expressions derived are useful to compute the dynamic range of the link weights guaranteeing almost sure convergence. Further, we show that our general formulation subsumes known protocols found in literature and derive additional closed-form expressions for the optimum parameters for particular cases of links with equal probability of connection. For simplicity, "random networks with directed topology" refers to networks with random directed instantaneous links, whereas "random networks with undirected topology" refers to networks with random undirected instantaneous links.

Outline of the Chapter

The chapter is organized as follows. Section 6.2 presents the network model with spatial correlation among links, and in Section 6.3 convergence in expectation to the average consensus in random networks is shown, considering both directed and undirected instantaneous links. An optimization criterion to reduce the convergence time which sets a sufficient condition for almost sure convergence is presented in Section 6.4, whereas in Section 6.5 convexity of the optimization problem is shown, as well as the existence of an optimum solution. In Section 6.6 closed-form expressions useful for the minimization of the convergence time are derived for undirected and for directed random networks. These closed-form expressions are further validated in Section 6.7, where theoretical results for known existing protocols found in literature as well as additional closed-form expressions for further particular cases are derived. Simulation results for three different scenarios, namely a random geometric expected network[1], a small-world network and a randomized transmission power network are included in Section 6.8, whereas the conclusions for the chapter are included in Section 6.9.

6.2 Network and Spatial Correlation Models

Consider a WSN composed of N nodes with a topology characterized by a random graph $\mathcal{G}(k) = \{\mathcal{V}, \mathcal{E}(k)\}$ where a communication link e_{ij} between two nodes $\{i, j\}$ exist with a given probability $0 \leq p_{ij} \leq 1$. Assuming a connection probability matrix with entries $\mathbf{P}_{ij} = p_{ij}$ where $p_{ij} = p_{ji}$ for all $\{i, j\} \in \mathcal{V}$, the instantaneous adjacency matrix $\mathbf{A}(k)$ is

[1] The random geometric network refers to a random geometric graph as defined in Chapter 2.

random with expected value matrix $\mathbb{E}[\mathbf{A}(k)] = \bar{\mathbf{A}} = \mathbf{P}$. The entries of the adjacency matrix $\mathbf{A}(k)$ are temporally independent but might be spatially correlated, and the information about the spatial correlation among pairs of links $\{e_{ij}, e_{qr}\}$ is arranged in the matrix $\mathbf{C} \in \mathbb{R}^{N^2 \times N^2}$, with $\{st\}^{th}$ entry given by

$$\mathbf{C}_{st} = \begin{cases} 0, & s = t \\ \mathbb{E}[a_{ij}a_{qr}] - p_{ij}p_{qr}, & s \neq t \end{cases} \quad \text{for} \quad \begin{cases} s = i + (j-1)N \\ t = q + (r-1)N \end{cases} \tag{6.1}$$

where $a_{ij} = [\mathbf{A}(k)]_{ij}$, and the iteration indexing is omitted since correlation is assumed time invariant. That is, the off-diagonal entries of $\mathbf{C}$ are the covariance between the links e_{ij} and e_{qr} for the nodes $\{i, j, q, r\}$, whereas the entries of the main diagonal are set equal to 0. Note that $\mathbf{C}$ is not a covariance matrix but it will be useful later in Section 6.6 to derive closed-form expressions in the convergence analysis.

We analyze the convergence of the state vector $\mathbf{x}(k)$ given by

$$\mathbf{x}(k) = \mathbf{W}(k-1)\mathbf{x}(k-1), \quad k > 0, \tag{6.2}$$

where $\mathbf{x}(0)$ is the vector of initial measurements and the weight matrix $\mathbf{W}(k) \in \mathbb{R}^{N \times N}$ is modeled as

$$\mathbf{W}(k) = \mathbf{I} - \epsilon\mathbf{L}(k), \quad k \geq 0 \tag{6.3}$$

with random Laplacian $\mathbf{L}(k) = \mathbf{D}(k) - \mathbf{A}(k)$ and constant $\epsilon > 0$, equal for all the links. The matrices $\{\mathbf{W}(k), \forall k \geq 0\}$ in (6.2) are random and independent of each other with one eigenvalue equal to 1 and associated right eigenvector $\mathbf{1}$. For undirected instantaneous topologies, $\mathbf{L}(k)$ is symmetric and $\mathbf{W}(k)$ satisfies:

$$\begin{array}{rcl} \mathbf{W}(k)\mathbf{1} & = & \mathbf{1} \\ \mathbf{1}^T\mathbf{W}(k) & = & \mathbf{1}^T \end{array}, \quad \forall k \geq 0 \tag{6.4}$$

whereas for directed instantaneous topologies, $\mathbf{L}(k)$ is non-symmetric and $\mathbf{W}(k)$ satisfies:

$$\begin{array}{rcl} \mathbf{W}(k)\mathbf{1} & = & \mathbf{1} \\ \mathbf{1}^T\mathbf{W}(k) & \neq & \mathbf{1}^T \end{array}, \quad \forall k \geq 0 \tag{6.5}$$

i.e., the left eigenvector associated with the eigenvalue 1 may vary randomly from realization to realization. The expected weight matrix $\bar{\mathbf{W}}$ is assumed to satisfy the following properties

$$\begin{aligned} \bar{\mathbf{W}}\mathbf{1} &= \mathbf{1} \\ \mathbf{1}^T\bar{\mathbf{W}} &= \mathbf{1}^T \\ \rho(\bar{\mathbf{W}} - \mathbf{J}_N) &< 1 \end{aligned} \tag{6.6}$$

where $\mathbf{J}_N = \mathbf{1}\mathbf{1}^T/N$ is the normalized all-ones matrix, in both the instantaneous undirected and the instantaneous directed case. In other words, the weight matrices $\{\mathbf{W}(k), \forall k \geq 0\}$ are in expectation balanced with only one largest eigenvalue equal to one. Using (6.3), the expected weight matrix is $\bar{\mathbf{W}} = \mathbf{I} - \epsilon\bar{\mathbf{L}}$. The matrix $\bar{\mathbf{L}}$ can be seen as the Laplacian of an undirected graph, defined as the expected graph $\mathbb{E}[\mathcal{G}(k)] = \bar{\mathcal{G}}$, and is symmetric due to the assumption of a symmetric connection probability matrix $\mathbf{P}$. We assume that $\bar{\mathcal{G}}$ is connected such that $\lambda_N(\bar{\mathbf{L}}) = 0$ with algebraic multiplicity one. $\bar{\mathbf{W}}$ is therefore symmetric by construction and double-stochastic for $\epsilon \in (0, 2/\lambda_1(\bar{\mathbf{L}})]$, and satisfies the convergence conditions in (6.6). These conditions will be important for the derivation of closed-form results later in Section 6.6.

6.3 Convergence in Expectation to the Average Consensus

It is not difficult to see that due to the conditions on the expected weight matrix in (6.6), the estimation in (6.2) is asymptotically unbiased and the consensus value $c \in \mathbb{R}$ is, in expectation, equal to the average consensus, since

$$\begin{aligned} \mathbb{E}[\mathbf{x}(k)] &= \mathbb{E}\left[\mathbf{W}(k-1)\mathbf{x}(k-1)\right] \\ &= \mathbb{E}\left[\prod_{l=1}^{k} \mathbf{W}(k-l)\mathbf{x}(0)\right] \\ &= \bar{\mathbf{W}}^k\mathbf{x}(0) \end{aligned}$$

where in the last equality the assumption of temporally independent weight matrices is used. Taking the limit of the expression above yields

$$\begin{aligned} \lim_{k\to\infty} \mathbb{E}[\mathbf{x}(k)] &= \mathbf{J}_N\mathbf{x}(0) \\ &= \mathbf{x}_{ave}. \end{aligned}$$

In order to minimize the rate at which the network reaches a consensus, in the following section we present a criterion for minimizing the convergence time which establishes a sufficient condition for almost sure convergence under these topology conditions.

6.4 Almost Sure Convergence to a Consensus

We say that the entries of the state vector $\mathbf{x}(k)$ in (6.2) converge *almost surely* (*with probability one*) to a consensus value c if

$$\Pr\left\{\lim_{k\to\infty}\mathbf{x}(k)=c\mathbf{1}\right\}=1.$$

In this section, we study the convergence of the random vector $\mathbf{x}(k)$ to the agreement space $\mathcal{A}=span\{\mathbf{1}\}$ using notions of stochastic stability. Using the results in [Zho09], we say that almost sure consensus is achieved if for all $\delta>0$

$$\lim_{k_0\to\infty}\Pr\left\{\sup_{k>k_0}\inf_{\mathbf{x}_a\in\mathcal{A}}\|\mathbf{x}(k)-\mathbf{x}_a\|_2^2>\delta\right\}=0. \tag{6.7}$$

Note that the vector $\mathbf{x}_a$ in $\mathcal{A}$ minimizing the norm above is given by

$$\inf_{\mathbf{x}_a\in\mathcal{A}}\|\mathbf{x}(k)-\mathbf{x}_a\|_2^2=\|\mathbf{x}(k)-\bar{\mathbf{x}}(k)\|_2^2,\quad\forall k$$

where $\bar{\mathbf{x}}(k)=\mathbf{J}_N\mathbf{x}(k)$ is the orthogonal projection of $\mathbf{x}(k)$ onto the agreement space $\mathcal{A}$. We define the deviation vector at time k as

$$\begin{aligned}\mathbf{d}(k)&=\mathbf{x}(k)-\bar{\mathbf{x}}(k)\\&=(\mathbf{I}-\mathbf{J}_N)\mathbf{x}(k)\end{aligned} \tag{6.8}$$

which specifies the distance to the average $\bar{\mathbf{x}}(k)$, i.e., $\mathbf{d}(k)$ specifies how far the nodes are from a consensus at time k. In other words, the expression in (6.7) is equivalent to

$$\Pr\left\{\sup_{k>k_0}\|\mathbf{d}(k)\|_2^2\right\}\to 0. \tag{6.9}$$

The discrete-time consensus algorithm in (6.2) asymptotically reaches almost sure consensus if and only if the equilibrium point $\mathbf{0}$ is almost surely asymptotically stable for the error vector $\mathbf{d}(k)$, and therefore we analyze the evolution of $\mathbb{E}\left[\|\mathbf{d}(k)\|_2^2\right]$. Remark that in the undirected case, the algorithm is sum-preserving and $\mathbb{E}\left[\|\mathbf{d}(k)\|_2^2\right]=\mathbb{E}\left[\|\mathbf{x}(k)-\mathbf{x}_{ave}\|_2^2\right]$ is the expected error with respect to the average consensus. Using the fact that the row-sums of $\mathbf{W}(k)$ are equal to 1, we have that

$$\begin{aligned}\mathbf{d}(k+1)&=(\mathbf{I}-\mathbf{J}_N)\mathbf{W}(k)\mathbf{x}(k)\\&=(\mathbf{I}-\mathbf{J}_N)\mathbf{W}(k)(\mathbf{I}-\mathbf{J}_N)\mathbf{x}(k)\\&=(\mathbf{I}-\mathbf{J}_N)\mathbf{W}(k)\mathbf{d}(k).\end{aligned}$$

For convenience, let's define the matrix $\mathbf{\Omega}(k) = (\mathbf{I} - \mathbf{J}_N)\mathbf{W}(k)$, and remark that [Pap65]

$$\mathbb{E}\left[\left\|\mathbf{d}(k)\right\|_2^2\right] = \mathbb{E}_{\mathbf{d}(k-1)}\left[\mathbb{E}_{\mathbf{d}(k)}\left[\left\|\mathbf{d}(k)\right\|_2^2 \middle| \mathbf{d}(k-1)\right]\right]. \tag{6.10}$$

The expected squared norm of $\mathbf{d}(k)$ given $\mathbf{d}(k-1)$ is given by

$$\begin{aligned}\mathbb{E}\left[\left\|\mathbf{d}(k)\right\|_2^2 \middle| \mathbf{d}(k-1)\right] &= \mathbf{d}(k-1)^T \mathbb{E}\left[\mathbf{\Omega}(k)^T \mathbf{\Omega}(k)\right] \mathbf{d}(k-1) \\ &= \mathbf{d}(k-1)^T \mathcal{W} \mathbf{d}(k-1) \\ &\leq \lambda_1(\mathcal{W}) \left\|\mathbf{d}(k-1)\right\|_2^2 \end{aligned} \tag{6.11}$$

where in the last inequality we have used the fact that for any vector $\mathbf{u}$ of unit norm yields $\mathbf{u}^T \mathbf{X} \mathbf{u} \leq \lambda_1(\mathbf{X}) \mathbf{u}^T \mathbf{u}$ [Hor06, Theorem 4.2.2], and

$$\begin{aligned}\mathcal{W} &= \mathbb{E}\left[\mathbf{\Omega}(k)^T \mathbf{\Omega}(k)\right] \\ &= \mathbb{E}\left[\mathbf{W}(k)^T (\mathbf{I} - \mathbf{J}_N) \mathbf{W}(k)\right]\end{aligned} \tag{6.12}$$

where we have considered that $(\mathbf{I} - \mathbf{J}_N)$ is symmetric and idempotent. Therefore, we have that the expected squared norm of $\mathbf{d}(k)$ given $\mathbf{d}(k-1)$ is bounded above. Substituting (6.11) in (6.10) yields

$$\begin{aligned}\mathbb{E}\left[\left\|\mathbf{d}(k)\right\|_2^2\right] &\leq \mathbb{E}\left[\lambda_1(\mathcal{W}) \left\|\mathbf{d}(k-1)\right\|_2^2\right] \\ &= \lambda_1(\mathcal{W}) \mathbb{E}\left[\left\|\mathbf{d}(k-1)\right\|_2^2\right].\end{aligned} \tag{6.13}$$

Repeatedly conditioning and replacing iteratively for $\mathbf{d}(k)$ we obtain

$$\mathbb{E}\left[\left\|\mathbf{d}(k)\right\|_2^2\right] \leq \lambda_1^k(\mathcal{W}) \left\|\mathbf{d}(0)\right\|_2^2. \tag{6.14}$$

The right hand side in (6.14) is an upper bound for the expected square norm of the deviation vector, and this upper bound will converge as time evolves whenever

$$\lambda_1(\mathcal{W}) < 1. \tag{6.15}$$

Clearly, $\lambda_1(\mathcal{W})$ governs the rate at which the upper bound for the error decays to zero, where a smaller value of $\lambda_1(\mathcal{W})$ will result in a faster convergence of the upper bound in (6.14). The convergence time of the upper bound is then given by

$$\tau_{bound} = -\frac{1}{\log\left(\lambda_1(\mathcal{W})\right)}. \tag{6.16}$$

$\lambda_1(\mathcal{W})$ is denoted as the per-step mean square convergence factor in [Zho09], defined as

$$r_s = \sup_{\mathbf{d}(k) \neq \mathbf{0}} \frac{\mathbb{E}[\|\mathbf{d}(k+1)\|_2^2 | \mathbf{d}(k)]}{\|\mathbf{d}(k)\|_2^2}.$$

and is the stochastic equivalent of the per-step convergence factor defined in (3.26). From the results in [Zho09, Lemma 2(i)], the per-step mean square convergence factor r_s is greater than the asymptotic mean square convergence factor r_a, defined as

$$r_a = \sup_{\mathbf{d}(k) \neq \mathbf{0}} \lim_{k \to \infty} \left(\frac{\mathbb{E}[\|\mathbf{d}(k)\|_2^2 | \mathbf{d}(0)]}{\|\mathbf{d}(0)\|_2^2} \right)^{1/k}$$

which considers instead the distance to the average consensus, and is the stochastic equivalent of the asymptotic convergence factor in (3.25). A necessary and sufficient condition for mean square stability, which is only sufficient for almost sure asymptotic stability [Fen92][2], is that $r_a < 1$. Since $r_s > r_a$, $r_s < 1$ is sufficient to ensure asymptotic almost sure stability. Therefore, according to [Zho09, Lemma 2(ii)], (6.15) is sufficient to ensure asymptotic almost sure stability and implies almost sure convergence of $\mathbf{x}(k)$ in (6.2) to the agreement space $\mathcal{A}$. The minimization of $\lambda_1(\mathcal{W})$, which further minimizes the convergence time of the upper bound in (6.14), is the optimization criterion chosen to reduce the convergence time of the consensus algorithm in random networks while ensuring almost sure convergence to a consensus. Consequently, we focus on finding the value of ϵ in (6.3) that minimizes the value of $\lambda_1(\mathcal{W})$.

6.5 Minimizing the Convergence Time of the Upper Bound

According to the results in Section 6.4, a sufficient condition for almost sure consensus is that (6.15) is satisfied. Note that after some matrix manipulations, the matrix $\mathcal{W}$ defined in (6.12) can be expressed as

$$\mathcal{W} = \mathbb{E}\left[\mathbf{L}(k)^T (\mathbf{I} - \mathbf{J}_N) \mathbf{L}(k)\right] \epsilon^2 - 2\bar{\mathbf{L}}\epsilon + \mathbf{I} - \mathbf{J}_N \tag{6.17}$$

where we have replaced for (6.3) and used the fact that $\bar{\mathbf{L}}$ has $\mathbf{1}$ as both the left and the right eigenvectors associated with $\lambda_N(\bar{\mathbf{L}}) = 0$. The matrix $\mathcal{W}$ is positive semidefinite

[2] In general, necessary and sufficient conditions for stochastic stability can be derived using the results in [Hib96] and [Ben97].

by definition with row-sums equal to 1 and nonnegative real eigenvalues. Since $\mathcal{W}$ is symmetric, it satisfies also $\mathbf{1}^T\mathcal{W} = \mathbf{1}^T$.

The minimization of $\lambda_1(\mathcal{W})$ with respect to ϵ is a convex optimization problem for both undirected and directed instantaneous topologies, as shown in the following theorem.

Theorem 6.1. *Consider the consensus algorithm in (6.2) with spatially correlated random links, $\mathbf{W}(k)$ defined in (6.3) and satisfying $\mathbf{W}(k)\mathbf{1} = \mathbf{1}$, and $\mathcal{W}$ defined in (6.12) with largest eigenvalue $\lambda_1(\mathcal{W})$. The minimization problem*

$$\begin{aligned} \underset{\epsilon}{min} \quad & f(\epsilon) = \lambda_1(\mathcal{W}) \\ s.t. \quad & \epsilon \geq 0, \quad \mathcal{W} \in \mathbb{S}^+ \end{aligned} \tag{6.18}$$

is convex on ϵ.

Proof. Let the objective function be expressed as $f(\epsilon) = h(G(\epsilon))$, $\mathbf{dom}\ f = \{\epsilon \in \mathbf{dom}\ G | G(\epsilon) \in \mathbf{dom}\ h\}$ where $h : \mathbb{R}^{N\times N} \to \mathbb{R}$ denotes the maximum eigenvalue function with domain $\mathbf{dom}\ h = \mathbb{S}$ and the function $\mathcal{W} = G : \mathbb{R} \to \mathbb{R}^{N\times N}$ is given by

$$G(\epsilon) = \boldsymbol{\Gamma}\epsilon^2 + 2\boldsymbol{\Delta}\epsilon + \boldsymbol{\Theta} \tag{6.19}$$

with $\boldsymbol{\Gamma} = \mathbb{E}\left[\mathbf{L}(k)^T(\mathbf{I} - \mathbf{J}_N)\mathbf{L}(k)\right]$, $\boldsymbol{\Delta} = -\bar{\mathbf{L}}$ and $\boldsymbol{\Theta} = \mathbf{I} - \mathbf{J}_N$, coinciding with (6.17). We show that the composition $f = h \circ G$ is convex, and we do so in three steps: **a)** show convexity of G, **b)** show matrix monotonicity of h and **c)** show convexity of the composition $f = h \circ G$.

a) To show convexity of $G(\epsilon)$ we start showing that $\boldsymbol{\Gamma} \in \mathbb{S}^+$, since for any non-zero unitary vector $\{\mathbf{v} \in \mathbb{R}^{N\times 1}, \|\mathbf{v}\| = 1\}$ we have

$$\mathbf{v}^T\mathbb{E}\left[\mathbf{L}(k)^T(\mathbf{I} - \mathbf{J}_N)\mathbf{L}(k)\right]\mathbf{v} = \mathbb{E}\left[\mathbf{v}^T\mathbf{L}(k)^T(\mathbf{I} - \mathbf{J}_N)^T(\mathbf{I} - \mathbf{J}_N)\mathbf{L}(k)\mathbf{v}\right] \geq 0$$

Further, observe that for all $x, y \in \mathbf{dom}\ G$ we have

$$\boldsymbol{\Gamma}(\tau x + (1-\tau)y)^2 \preceq \boldsymbol{\Gamma}(\tau x^2 + (1-\tau)y^2)$$

where $\mathbf{A} \preceq \mathbf{B}$ means that $\mathbf{B} - \mathbf{A} \in \mathbb{S}^+$. Adding $2\boldsymbol{\Delta}(\tau x + (1-\tau)y) + \boldsymbol{\Theta}$ on both sides yields

$$\begin{gathered} \boldsymbol{\Gamma}(\tau x + (1-\tau)y)^2 + 2\boldsymbol{\Delta}(\tau x + (1-\tau)y) + \boldsymbol{\Theta} \preceq \tau\boldsymbol{\Gamma}x^2 + (1-\tau)\boldsymbol{\Gamma}y^2 + \tau 2\boldsymbol{\Delta}x + (1-\tau)2\boldsymbol{\Delta}y + \boldsymbol{\Theta} \\ G(\tau x + (1-\tau)y) \preceq \tau G(x) + (1-\tau)G(y). \end{gathered}$$

b) Next, we analyze the function h. This function is matrix convex [Boy04] and we will show that it is also matrix monotone:

A function $h : \mathbb{R}^{N\times N} \rightarrow \mathbb{R}$ *is matrix monotone with respect to the set* $\mathbb{S}$ *if for any pair* $\mathbf{X}, \mathbf{Y} \in \mathbb{S}$, $\mathbf{X} \preceq \mathbf{Y}$ *yields* $h(\mathbf{X}) \leq h(\mathbf{Y})$.

Since the maximum eigenvalue of $\mathbf{X}$ can be seen as the point-wise supremum of a family of linear functions of $\mathbf{X}$, we have $\lambda_1(\mathbf{X}) = \sup\{\mathbf{u}^T\mathbf{X}\mathbf{u} \mid \|\mathbf{u}\|_2 = 1\}$. Let's denote by $\mathbf{u}$ and $\mathbf{v}$ the eigenvectors associated with the largest eigenvalue of $\mathbf{X}$ and $\mathbf{Y}$ respectively. If $\mathbf{X} \preceq \mathbf{Y}$, for any $\mathbf{u} \in \mathbb{R}^{N\times 1} \Rightarrow \mathbf{u}^T\mathbf{X}\mathbf{u} \leq \mathbf{u}^T\mathbf{Y}\mathbf{u}$. Moreover, if $\mathbf{u}$ is the eigenvector associated with $\lambda_1(\mathbf{X})$ we have $\lambda_1(\mathbf{X}) = \mathbf{u}^T\mathbf{X}\mathbf{u} \leq \mathbf{u}^T\mathbf{Y}\mathbf{u} \leq \lambda_1(\mathbf{Y})$ and we can conclude that $\lambda_1(\mathbf{X}) \leq \lambda_1(\mathbf{Y})$.

c) Combining the convexity of G from **a)** and the matrix monotonicity of h from **b)** yields

$$h(G(\tau x + (1-\tau)y)) \leq h(\tau G(x) + (1-\tau)G(y)) \tag{6.20}$$

and recalling the matrix convexity of h we have

$$h(\tau G(x) + (1-\tau)G(y)) \leq \tau h(G(x)) + (1-\tau)h(G(y)) \tag{6.21}$$

Combining the left-hand side of (6.20) and the right-hand side of (6.21) yields

$$f(\tau x + (1-\tau)) \leq \tau f(x) + (1-\tau)f(y)$$

which completes the proof. □

6.5.1 Dynamic Range for the Uniform Link Weights

Theorem 6.1 shows that the function $f(\epsilon)$ in (6.18) is convex. The next step consists in showing that the dynamic range of ϵ for which (6.15) is satisfied, exists. In fact, in this section we show that the minimization of $f(\epsilon)$ results in a value of $\lambda_1(\mathcal{W})$ satisfying the sufficient condition in (6.15) that guarantees almost sure consensus.

Using (6.19), the function $f(\epsilon)$ in (6.18) can be rewritten as follows

$$\begin{aligned} f(\epsilon) &= \sup_{\|\mathbf{u}\|_2^2=1} \{\mathbf{u}^T\boldsymbol{\Gamma}\mathbf{u}\epsilon^2 + 2\mathbf{u}^T\boldsymbol{\Delta}\mathbf{u}\epsilon + \mathbf{u}^T\boldsymbol{\Theta}\mathbf{u}\} \\ &= \sup_{\|\mathbf{u}\|_2^2=1} \{\gamma\epsilon^2 + 2\delta\epsilon + \theta\} \end{aligned} \tag{6.22}$$

where $\gamma = \mathbf{u}^T \boldsymbol{\Gamma} \mathbf{u}, \delta = \mathbf{u}^T \boldsymbol{\Delta} \mathbf{u}$ and $\theta = \mathbf{u}^T \boldsymbol{\Theta} \mathbf{u}$. Observe that $\boldsymbol{\Theta}$ has eigenvalues 0 with algebraic multiplicity one and 1 with algebraic multiplicity $N-1$. Thus, for $\epsilon = 0$ we have $\lambda_i(\mathcal{W}) = \lambda_i(\boldsymbol{\Theta})$ for all i with $\lambda_1(\mathcal{W}) = 1$. For values of ϵ in the proximity of zero, the quadratic term in (6.22) is negligible and $f(\epsilon)$ can be approximated as

$$f(\epsilon)\Big|_{\epsilon \sim 0} \approx \sup_{\|\mathbf{u}\|_2^2 = 1} \{2\delta\epsilon + \theta\}. \tag{6.23}$$

Since $\boldsymbol{\Theta}$ has eigenvalues 0 and 1, for all vectors $\mathbf{u}$ with $\|\mathbf{u}\|_2^2 = 1$ we have $0 \leq \theta \leq 1$, while since $\boldsymbol{\Delta} \in \mathbb{S}^-$, we have $\delta \leq 0$. Therefore, the value in (6.23) is always less than or equal to 1. Note however that $\theta = 1$ if and only if $\mathbf{u}$ is one eigenvector associated with the eigenvalue 1. Analogously, $\delta = 0$ if and only if $\mathbf{u}$ is the eigenvector associated with the eigenvalue 0, i.e., $\mathbf{u} = \mathbf{1}$. However, $\mathbf{1}$ is also the eigenvector associated with the eigenvalue 0 of $\boldsymbol{\Theta}$. Therefore, there is no $\mathbf{u}$ such that $\delta = 0$ and $\theta = 1$ at the same time, showing that the approximation in (6.23) is always below 1. On the other hand, as $\epsilon \to \infty$ the quadratic term in (6.22) becomes predominant and $\lambda_1(\mathcal{W}) \to \infty$. Combining these results with the fact that $f(\epsilon)$ is convex, we conclude that there exists an interval of positive values of ϵ for which (6.15) holds, ensuring therefore (6.9). In addition, the value of ϵ that minimizes $f(\epsilon)$ satisfies the sufficient condition for almost sure convergence in (6.15).

In summary, there exists a range of positive values of ϵ ensuring the convergence of the norm of the deviation vector to zero, and there exists a positive value of ϵ minimizing the convergence time of the upper bound in (6.14). Theorem 6.1 shows that the optimum ϵ is the solution of a convex optimization problem, and this value can be computed using the subgradient algorithm [Boy06, Jak10]. In order to find the optimum ϵ analytically we derive closed-form expressions for the matrix $\mathcal{W}$ for both the case of instantaneous undirected topologies and the case of instantaneous directed topologies with links exhibiting spatial correlation.

6.6 Derivation of Closed-Form Expressions for the Matrix $\mathcal{W}$

The next step consists in analyzing the matrix $\mathcal{W}$ where we focus on uniform link weights with matrices satisfying either (6.4) or (6.5) and a network connected in average over time whose instantaneous links are independent in time but correlated in space according to the model in (6.1). In both cases, we assume that the expected weight matrix satisfies the convergence conditions in (6.6). We derive closed-form expressions for $\mathcal{W}$ for undirected random networks in Section 6.6.1 and for directed random networks in Section 6.6.2.

6.6.1 Closed-form Expressions for $\mathcal{W}$ in Undirected Topologies

In this section we consider instantaneous symmetric links, i.e., weight matrices satisfying (6.4). When the links are bidirectional, the matrix $\mathcal{W}$ in (6.12) can be rewritten as follows

$$\mathcal{W} = \mathbb{E}[\mathbf{W}(k)^T\mathbf{W}(k)] - \mathbf{J}_N. \tag{6.24}$$

Theorem 6.2. *Consider the consensus algorithm in (6.2) with spatially correlated random links,* $\mathbf{W}(k)$ *defined in (6.3) and satisfying (6.4) and (6.6). The matrix* $\mathcal{W}$ *in (6.24) has a closed-form expression given by*

$$\mathcal{W} = \left(\bar{\mathbf{L}}^2 + 2(\bar{\mathbf{L}} - \tilde{\mathbf{L}}) + \mathbf{R}'\right)\epsilon^2 - 2\bar{\mathbf{L}}\epsilon + \mathbf{I} - \mathbf{J}_N \tag{6.25}$$

where $\bar{\mathbf{L}}$ *is the Laplacian of the expected underlying graph,*

$$\tilde{\mathbf{L}} = \tilde{\mathbf{D}} - \mathbf{P} \odot \mathbf{P} \tag{6.26}$$

$\tilde{\mathbf{D}}$ *is diagonal with* $\{ii\}^{th}$ *entry* $\tilde{\mathbf{D}}_{ii} = [(\mathbf{P} \odot \mathbf{P})\mathbf{1}]_i \ \forall i = \{1, \cdots, N\}$ *and connection probability matrix* $\mathbf{P}$*, and* $\mathbf{R}'$ *is an* $N \times N$ *symmetric matrix built with the covariance terms in* $\mathbf{C}$ *as follows*

$$\begin{aligned} \mathbf{R}'_{mm} &= \mathbf{g}_m^T \mathbf{C}\, \mathbf{g}_m \\ \mathbf{R}'_{mn} &= \textstyle\sum_{i=1}^N \mathbf{e}_{in}^T \mathbf{C}\, \mathbf{e}_{im} - \mathbf{g}_n^T \mathbf{C}\, \mathbf{e}_{nm} - \mathbf{g}_m^T \mathbf{C}\, \mathbf{e}_{mn} \end{aligned} \tag{6.27}$$

with

$$\mathbf{g}_m = \mathbf{1} \otimes \mathbf{e}_m, \quad \mathbf{e}_{mn} = \mathbf{e}_m \otimes \mathbf{e}_n, \tag{6.28}$$

where $\mathbf{e}_i$ *is the* i^{th} *column of* $\mathbf{I}$ *and* $\otimes$ *denotes Kronecker product.*

Proof. See Appendix 6.A. □

Note that the matrix $\tilde{\mathbf{L}}$ in (6.26) has the structure of a Laplacian whose non-diagonal entries are the squared entries of $\mathbf{P}$. In general, $\tilde{\mathbf{L}}$ is not diagonalized by the eigenvectors of $\bar{\mathbf{L}}$ and a closed-form expression for $\lambda_1(\mathcal{W})$ can not be derived, except for the special case of links with equal probability of connection, as we will see in Section 6.7. Moreover, the matrix $\mathbf{R}'$ has row-sums equal to 1, as it can be verified using the covariance sums in (6.75) and (6.76) from Appendix 6.A.

Corollary 6.3. *Consider the consensus algorithm in (6.2) with spatially uncorrelated random links, $\mathbf{W}(k)$ defined in (6.3) and satisfying (6.4) and (6.6). The matrix $\mathcal{W}$ in (6.24) has a closed-form expression given by*

$$\mathcal{W} = \left(\bar{\mathbf{L}}^2 + 2(\bar{\mathbf{L}} - \tilde{\mathbf{L}})\right)\epsilon^2 - 2\bar{\mathbf{L}}\epsilon + \mathbf{I} - \mathbf{J}_N \tag{6.29}$$

where $\bar{\mathbf{L}}$ is the Laplacian of the expected underlying graph and $\tilde{\mathbf{L}}$ is as defined in (6.26).

6.6.2 Closed-form Expressions for $\mathcal{W}$ in Directed Topologies

This section addresses the case of instantaneous asymmetric random links, i.e., weight matrices satisfying (6.5). For the case of links with directionality, we consider the matrix $\mathcal{W}$ defined in (6.12).

Theorem 6.4. *Consider the consensus algorithm in (6.2) with N nodes and spatially correlated random links, $\mathbf{W}(k)$ defined in (6.3) and satisfying (6.5) and (6.6). The matrix $\mathcal{W}$ in (6.12) has a closed-form expression given by*

$$\mathcal{W} = \left(\bar{\mathbf{L}}^2 + \frac{2(N-1)}{N}(\bar{\mathbf{L}} - \tilde{\mathbf{L}}) + \mathbf{R}\right)\epsilon^2 - 2\bar{\mathbf{L}}\epsilon + \mathbf{I} - \mathbf{J}_N \tag{6.30}$$

where $\bar{\mathbf{L}}$ is the Laplacian of the expected underlying graph, $\tilde{\mathbf{L}}$ is as defined in (6.26) and

$$\begin{aligned} \mathbf{R}_{mm} &= \mathbf{g}_m^T \mathbf{C}\, \mathbf{g}_m + \tfrac{1}{N}(\mathbf{g}_m^T - \mathbf{q}_m^T)\mathbf{C}\,(\mathbf{q}_m - \mathbf{g}_m) \\ \mathbf{R}_{mn} &= \textstyle\sum_{i=1}^N \mathbf{e}_{in}^T \mathbf{C}\, \mathbf{e}_{im} - \mathbf{g}_n^T \mathbf{C}\, \mathbf{e}_{nm} - \mathbf{g}_m^T \mathbf{C}\, \mathbf{e}_{mn} + \tfrac{1}{N}(\mathbf{g}_n^T - \mathbf{q}_n^T)\mathbf{C}\,(\mathbf{q}_m - \mathbf{g}_m) \end{aligned} \tag{6.31}$$

with

$$\mathbf{q}_m = \mathbf{e}_m \otimes \mathbf{1} \tag{6.32}$$

and $\mathbf{g}_m$, $\mathbf{e}_{mn}$ as defined in (6.28).

Proof. See Appendix 6.B. □

Remark that the matrix $\mathbf{R}'$ defined in (6.27) for the undirected case is different from the matrix $\mathbf{R}$ defined in (6.31) for the directed case. Analogously to the undirected case, the matrix $\mathbf{R}$ has row-sums equal one, as it can be verified using the covariance sums in (6.75) and (6.76) from Appendix 6.A, combined with the covariance sums in (6.80) and (6.81) from Appendix 6.B.

Corollary 6.5. *Consider the consensus algorithm in (6.2) with spatially uncorrelated random links, $\mathbf{W}(k)$ defined in (6.3) and satisfying (6.5) and (6.6). The matrix $\mathcal{W}$ in (6.12) has a closed-form expression given by*

$$\mathcal{W} = \left(\bar{\mathbf{L}}^2 + \frac{2(N-1)}{N}(\bar{\mathbf{L}} - \tilde{\mathbf{L}})\right)\epsilon^2 - 2\bar{\mathbf{L}}\epsilon + \mathbf{I} - \mathbf{J}_N \tag{6.33}$$

where $\bar{\mathbf{L}}$ is the Laplacian of the expected underlying graph and $\tilde{\mathbf{L}}$ is as defined in (6.26).

Corollary 6.3 and Corollary 6.5 follow from considering zero matrices $\mathbf{R}'$ and $\mathbf{R}$ when computing (6.25) and (6.30) respectively, which result from considering a zero matrix $\mathbf{C}$.

The closed-form expressions derived in this section are useful to find the optimum ϵ minimizing $\lambda_1(\mathcal{W})$ analytically. Furthermore, starting from the expressions in (6.25), (6.30) and (6.33), we are able to derive closed-form expressions for existing protocols found in literature, showing therefore that they can be seen as particular cases of the general expressions derived in this chapter.

6.7 Particularization of General Expressions to Special Cases

We particularize now the closed-form expressions derived in the previous section for the case of links with equal probability of connection. This is useful not only to further validate the main results of Section 6.6, but also to gain insight into the impact of the spatial correlation on the convergence time of the consensus algorithm. In Section 6.7.1, we particularize the general formulations for known protocols whose parameters have a closed-form expression, whereas in Section 6.7.2 we derive additional closed-form results for further particular cases.

6.7.1 Derivation of Expressions for Existing Protocols

We particularize the general formulations in (6.25), (6.30) and (6.33) to obtain four protocols found in literature: the broadcast gossip algorithm [Ays09], the pair-wise gossip algorithm [Boy06], random consensus in undirected networks with correlated links [Jak10] and random consensus in directed Erdős-Rényi networks with uncorrelated links [Sil09a], also included in Chapter 5 -Section 5.5-.

6.7.1.1 The Broadcast Gossip Algorithm

The broadcast gossip algorithm is a particular case of random consensus with spatially correlated links in directed topologies. In broadcast gossiping, a node wakes up randomly and broadcast its state to all its neighbors within its connectivity radius, where the probability of being activated is the same for all nodes and equal to $p = 1/N$. While standard gossip preserves the sum from iteration to iteration, the broadcast gossip algorithm converges to the average consensus only in expectation, since the connectivity radius is assumed equal for all nodes. The difference with the more general consensus model used in Section 6.6 is that, instead of having several nodes transmitting at the same time, in gossip algorithms only one node is transmitting at each time instant with probability p. Then, with probability p, the instantaneous weight matrix has $\{jk\}^{th}$ entry given by

$$[\mathbf{W}(k)]_{jk} = \begin{cases} 1, & j \notin \mathcal{N}_i, k = j \\ 1-\epsilon, & j \in \mathcal{N}_i, k = j \\ \epsilon, & j \in \mathcal{N}_i, k = i \\ 0, & \text{otherwise} \end{cases}$$

where $\epsilon \in (0,1)$. According to the results in [Ays09], the matrix $\mathcal{W}$ in (6.12) for the broadcast gossip algorithm, denoted hereafter $\mathcal{W}_{BG}$, is given by

$$\mathcal{W}_{BG} = \left(2\bar{\mathbf{L}} - \bar{\mathbf{L}}^2\right)\epsilon^2 - 2\bar{\mathbf{L}}\epsilon + \mathbf{I} - \mathbf{J}_N \tag{6.34}$$

with $\bar{\mathbf{L}} = p\mathbf{L}$ and $\bar{\mathbf{A}} = p\mathbf{A}$ where, adopting the notation in [Ays09], $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{L}$ are respectively the adjacency matrix, the degree matrix and the Laplacian matrix of the underlying fixed graph. We start defining the entries of the matrix $\mathbf{C}$ in (6.1), needed for the computation of $\mathbf{R}$ in (6.31).

Covariance terms Let $a_{ij} = [\mathbf{A}(k)]_{ij}$ as before and note that $\mathbb{E}[a_{ij}] = p$ for $j \in \mathcal{N}_i$ and for all k. Then,

$$\mathbb{E}[a_{ij}a_{qr}] = \begin{cases} p, & \text{if } q = i,\ r = j,\ j \in \mathcal{N}_i \\ p, & \text{if } r = j, q \neq i \text{ and } j \in \{\mathcal{N}_i \bigcap \mathcal{N}_q\} \\ 0, & \text{otherwise} \end{cases} \tag{6.35}$$

for all $\{i, j, q, r\} \in \{1, \cdots, N\}$. For notation clarity, we let $\mathbf{C}_{\substack{ij\\qr}}$ denote the $\{st\}^{th}$ entry of $\mathbf{C}$, with $\{s, t\}$ as defined in (6.1) for nodes $\{i, j, q, r\}$ such that

$$\mathbf{C}_{\substack{ij\\qr}} = \begin{cases} p(1-p) & \text{if } r = j,\ q \neq i \text{ and } j \in \{\mathcal{N}_i \bigcap \mathcal{N}_q\} \\ -p^2 & \text{if } r \neq j,\ j \in \mathcal{N}_i \text{ and } r \in \mathcal{N}_q \\ 0 & r = j \text{ and } q = i \\ 0 & \text{otherwise} \end{cases} \tag{6.36}$$

where we have used (6.35) and considered that the diagonal entries of $\mathbf{C}$ are equal to 0 by definition. A closed-form expression for $\mathbf{R}$ is given in the following Lemma:

Lemma 6.1. *Consider the broadcast gossip algorithm with N nodes and $\mathbf{C}$ as defined in (6.36). The matrix of correlation terms in (6.31) has a closed-form expression given by*

$$\mathbf{R} = 2\tilde{\mathbf{L}} - \bar{\mathbf{L}}^2 + \frac{1}{N}\left(2(\bar{\mathbf{L}} - \tilde{\mathbf{L}}) - p\mathbf{L}^2\right). \tag{6.37}$$

Proof. Consider the covariance summations in (6.75) and (6.76) from Appendix 6.A and (6.80) and (6.81) from Appendix 6.B for the diagonal and the non-diagonal entries of $\mathbf{R}$ respectively, and let $\mathbf{R}$ be expressed as

$$\mathbf{R} = \mathbf{R}_{diag} + \mathbf{R}_{off}$$

where $\mathbf{R}_{diag}$ is diagonal and $\mathbf{R}_{off}$ has zero entries on the main diagonal. Since we assume $\mathbf{C}_{\substack{ii\\qr}} = \mathbf{C}_{\substack{ij\\qq}} = \mathbf{C}_{\substack{ij\\ij}} = 0$ for all $\{i, j, q, r\}$, we observe that the $\{mm\}^{th}$ entry of $\mathbf{R}$ is given by

$$\mathbf{R}_{mm} = \sum_i \sum_j \mathbf{C}_{\substack{mi\\mj}} + \frac{1}{N}\left(-\sum_i \sum_j \mathbf{C}_{\substack{mi\\mj}} + \sum_i \sum_j \mathbf{C}_{\substack{im\\mj}} + \sum_i \sum_j \mathbf{C}_{\substack{mi\\jm}} - \sum_i \sum_j \mathbf{C}_{\substack{im\\jm}}\right)$$

whereas the non-diagonal $\{mn\}^{th}$ entry is given by

$$\mathbf{R}_{mn} = \sum_i \mathbf{C}_{\substack{im\\in}} - \sum_i \mathbf{C}_{\substack{mi\\mn}} - \sum_i \mathbf{C}_{\substack{ni\\nm}} + \frac{1}{N}\left(\sum_i\sum_j \mathbf{C}_{\substack{im\\nj}} - \sum_i\sum_j \mathbf{C}_{\substack{im\\jn}} - \sum_i\sum_j \mathbf{C}_{\substack{mi\\nj}} + \sum_i\sum_j \mathbf{C}_{\substack{mi\\jn}}\right)$$

where the sums are from $\{i,j\} = 1$ to N. Now, for notation simplicity let $\mathbf{D}_m$ denote the $\{mm\}^{th}$ entry of the degree matrix of the fixed topology and observe that the summations for the diagonal entries can be expressed as

$$\begin{array}{rclrcl} \sum_i\sum_j \mathbf{C}_{\substack{mi\\mj}} &=& -p^2\mathbf{D}_m(\mathbf{D}_m-1) & \sum_i\sum_j \mathbf{C}_{\substack{im\\mj}} &=& -p^2\mathbf{D}_m^2 \\ \sum_i\sum_j \mathbf{C}_{\substack{im\\jm}} &=& p(1-p)\mathbf{D}_m(\mathbf{D}_m-1) & \sum_i\sum_j \mathbf{C}_{\substack{mi\\jm}} &=& -p^2\mathbf{D}_m^2 \end{array}$$

whereas the summations for the non-diagonal can be expressed as

$$\begin{array}{rclrcl} \sum_i \mathbf{C}_{\substack{im\\in}} &=& -p^2(\mathcal{N}_m \bigcap \mathcal{N}_n) = -p^2[\mathbf{A}^2-\mathbf{D}]_{mn} & \sum_i \mathbf{C}_{\substack{mi\\mn}} &=& -p^2\mathbf{A}_{mn}(\mathbf{D}_m-1) \\ \sum_i \mathbf{C}_{\substack{ni\\nm}} &=& -p^2\mathbf{A}_{mn}(\mathbf{D}_n-1) & \sum_i\sum_j \mathbf{C}_{\substack{im\\jn}} &=& -p^2\mathbf{D}_m\mathbf{D}_n \end{array}$$

$$\begin{aligned} \sum_i\sum_j \mathbf{C}_{\substack{mi\\nj}} &= -p^2\mathbf{D}_m\mathbf{D}_n - p^2(\mathcal{N}_m \bigcap \mathcal{N}_n) + p(1-p)(\mathcal{N}_m \bigcap \mathcal{N}_n) \\ \sum_i\sum_j \mathbf{C}_{\substack{mi\\jn}} &= -p^2(\mathbf{D}_m\mathbf{D}_n - \mathbf{A}_{mn}\mathbf{D}_n) + p(1-p)\mathbf{A}_{mn}(\mathbf{D}_n-1) \\ \sum_i\sum_j \mathbf{C}_{\substack{im\\nj}} &= -p^2(\mathbf{D}_m\mathbf{D}_n - \mathbf{D}_m\mathbf{A}_{mn}) + p(1-p)\mathbf{A}_{mn}(\mathbf{D}_m-1). \end{aligned}$$

Then, $\mathbf{R}_{diag}$ can be computed as follows

$$\begin{aligned} \mathbf{R}_{diag} &= -p^2(\mathbf{D}^2-\mathbf{D}) + \frac{1}{N}\left(p^2(\mathbf{D}^2-\mathbf{D}) - 2p^2\mathbf{D}^2 - p(1-p)(\mathbf{D}^2-\mathbf{D})\right) \\ &= p^2(\mathbf{D}-\mathbf{D}^2) + \frac{1}{N}\left(p(\mathbf{D}-\mathbf{D}^2) - 2p^2\mathbf{D}\right) \end{aligned} \tag{6.38}$$

whereas $\mathbf{R}_{off}$ can be computing gathering together the terms multiplying $-p^2$ and the terms multiplying $p(1-p)$ as follows

$$\begin{aligned} \mathbf{R}_{off} &= -p^2(\mathbf{A}^2-\mathbf{D}-\mathbf{AD}-\mathbf{DA}+2\mathbf{A}) + \frac{1}{N}\left(-p^2(\mathbf{A}^2-\mathbf{D}-\mathbf{AD}-\mathbf{DA})\right) \\ &\quad + \frac{1}{N}p(1-p)(\mathbf{AD}+\mathbf{DA}-2\mathbf{A}-\mathbf{A}^2+\mathbf{D}) \\ &= p^2(-\mathbf{A}^2+\mathbf{D}+\mathbf{AD}+\mathbf{DA}-2\mathbf{A}) + \frac{1}{N}\left(p(\mathbf{AD}+\mathbf{DA}-2\mathbf{A}-\mathbf{A}^2+\mathbf{D}) + 2p^2\mathbf{A}\right) \end{aligned} \tag{6.39}$$

where an extra $\mathbf{D}$ is added to compensate for the contribution of $\mathbf{A}^2$ on the main diagonal. Finally, adding (6.38) and (6.39) and substituting for $\bar{\mathbf{L}} = p(\mathbf{D}-\mathbf{A})$ and $\tilde{\mathbf{L}} = p\bar{\mathbf{L}}$, we obtain the closed-form expression for $\mathbf{R}$ in (6.37), completing the proof. □

Replacing (6.37) in (6.30) we obtain the expression for $\mathcal{W}_{BG}$ in (6.34), and we can conclude that the closed-form results for the $\mathcal{W}$ matrix derived in [Ays09] can be obtained using the general formulation in (6.30). The eigenvalues of $\mathcal{W}_{BG}$ can be expressed as functions of the eigenvalues of $\bar{\mathbf{L}}$ as follows

$$f_i(\epsilon) = \left(2\lambda_i(\bar{\mathbf{L}}) - \lambda_i(\bar{\mathbf{L}})^2\right)\epsilon^2 - 2\lambda_i(\bar{\mathbf{L}})\epsilon + 1, \quad \forall i \in \{1, \cdots, N-1\}$$

where we have replaced $\lambda_i(\mathbf{I} - \mathbf{J}_N) = 1$ and considered that $f_N(\epsilon) = 0$. Note that these are quadratic functions of ϵ where the subindex i is in one to one correspondence with the ordering of the eigenvalues of $\bar{\mathbf{L}}$ but not with the ordering of the eigenvalues of $\mathcal{W}_{BG}$. This is due to the fact that, as the eigenvalues of $\bar{\mathbf{L}}$ increase in magnitude, the terms on ϵ increase in magnitude as well and the quadratic curves become narrower with a vertex approaching the abscissa while moving away from the origin. Since at $\epsilon = 0$, $f_i(\epsilon)$ is equal to one for all $i \in \{1, \cdots, N-1\}$ and the curve experiencing the slowest decay is the one corresponding to $i = N-1$, the largest eigenvalue of $\mathcal{W}_{BG}$ is given by[3]

$$f_{N-1}(\epsilon) = \left(2\lambda_{N-1}(\bar{\mathbf{L}}) - \lambda_{N-1}(\bar{\mathbf{L}})^2\right)\epsilon^2 - 2\lambda_{N-1}(\bar{\mathbf{L}})\epsilon + 1. \tag{6.40}$$

The optimum ϵ is therefore the value minimizing the function in (6.40), and is given by

$$\epsilon^* = \frac{1}{2 - \lambda_{N-1}(\bar{\mathbf{L}})}.$$

Replacing for $\epsilon = 1 - \gamma$, $\bar{\mathbf{L}}$, and p we obtain the optimum mixing parameter γ^* derived in [Ays09, Corollary 1]. A closed-form expression for $\lambda_1(\mathcal{W}_{BG})$ is obtained substituting for ϵ^* in (6.40) and yields

$$\lambda_1(\mathcal{W}_{BG})\big|_{\epsilon^*} = \frac{2(1 - \lambda_{N-1}(\bar{\mathbf{L}}))}{2 - \lambda_{N-1}(\bar{\mathbf{L}})} = \frac{2(N - \lambda_{N-1}(\mathbf{L}))}{2N - \lambda_{N-1}(\mathbf{L})}.$$

6.7.1.2 The Pair-Wise Gossip Algorithm

The pair-wise gossip algorithm is a particular case of random consensus with spatially correlated links in undirected topologies. As for the case of the broadcast gossip algorithm,

[3] The procedure to find the optimum ϵ minimizing $\lambda_1(\mathcal{W})$ under similar conditions is explained in detail in Section 6.7.2.3 using the general formulations.

a node i wakes up randomly at time k with probability $1/N$ but establishes instead a bidirectional link with one of its neighboring nodes with probability $1/\mathcal{N}_i$, where $\mathcal{N}_i$ is the set of neighbors of node i. The rest of the nodes remain silent, such that the instantaneous weight matrix for this protocol is given by

$$\mathbf{W}(k) = \mathbf{I} - \frac{1}{2}(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T$$

coinciding with the definition in Chapter 3 -Section 3.4-. Remark that the weight matrix $\mathbf{W}(k)$ for this protocol has the form in (6.3) where $\mathbf{L}(k) = (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T$ is the instantaneous Laplacian matrix for a pair of communicating nodes $\{i, j\}$ and $\epsilon = 1/2$. The probability of a link between nodes i and j in this case is

$$p_{ij} = \frac{1}{N}\left(\frac{1}{\mathcal{N}_i} + \frac{1}{\mathcal{N}_j}\right) \tag{6.41}$$

where for symmetry, $p_{ij} = p_{ji}$.

Covariance terms Let $a_{ij} = [\mathbf{A}(k)]_{ij}$ as before and observe that

$$\mathbb{E}\left[a_{ij}a_{qr}\right] = \begin{cases} p_{ij}, & \text{if } r = i,\ q = j,\ j \in \mathcal{N}_i \\ 0, & \text{otherwise} \end{cases} \qquad \forall i, j \in \{1, \cdots, N\}.$$

Again, we let $\mathbf{C}_{\substack{ij\\qr}}$ denote the $\{st\}^{th}$ entry of $\mathbf{C}$, with $\{s, t\}$ as defined in (6.1) for nodes $\{i, j, q, r\}$. The entries of the matrix $\mathbf{C}$ are therefore

$$\mathbf{C}_{\substack{ij\\qr}} = \begin{cases} p_{ij}(1 - p_{ij}) & \text{if } q = j, r = i, j \in \mathcal{N}_i \\ -p_{ij}p_{qr} & \text{if } j \in \mathcal{N}_i \text{ and } r \in \mathcal{N}_q \\ 0 & r = j \text{ and } q = i \\ 0 & \text{otherwise} \end{cases} \tag{6.42}$$

with zero elements on the main diagonal. The matrix of correlation terms $\mathbf{R}'$ defined in (6.27) has diagonal and non-diagonal entries

$$\begin{aligned} \mathbf{R}'_{mm} &= \textstyle\sum_i \sum_j \mathbf{C}_{\substack{mi\\mj}} \\ \mathbf{R}'_{mn} &= \textstyle\sum_i \mathbf{C}_{\substack{im\\in}} - \sum_i \mathbf{C}_{\substack{mi\\mn}} - \sum_i \mathbf{C}_{\substack{ni\\nm}} \end{aligned} \tag{6.43}$$

and can be computed using the values given in (6.42). For notation simplicity let $\bar{\mathbf{D}}_m$ denote the $\{mm\}^{th}$ entry of the matrix $\bar{\mathbf{D}} = diag(\mathbf{P} \cdot \mathbf{1})$ and $\tilde{\mathbf{D}}_m$ denote the $\{mm\}^{th}$ entry of the matrix $\tilde{\mathbf{D}} = diag((\mathbf{P} \odot \mathbf{P})\mathbf{1})$, where $\mathbf{P}$ is the connection probability matrix with $\{ij\}^{th}$ entry p_{ij} defined in (6.41). Then we have

$$\begin{aligned}
\sum_i \sum_j \mathbf{C}_{\substack{mi \\ mj}} &= -\left(\sum_{i \in \mathcal{N}_m} \left(\frac{1}{N\mathcal{N}_m} + \frac{1}{N\mathcal{N}_i}\right)\right)^2 + \sum_{i \in \mathcal{N}_m} \left(\frac{1}{N\mathcal{N}_m} + \frac{1}{N\mathcal{N}_i}\right)^2 \\
&= -\bar{\mathbf{D}}_m^2 + \tilde{\mathbf{D}}_m \\
\sum_i \mathbf{C}_{\substack{im \\ in}} &= \sum_{\{m,n\} \in \mathcal{N}_i} \left(\frac{1}{N\mathcal{N}_i} + \frac{1}{N\mathcal{N}_m}\right) \cdot \left(\frac{1}{N\mathcal{N}_i} + \frac{1}{N\mathcal{N}_n}\right) \\
&= -[\mathbf{P}^2]_{mn} \\
\sum_i \mathbf{C}_{\substack{mi \\ mn}} &= -\sum_{\{i,n\} \in \mathcal{N}_m} \left(\frac{1}{N\mathcal{N}_m} + \frac{1}{N\mathcal{N}_i}\right) \cdot \left(\frac{1}{N\mathcal{N}_m} + \frac{1}{N\mathcal{N}_n}\right) \\
&= -[\bar{\mathbf{D}}\mathbf{P}]_{mn} + [\mathbf{P} \odot \mathbf{P}]_{mn} \\
\sum_i \mathbf{C}_{\substack{ni \\ nm}} &= -\sum_{\{i,m\} \in \mathcal{N}_n} \left(\frac{1}{N\mathcal{N}_n} + \frac{1}{N\mathcal{N}_i}\right) \cdot \left(\frac{1}{N\mathcal{N}_n} + \frac{1}{N\mathcal{N}_m}\right) \\
&= -[\mathbf{P}\bar{\mathbf{D}}]_{mn} + [\mathbf{P} \odot \mathbf{P}]_{mn}
\end{aligned}$$

Combining the expressions above as in (6.43) we have

$$\mathbf{R}' = -\bar{\mathbf{D}}^2 - \mathbf{P}^2 + \bar{\mathbf{D}}\mathbf{P} + \mathbf{P}\bar{\mathbf{D}} + 2(\tilde{\mathbf{D}} - \mathbf{P} \odot \mathbf{P})$$

where an extra term $\tilde{\mathbf{D}}$ has been added to compensate for the contribution of $\mathbf{P}^2$ in the main diagonal. The expression above can be rewritten as

$$\mathbf{R}' = 2\tilde{\mathbf{L}} - \bar{\mathbf{L}}^2 \tag{6.44}$$

where $\bar{\mathbf{L}} = \bar{\mathbf{D}} - \mathbf{P}$. Replacing (6.44) in (6.25) yields

$$\mathcal{W} = 2\bar{\mathbf{L}}\epsilon^2 - 2\bar{\mathbf{L}}\epsilon + \mathbf{I} - \mathbf{J}_N. \tag{6.45}$$

Analogously to the previous case, the largest eigenvalue of the matrix $\mathcal{W}$ in (6.45) has a closed-form expression and is given by

$$\lambda_1(\mathcal{W}) = 2\lambda_{N-1}(\bar{\mathbf{L}})\epsilon^2 - 2\lambda_{N-1}(\bar{\mathbf{L}})\epsilon + 1.$$

Furthermore, the value minimizing the function above is

$$\epsilon^* = \frac{1}{2}$$

which is the optimum mixing parameter for pair-wise gossip algorithms. The value of $\lambda_1(\mathcal{W})$ for the optimum ϵ^* is therefore

$$\lambda_1(\mathcal{W}) = 1 - \frac{1}{2}\lambda_{N-1}(\bar{\mathbf{L}}).$$

Summing up, the closed-form expression in (6.25) can be particularized for the pair-wise gossip algorithm, showing that this protocol is a particular case of consensus with spatially correlated undirected communication links.

6.7.1.3 Erdős-Rényi Undirected Topologies with Correlated Links

We consider now an Erdős-Rényi graph composed of N nodes, where any pair of nodes is connected with probability p, and the links have equal covariance $p(1-p)\upsilon$ with $0 < \upsilon < 1$. The instantaneous links are undirected, such that the weight matrix satisfies (6.4) and $\mathcal{W}$ is given by the expression in (6.25). Assuming an all-ones matrix $\tilde{\mathbf{J}} \in \mathbb{R}^{N^2 \times N^2}$ and an identity matrix $\tilde{\mathbf{I}} \in \mathbb{R}^{N^2 \times N^2}$, the matrix $\mathbf{C}$ defined in (6.1) is in this case given by

$$\mathbf{C} = p(1-p)\upsilon(\tilde{\mathbf{J}} - \tilde{\mathbf{I}} - \mathbf{T}) \tag{6.46}$$

where

$$\mathbf{T}_{st} = \begin{cases} 1 & \text{if } s = i + (i-1)N \text{ or } t = j + (j-1)N \\ 0 & \text{otherwise} \end{cases}$$

for all $\{i, j\} \in \{1, \cdots, N\}$. When the entries of the matrix $\mathbf{P}$ are all equal and non-zero, i.e., $\mathbf{P}_{ij} = p$ and $\mathbf{P}_{ii} = 0$ for all $i \neq j$, the expected Laplacian is equal to $\bar{\mathbf{L}} = pN(\mathbf{I} - \mathbf{J}_N)$ and $\tilde{\mathbf{L}} = p^2 N(\mathbf{I} - \mathbf{J}_N)$. $\bar{\mathbf{L}}$ has one eigenvalue equal to 0 with algebraic multiplicity one, and one eigenvalue equal to Np with algebraic multiplicity $N-1$, whereas $\tilde{\mathbf{L}}$ has eigenvalue equal to 0 with algebraic multiplicity one, and one eigenvalue equal to Np^2 with algebraic multiplicity $N-1$. Moreover, the entries of the matrix $\mathbf{R}'$ and its corresponding eigenvalues can be easily computed, since the nonzero entries of $\mathbf{C}$ in (6.46) are all equal. According to (6.27) we have

$$\mathbf{R}'_{mm} = \sum_i \sum_j \mathbf{C}_{\substack{mi \\ mj}} = (N-1)(N-2)p(1-p)\upsilon$$

$$\mathbf{R}'_{mn} = \sum_i \mathbf{C}_{\substack{im \\ in}} - \sum_i \mathbf{C}_{\substack{mi \\ mn}} - \sum_i \mathbf{C}_{\substack{ni \\ nm}} = \Big((N-2) - 2(N-2)\Big)p(1-p)\upsilon = -(N-2)p(1-p)\upsilon$$

such that

$$\mathbf{R}' = N(N-2)p(1-p)\upsilon(\mathbf{I} - \mathbf{J}_N).$$

The eigenvalues of $\mathbf{R}'$ are therefore 0 with algebraic multiplicity one, and $N(N-2)p(1-p)\upsilon$ with algebraic multiplicity $N-1$. Under these connectivity conditions $\bar{\mathbf{L}}, \tilde{\mathbf{L}}$ and $\mathbf{R}'$ are diagonalized by the same set of eigenvectors and the matrix $\mathcal{W}$ has largest eigenvalue

$$\lambda_1(\mathcal{W}) = (Np + (1-p)(2 + \upsilon(N-2)))Np\epsilon^2 - 2Np\epsilon + 1 \tag{6.47}$$

where we have used (6.25). The minimum of $\lambda_1(\mathcal{W})$ is attained when the function above reaches its minimum. Taking the derivative of (6.47) and solving for ϵ we obtain

$$\epsilon^* = \frac{1}{Np + (1-p)(2 + \upsilon(N-2))} \tag{6.48}$$

coinciding with the results in [Jak10]. Substituting (6.48) in (6.47) we obtain the value of $\lambda_1(\mathcal{W})$, which in this case is

$$\lambda_1(\mathcal{W})\big|_{\epsilon^*} = \frac{(1-p)(2 + \upsilon(N-2))}{Np + (1-p)(2 + \upsilon(N-2))}. \tag{6.49}$$

Finally, since the function in (6.47) is quadratic on ϵ, the dynamic range is given by

$$\epsilon \in \left(0, \frac{2}{Np + (1-p)(2 + \upsilon(N-2))}\right). \tag{6.50}$$

This example is useful to observe that as the correlation increases, the value of $\lambda_1(\mathcal{W})$ increases and the convergence time of the upper bound defined in (6.16), decreases.

6.7.1.4 Erdős-Rényi Directed Topologies with Uncorrelated Links

Analogous to the previous case we assume an Erdős-Rényi graph composed of N nodes connected with probability p, but consider instead spatially independent links. The procedure to derive ϵ^* and its dynamic range is similar to the previous one where, using instead (6.33), the largest eigenvalue of $\mathcal{W}$ is given by

$$\lambda_1(\mathcal{W}) = \left(N^2p + 2(N-1)(1-p)\right)p\epsilon^2 - 2Np\epsilon + 1. \tag{6.51}$$

Taking the derivative of (6.51) and solving for ϵ we obtain

$$\epsilon^* = \frac{N}{N^2p + 2(N-1)(1-p)} \tag{6.52}$$

and substituting ϵ^* in (6.51) we obtain the value of $\lambda_1(\mathcal{W})$, given by

$$\lambda_1(\mathcal{W})\big|_{\epsilon^*} = \frac{2(N-1)(1-p)}{N^2p + 2(N-1)(1-p)}. \tag{6.53}$$

Again, since the function in (6.51) is quadratic on ϵ, the dynamic range of ϵ is given by

$$\epsilon \in \left(0, \frac{2N}{N^2p + 2(N-1)(1-p)}\right). \tag{6.54}$$

Remark 6.1. *The optimum link weight in (6.52) as well as its dynamic range in (6.54), coincide with the expressions derived in Chapter 5 -Section 5.5- minimizing the mean square convergence of the consensus algorithm for the case of Erdős-Rényi directed random networks with spatially independent links.*

6.7.2 Derivation of Expressions for Further Particular Cases

In this section we analyze particular cases of networks with links of equal probability of connection whose optimum link weights allow a closed-form expression. In case of no correlation among pairs of links, the matrix $\mathcal{W}$ is given by (6.29) for undirected topologies and by (6.33) for directed topologies. In Section 6.7.2.1 we derive results for Erdős-Rényi directed topologies with correlated links whereas in Section 6.7.2.2 we focus on Erdős-Rényi undirected topologies with uncorrelated links. These examples will be useful along with some results from Section 6.7.1 to evaluate the impact of correlation on the convergence rate. The optimum value of ϵ minimizing the convergence time of the consensus algorithm in networks with generic expected topologies and its dynamic range are found analytically for both the directed and the undirected case in Sections 6.7.2.3 and 6.7.2.4 respectively.

6.7.2.1 Erdős-Rényi Directed Topologies with Correlated Links

Consider an Erdős-Rényi graph composed of N nodes connected with probability p, and equal covariance among links according to the model in (6.46). This is the generalization of the example in Section 6.7.1.3 to directed topologies, or equivalently, the counterpart of the model in Section 6.7.1.4 with spatially correlated links, where again we have that $\bar{\mathbf{L}} = pN(\mathbf{I} - \mathbf{J}_N)$ and $\tilde{\mathbf{L}} = p^2N(\mathbf{I} - \mathbf{J}_N)$. The entries of the matrix $\mathbf{R}$ and its corresponding eigenvalues can be easily computed using (6.31) as follows

$$\mathbf{R}_{mm} = \left((N-1)(N-2) + 2\left(1 - \frac{1}{N}\right)\right) p(1-p)\upsilon$$
$$\mathbf{R}_{mn} = \left(-(N-2) - \frac{2}{N}\right) p(1-p)\upsilon$$

such that

$$\mathbf{R} = ((N-1)^2 + 1)p(1-p)\upsilon(\mathbf{I} - \mathbf{J}_N).$$

The eigenvalues of $\mathbf{R}$ are therefore 0 with algebraic multiplicity one, and $((N-1)^2 + 1)p(1-p)\upsilon$ with algebraic multiplicity $N-1$. Under these connectivity conditions $\bar{\mathbf{L}}, \tilde{\mathbf{L}}$ and $\mathbf{R}$ are diagonalized by the same set of eigenvectors and the matrix $\mathcal{W}$ has largest eigenvalue

$$\lambda_1(\mathcal{W}) = \left(Np + (1-p)\left(\frac{2(N-1)}{N} + \upsilon\left(N - 2 + \frac{2}{N}\right)\right)\right) Np\epsilon^2 - 2Np\epsilon + 1$$

where we have used (6.30). Taking the derivative of the function above with respect to ϵ and solving, we obtain the optimum ϵ given by

$$\epsilon^* = \frac{N}{N^2 p + (1-p)(2(N-1) + \upsilon\left((N-1)^2 + 1\right))} \tag{6.55}$$

and substituting for ϵ^* in the expression for $\lambda_1(\mathcal{W})$ we obtain

$$\lambda_1(\mathcal{W})\big|_{\epsilon^*} = \frac{(1-p)(2(N-1) + \upsilon\left((N-1)^2 + 1\right))}{N^2 p + (1-p)(2(N-1) + \upsilon\left((N-1)^2 + 1\right))}. \tag{6.56}$$

The dynamic range is obtained as before and yields

$$\epsilon \in \left(0, \frac{2N}{N^2 p + (1-p)(2(N-1) + \upsilon\left((N-1)^2 + 1\right))}\right). \tag{6.57}$$

Remark 6.2. *Comparing these results with the ones in 6.7.1.4 for the case of uncorrelated links, we can state that spatial correlation is detrimental to the convergence rate of consensus algorithms in random directed topologies. This is clear comparing the expressions for $\lambda_1(\mathcal{W})$ in (6.53) and (6.56). Further, comparing (6.54) and (6.57) it can be observed that correlation reduces the dynamic range of the link weights. In addition, particularizing the results of this section for $\upsilon = 0$, we obtain the results of Section 6.7.1.4.*

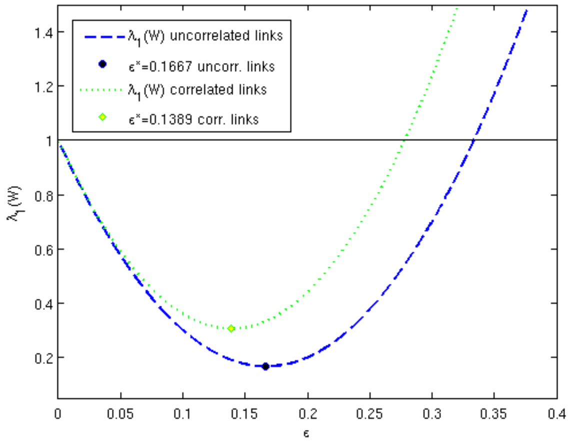


Figure 6.1: $\lambda_1(\mathcal{W})$ as a function of ϵ for uncorrelated links and correlated links with $\upsilon = 0.3$ respectively along with ϵ^* for a network with $N = 20$ and $p = 0.5$.

The curve for $\lambda_1(\mathcal{W})$ as a function of ϵ for an example deployment of $N = 20$ nodes and probability of connection $p = 0.5$ is depicted in Fig. 6.1 for both spatially uncorrelated and spatially correlated links with $\upsilon = 0.3$. The optimum ϵ for the uncorrelated case is $\epsilon^* = 0.1667$, whereas the optimum ϵ for the correlated case is $\epsilon^* = 0.1389$.

6.7.2.2 Erdős-Rényi Undirected Topologies with Uncorrelated Links

In this section we consider the case of uncorrelated links in undirected topologies, that is, the counterpart of Section 6.7.1.3 with uncorrelated links. Consider an Erdős-Rényi graph composed of N nodes connected with probability p, and spatially uncorrelated communication links. We proceed as before but taking into account that $\mathbf{R}'$ is a zero matrix, which leads to further simplified computations. The matrix $\mathcal{W}$ for this case has largest eigenvalue

$$\lambda_1(\mathcal{W}) = (Np + 2(1-p))\, Np\epsilon^2 - 2Np\epsilon + 1 \tag{6.58}$$

where we have used (6.29). Taking the derivative and solving for ϵ we obtain the value of ϵ minimizing (6.58), i.e.,

$$\epsilon^* = \frac{1}{Np + 2(1-p)} \tag{6.59}$$

and substituting ϵ^* in (6.58) we obtain

$$\lambda_1(\mathcal{W})\big|_{\epsilon^*} = \frac{2(1-p)}{Np + 2(1-p)}. \tag{6.60}$$

Analogously to the previous cases, the dynamic range is given by

$$\epsilon \in \left(0, \frac{2}{Np + 2(1-p)}\right). \tag{6.61}$$

Note that another way to obtain the expressions in (6.59) and (6.61) is using (6.47) with $\upsilon = 0$.

In the following sections we assume generic expected topologies where the communications are restricted to a smaller number of nodes, that is, not every pair of nodes may exchange information, and spatially uncorrelated links existing with the same probability. The optimum ϵ minimizing the convergence time of the algorithm can be found analytically as well for both the directed and the undirected case, as described below.

6.7.2.3 Generic Directed Topologies with Uncorrelated Links

We consider a WSN with generic expected topology, where the instantaneous links are random directed and spatially uncorrelated, existing with probability p. Recall that for the case of equally probable links, the matrices $\bar{\mathbf{L}}$ and $\tilde{\mathbf{L}}$ are diagonalized by the same set of eigenvectors because $\tilde{\mathbf{L}} = p\bar{\mathbf{L}}$. Therefore, the eigenvalues are related by $\lambda_i(\tilde{\mathbf{L}}) = p\lambda_i(\bar{\mathbf{L}})$. Using (6.33), we can express the eigenvalues of $\mathcal{W}$ as functions of the eigenvalues of $\bar{\mathbf{L}}$ as follows

$$f_i(\epsilon) = \left(\lambda_i(\bar{\mathbf{L}})^2 + \lambda_i(\bar{\mathbf{L}})\eta\right)\epsilon^2 - 2\lambda_i(\bar{\mathbf{L}})\epsilon + 1, \quad \forall i \in \{1, \cdots, N-1\} \tag{6.62}$$

where

$$\eta = \frac{2(N-1)}{N}(1-p) \tag{6.63}$$

and we have considered that $f_N(\epsilon) = 0$. Note that these are quadratic functions of ϵ as well. First of all, recall that the subindex i of the functions $f_i(\epsilon)$ are in one to one correspondence with the ordering of the eigenvalues of $\bar{\mathbf{L}}$ but not with the ordering of the eigenvalues of $\mathcal{W}$. This is due to the fact that, as the eigenvalues of $\bar{\mathbf{L}}$ increase in magnitude, the terms on ϵ in (6.62) increase in magnitude as well and the quadratic

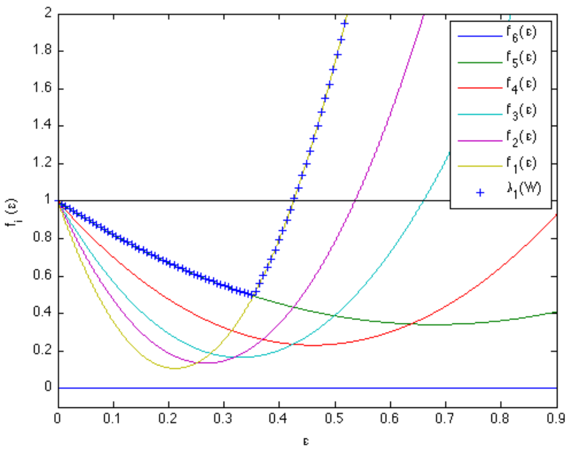


Figure 6.2: $\lambda_{1(\mathcal{W})}$ as a function of ϵ for uncorrelated links along with the curves for $f_i(\epsilon)$ in (6.62) for a random geometric network with $N=6$ and $p=0.7$.

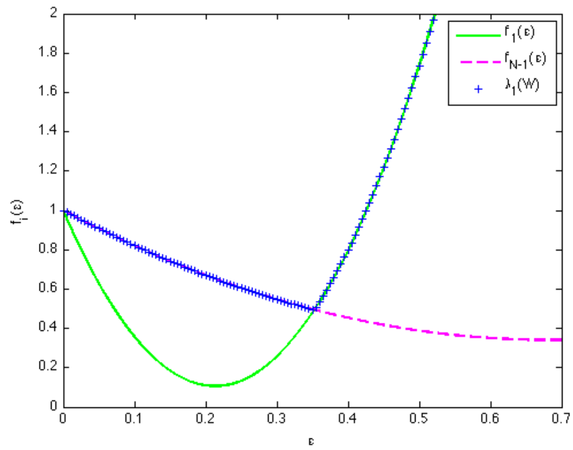


Figure 6.3: $f_1(\epsilon)$ and $f_{N-1}(\epsilon)$ as a function of ϵ along with the curve for $\lambda_1(\mathcal{W})$ for the same random geometric network with uncorrelated links with $N=6$ and $p=0.7$.

curves become narrower with a vertex approaching the abscissa while moving towards the origin. For the sake of clarity, Fig. 6.2 depicts the set of functions in (6.62) along with $\lambda_1(\mathcal{W})$ for a particular network composed of $N=6$. We relate the functions in (6.62) with $\lambda_1(\mathcal{W})$ in four steps:

i) First, we evaluate their slope. Note that at $\epsilon=0$, $f_i(\epsilon)$ is equal to one for all $i \in \{1, \cdots, N-1\}$, and the slope evaluated at that point is equal to $-2\lambda_i(\bar{\mathbf{L}})$. The curve experiencing the slowest decay will be the one corresponding to $i=N-1$. In other words, in the proximity of $\epsilon=0$ the largest eigenvalue of $\mathcal{W}$ will be given by $f_{N-1}(\epsilon)$, in correspondence with the results of Section 6.5.1.

ii) Next, we have to evaluate the intersections of $f_{N-1}(\epsilon)$ with the remaining $f_j(\epsilon)$ for $j \in \{1, \cdots, N-2\}$, i.e., $f_{N-1}(\epsilon) = f_j(\epsilon)$, which occur at the point

$$\epsilon = \frac{2(\lambda_{N-1}(\bar{\mathbf{L}}) - \lambda_j(\bar{\mathbf{L}}))}{(\lambda_{N-1}^2(\bar{\mathbf{L}}) - \lambda_j^2(\bar{\mathbf{L}})) + \eta(\lambda_{N-1}(\bar{\mathbf{L}}) - \lambda_j(\bar{\mathbf{L}}))} = \frac{2}{\lambda_{N-1}(\bar{\mathbf{L}}) + \lambda_j(\bar{\mathbf{L}}) + \eta} \qquad (6.64)$$

for a given j. The first intersection of $f_{N-1}(\epsilon)$ will take place when the value of ϵ in (6.64) is minimum, and this happens for $j=1$. At this point, substituting for $j=1$ in (6.64) we obtain

$$\epsilon_{int} = \frac{2}{\lambda_{N-1}(\bar{\mathbf{L}}) + \lambda_1(\bar{\mathbf{L}}) + \eta}. \quad (6.65)$$

iii) It is not difficult to check that no other curve intersects $f_1(\epsilon)$ for $\epsilon > \epsilon_{int}$ since the crossing points fulfill $2/(\lambda_1(\bar{\mathbf{L}}) + \lambda_j(\bar{\mathbf{L}}) + \eta) < \epsilon_{int}$, for $j \in \{2, \cdots, N-2\}$. Summing up, $\lambda_1(\mathcal{W})$ is given by $f_{N-1}(\epsilon)$ for $\epsilon \in (0, \epsilon_{int})$, whereas from ϵ_{int} it is given by $f_1(\epsilon)$.

iv) The last step consists in checking if the minimum of $f_{N-1}(\epsilon)$ is attained before or after the intersection with $f_1(\epsilon)$. Note that the minimum of $f_{N-1}(\epsilon)$ appears at the point

$$\epsilon_{min} = \frac{1}{\lambda_{N-1}(\bar{\mathbf{L}}) + \eta} \quad (6.66)$$

and this value of ϵ is larger than the one where $f_1(\epsilon)$ attains its minimum, i.e., $1/(\lambda_1(\bar{\mathbf{L}})+\eta)$. Therefore, the optimum value of the link weights solving (6.18) will be given by

$$\epsilon^* = \min\left\{\frac{2}{\lambda_{N-1}(\bar{\mathbf{L}}) + \lambda_1(\bar{\mathbf{L}}) + \eta}, \frac{1}{\lambda_{N-1}(\bar{\mathbf{L}}) + \eta}\right\}. \quad (6.67)$$

Note that (6.65) will be the optimum when $\lambda_{N-1}(\bar{\mathbf{L}}) + \eta < \lambda_1(\bar{\mathbf{L}})$, and this will happen already from relatively small values of N and p (see Fig. 6.3). If the optimum link weight is given by ϵ_{int} in (6.65), the expression for $\lambda_1(\mathcal{W})$ is given by

$$\lambda_1(\mathcal{W})\Big|_{e*} = \frac{(\lambda_{N-1}(\bar{\mathbf{L}}) + \lambda_1(\bar{\mathbf{L}}) + \eta)^2 - 4\lambda_{N-1}(\bar{\mathbf{L}})\lambda_1(\bar{\mathbf{L}})}{(\lambda_{N-1}(\bar{\mathbf{L}}) + \lambda_1(\bar{\mathbf{L}}) + \eta)^2}.$$

On the other hand, if the optimum link weight is given by ϵ_{min} in (6.66), the expression for $\lambda_1(\mathcal{W})$ is

$$\lambda_1(\mathcal{W})\Big|_{e*} = \frac{\eta}{\lambda_{N-1}(\bar{\mathbf{L}}) + \eta}.$$

In order to specify the dynamic range of ϵ, observe that the curve for $f_1(\epsilon)$ is the narrower one, and since no other curves are crossing $f_1(\epsilon)$ for $\epsilon > \epsilon_{int}$, the upper bound for ϵ is found simply equating $f_1(\epsilon)$ to one. Solving for ϵ we obtain

$$\epsilon \in \left(0, \frac{2}{\lambda_1(\bar{\mathbf{L}}) + \eta}\right). \quad (6.68)$$

6.7.2.4 Generic Undirected Topologies with Uncorrelated Links

Analogous to the previous case, we consider a WSN with generic expected topology and equally probable links, but with undirected instantaneous links instead. Using (6.29), the set of functions used for the analysis are

$$f_i(\epsilon) = (\lambda_i(\bar{\mathbf{L}})^2 + \lambda_i(\bar{\mathbf{L}})\eta')\epsilon^2 - 2\lambda_i(\bar{\mathbf{L}})\epsilon + 1, \quad \forall i \in \{1, \cdots, N-1\}$$

where $\eta' = 2(1-p)$. The optimum value of the link weights under these connectivity conditions is given by

$$\epsilon^* = \min\left\{\frac{2}{\lambda_{N-1}(\bar{\mathbf{L}}) + \lambda_1(\bar{\mathbf{L}}) + \eta'}, \frac{1}{\lambda_{N-1}(\bar{\mathbf{L}}) + \eta'}\right\}. \tag{6.69}$$

If the first expression to the right hand side of (6.69) is the value of the optimum ϵ, then

$$\lambda_1(\mathcal{W})\big|_{e*} = \frac{(\lambda_{N-1}(\bar{\mathbf{L}}) + \lambda_1(\bar{\mathbf{L}}) + \eta')^2 - 4\lambda_{N-1}(\bar{\mathbf{L}})\lambda_1(\bar{\mathbf{L}})}{(\lambda_{N-1}(\bar{\mathbf{L}}) + \lambda_1(\bar{\mathbf{L}}) + \eta')^2}$$

whereas if the second expression to the right-hand side of (6.69) is the value of the optimum ϵ, then

$$\lambda_1(\mathcal{W})\big|_{e*} = \frac{\eta'}{\lambda_{N-1}(\bar{\mathbf{L}}) + \eta'}.$$

Analogously to the directed case in the previous section, the dynamic range of ϵ is given by

$$\epsilon \in \left(0, \frac{2}{\lambda_1(\bar{\mathbf{L}}) + \eta'}\right). \tag{6.70}$$

Remark 6.3. *When the probability of connection p is equal to 1, we obtain a deterministic system and the closed-form expressions in (6.67) and (6.68), and (6.69) and (6.70) coincide with the expressions for the optimum ϵ and its dynamic range derived in Xiao and Boyd [Xia03] for fixed topologies, that is*

$$\epsilon^* = \frac{2}{\lambda_{N-1}(\bar{\mathbf{L}}) + \lambda_1(\bar{\mathbf{L}})} \quad \textit{and} \quad \epsilon \in \left(0, \frac{2}{\lambda_1(\bar{\mathbf{L}})}\right).$$

6.8 Numerical Results

The analytical results obtained in the previous sections are supported with computer simulations of three different scenarios: a random geometric network [Bol01], a small-world network [Wat98], and a network with randomized transmission power as presented in Chapter 5 -Section 5.7-. The random geometric network is a general case where the links have different probabilities of connection and different spatial correlation, and is therefore a useful case to verify and support the analytical expressions derived in the chapter.

6.8.1 Random Geometric Network

We consider $N = 20$ nodes randomly deployed within a unit square and with fixed position, where each pair of neighboring nodes may have a connection only if the euclidean distance between them is smaller than a given threshold, chosen equal to 0.37 -see Fig. 6.4-. The entries of $\mathbf{x}(0)$ are modeled as Gaussian r.v.'s with mean $x_m = 20$ and variance $\sigma_0^2 = 5$. For the verification of the closed-form expressions, we choose a very general model with different probabilities of connection for the possible links and different correlation among pairs of links. Therefore, the instantaneous links among neighboring nodes i and j are generated as correlated Bernoulli r.v.'s with probability $p_{ij} = p_{ji}$ where p_{ij} is chosen uniformly at random in the interval $[0, 1]$. For the spatial correlation we consider the autoregressive model in [Lun98], included in Appendix 6.C for the sake of clarity. The matrix $\mathbf{C}$ in this case is given by

$$\mathbf{C}_{st} = \begin{cases} \psi(1-\psi)v^{|t-s|}\zeta_s\zeta_t, & s \neq t \\ 0, & s = t \end{cases} \quad \text{with} \quad \begin{cases} s = i + (j-1)N \\ t = m + (n-1)N \end{cases} \tag{6.71}$$

for nodes $\{i, j, m, n\}$, where ψ, v and ζ are as defined in Appendix 6.C.

A total of 10.000 independent realizations were run to obtain the expected squared norm of the deviation vector $\mathbb{E}\left[\|\mathbf{d}(k)\|_2^2\right]$, where the matrix $\mathbf{P}$ was kept fixed while a new $\mathbf{L}(k)$ was generated at each iteration. The closed-form expression in (6.30) has been verified with simulations, and the difference between the theoretical value of $\lambda_1(\mathcal{W})$ and

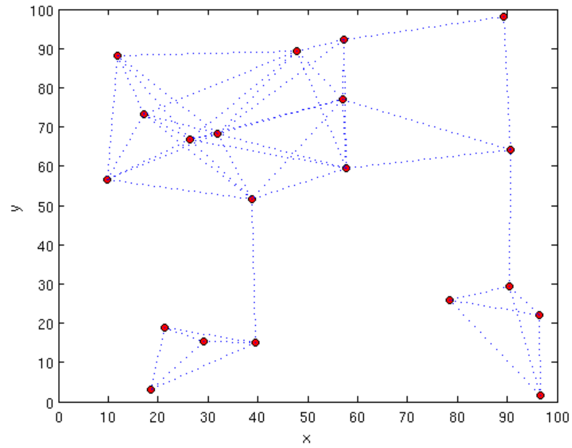


Figure 6.4: Deployment of $N = 20$ nodes in the unit square with connectivity radius equal to 0.37.

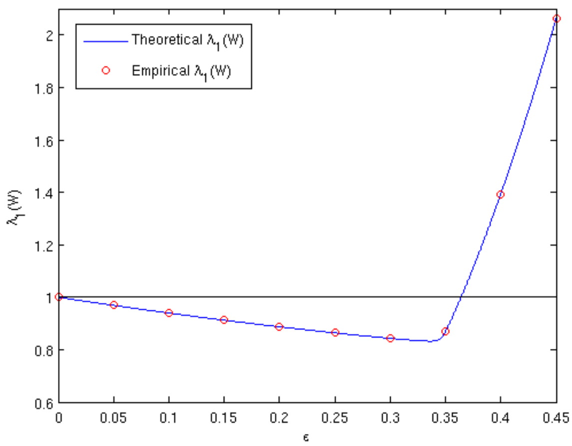


Figure 6.5: Theoretical and empirical $\lambda_1(\mathcal{W})$ as a function of ϵ for a network with $N = 20$.

the empirical one is around 1×10^{-5}. Fig. 6.5 depicts the theoretical value of $\lambda_1(\mathcal{W})$ (solid line) along with samples of the empirical value ('o') as a function of ϵ for the example deployment of Fig. 6.4.

Fig. 6.6 shows the expected squared norm of the deviation vector in log-linear scale as a function of the iteration index for three different values of ϵ:

(1) $\epsilon = 1/(N-1) = 0.0526$, considered as the worst case scenario (dotted line)

(2) $\epsilon_{bound} = 0.1850$, minimizing the upper bound for the MSE($x(k)$) derived in Section 5.6 (dashed line)

(3) $\epsilon^* = 0.3367$ minimizing $\lambda_1(\mathcal{W})$ (solid-line)

As expected, we observe that the choice of ϵ^* minimizes the convergence time of the algorithm. For the case of ϵ^*, the error is below 10^{-3} after 43 iterations, whereas for the cases $\epsilon = 1/(N-1)$ and ϵ_{bound}, the error is below that value after respectively 256 and 73 iterations.

We have also computed the empirical MSE($x(k)$) characterizing the deviation of the state with respect to the statistical mean of the observations, that is, determining the error caused by allowing asymmetric links. Since the network is connected and symmetric in average over time, the estimation is unbiased and the consensus value c is, in expectation, equal to the average consensus. Fig. 6.7 shows the empirical MSE($x(k)$) in log-linear scale for the three values of ϵ defined above, along with the benchmark value σ_0^2/N (solid line). Again, choosing ϵ^* we obtain fastest convergence whereas choosing the smallest ϵ the convergence is slower but as expected, the error curve is closest to the benchmark. The choice of ϵ^* outperforms also the performance of ϵ_{bound}.

The results depicted in Fig. 6.6 are useful to verify that the convergence using ϵ^* is faster, whereas the results in Fig. 6.7 are useful to evaluate the error with respect to the statistical mean of the initial measurements when choosing the different values of ϵ. We conclude that finding the ϵ that minimizes $\lambda_1(\mathcal{W})$ is a good criterion to reduce the convergence time of the consensus algorithm under these topology conditions.

6.8.2 Small-World Network

We consider a small-world network with $N = 20$, 4 nearest neighbors and shortcut probability 0.4, where the non-zero entries of the matrix $\mathbf{P}$ are set equal to $p = 0.4$. The links are assumed spatially uncorrelated and the optimum link weight is computed using (6.67). Fig. 6.8 depicts the empirical squared norm of the deviation vector in log-linear scale as a function of the iteration index for three different values of ϵ:

(1) $\epsilon = 0.0526$ (dotted line)

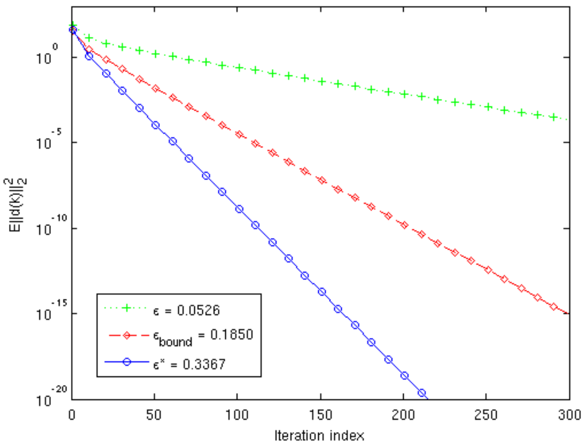


Figure 6.6: Expected squared norm of $\mathbf{d}(k)$ as a function of k in log-linear scale for different probabilities of connection and different values of ϵ in a random geometric graph with autoregressive correlated links.

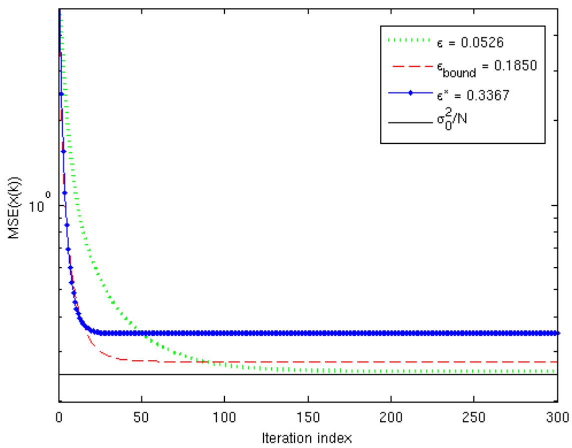


Figure 6.7: Empirical MSE($x(k)$) as a function of the iteration index k in log-linear scale for the same random geometric deployment with different probabilities of connection and different values of ϵ.

(2) $\epsilon_{bound} = 0.1686$ (dashed line)

(3) $\epsilon^* = 0.2633$ minimizing $\lambda_1(\mathcal{W})$ (solid line)

The expected squared norm of the deviation vector tends to zero, and as expected, choosing ϵ^* we achieve fastest convergence also in small-world networks. For the case of ϵ^*, the error is below 10^{-3} after only 13 iterations, whereas for the cases $\epsilon = 1/(N-1)$ and ϵ_{bound},

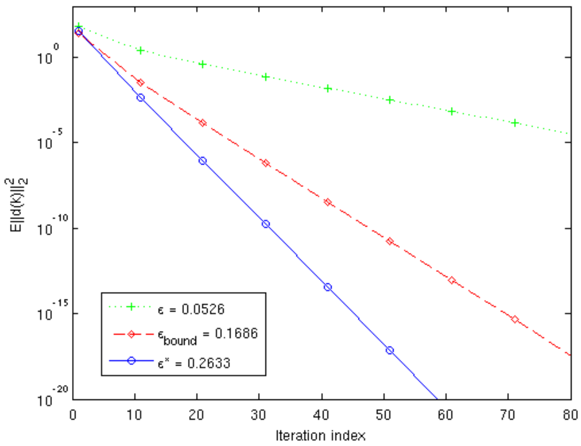


Figure 6.8: Expected squared norm of $\mathbf{d}(k)$ as a function of k in log-linear scale for different values of ϵ in a small-world network with $N = 20$, 4 neighboring nodes and $p = 0.4$.

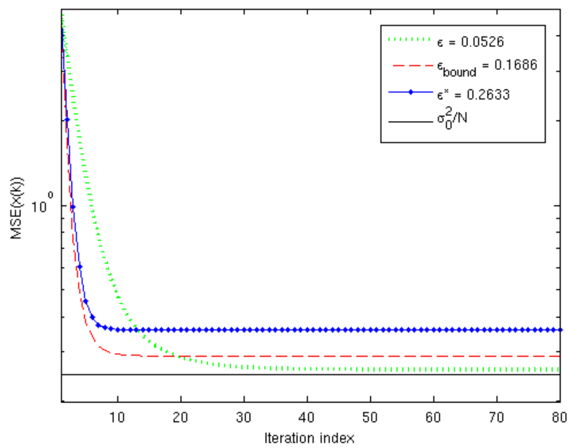


Figure 6.9: Empirical MSE$(x(k))$ as a function of k in log-linear scale for the small-world with different probabilities of connection and different values of ϵ.

the error is below that value after respectively 59 and 18 iterations.

Finally, Fig. 6.9 shows the empirical MSE$(x(k))$ in log-linear scale for the different values of ϵ, along with the benchmark value σ_0^2/N (solid line). Again, choosing ϵ^* we obtain fastest convergence whereas choosing the smallest ϵ the curve is closest to the benchmark. The choice of ϵ^* outperforms also the performance of the ϵ_{bound}, showing that the minimization of $\lambda_1(\mathcal{W})$ is a good criterion to reduce the convergence time of consensus algorithms in small-world networks.

6.8.3 Network with Randomized Transmission Power

For the last simulation we use the same deployment as the one in Section 6.8.1, i.e., the position of the nodes and the expected graph correspond to the random geometric network depicted in Fig. 6.4. In this case however, the nodes transmit using different power levels at each time instant, selected at random from a predefined range of values and independently of the rest of the nodes. In Chapter 5 -Section 5.7- we considered a similar transmission scheme but assuming a discrete set of power levels, and observed via simulations that the overall energy consumption of the network until convergence is strongly reduced when implementing the of the consensus algorithm. Here, we address a more general case where the power decays exponentially with distance and may take any value in a closed continuous interval.

Each power level describes approximately a circle of connectivity with radius $\rho \in (0, \rho_{max}]$ proportional to the square root of the transmission power level and centered at the transmitting node. Thus, if node i transmits with a power level at iteration k defining a connectivity radius ρ_i, we assume node j will receive data from node i at iteration k whenever the distance $d(i,j)$ between them is less or equal to ρ_i. The connectivity radius ρ at time k is the realization of a r.v. with exponential density function, i.e.,

$$\rho \in (0, \rho_{max}] \text{ with probability density function } f_\rho(\rho) = \frac{1}{\rho_{ave}} e^{-\frac{\rho}{\rho_{ave}}}$$

where ρ_{ave} is the statistical mean. For practical reasons we consider $\rho_{max} < \infty$. However, $\rho_{max} >> \rho_{ave}$ must be satisfied so that ρ_{ave} can be considered the mean. In fact, we choose a value of ρ_{max} such that 90% of the nodes are located within the maximum connectivity radius. The connection probability matrix $\mathbf{P}$ for this model has entries given by

$$\mathbf{P}_{ij} = \begin{cases} p_{ij} = \Pr\{\rho \geq d(i,j)\} & i \neq j \\ 0 & i = j \end{cases}$$

where

$$\Pr\{\rho \geq d(i,j)\} = \int_{d(i,j)}^{\rho_{max}} f_\rho(\rho) d\rho.$$

In order to compute the matrix **C** observe that, since the transmission power is chosen independently by each node at each time instant, two links are correlated only if they are receiving data from a third transmitting node no further away than ρ_{max}, such that

$$\mathbf{C}_{st} = \begin{cases} 0 & s = t \\ p_{im}(1 - p_{jm}) & \text{if } \rho_{max} \geq d(i,m) \geq d(j,m) \\ p_{jm}(1 - p_{im}) & \text{if } \rho_{max} \geq d(j,m) \geq d(i,m) \\ 0 & \text{otherwise} \end{cases}, \text{ for } \begin{cases} s = i + (m-1)N \\ t = j + (m-1)N \end{cases}$$

Note that this matrix is much more sparse than the general one in (6.1). Then, we compute the matrix $\mathcal{W}$ using (6.30) and find the value of ϵ minimizing $\lambda_1(\mathcal{W})$.

Fig. 6.10 shows the empirical squared norm of the deviation vector in log-linear scale as a function of the iteration index for three different values of ϵ as before:

(1) $\epsilon = 0.0526$ (dotted line)
(2) $\epsilon_{bound} = 0.1250$ (dashed line)
(3) $\epsilon^* = 0.2060$ (solid line)

The choice of ϵ^* provides a fast convergence of the deviation vector, although the difference with respect to the case $\epsilon = 1/(N-1)$ is not distinguishable in the high error regime for this particular deployment. For instance, the error is below 10^{-3} after only 12 iterations using ϵ^*, but the same results are obtained with ϵ_{bound}. For $\epsilon = 1/(N-1)$ however, the error is below that value after 30 iterations. This similarity in the results for the cases ϵ^* and ϵ_{bound} is observed until reducing the error to 10^{-5}, from where the performance improves using ϵ^*.

Fig. 6.11 shows the empirical MSE$(x(k))$ in log-linear scale along with the benchmark value σ_0^2/N (solid line). Again, the improvement when choosing ϵ^* for this example is not so significant as for the ones in the previous sections when comparing it to the performance of ϵ_{bound}. However, the MSE$(x(k))$ converges faster for ϵ^* when comparing it to the case of $\epsilon = 1/(N-1)$. We conclude that the minimization of $\lambda_1(\mathcal{W})$ is a good design criterion to reduce the convergence time of the consensus algorithm in WSNs with randomized transmission power.

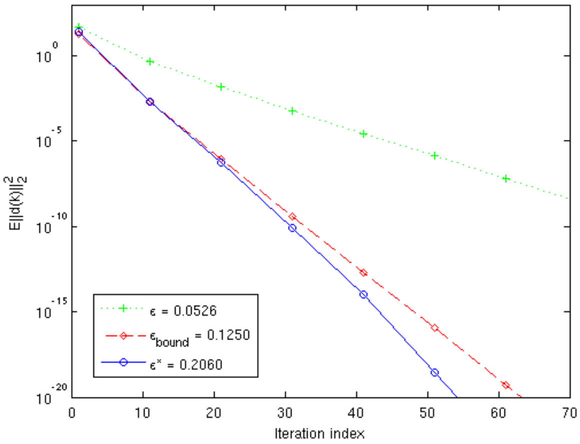


Figure 6.10: Expected squared norm of $\mathbf{d}(k)$ as a function of k in log-linear scale for a random geometric graph with randomized transmission power and different values of ϵ.

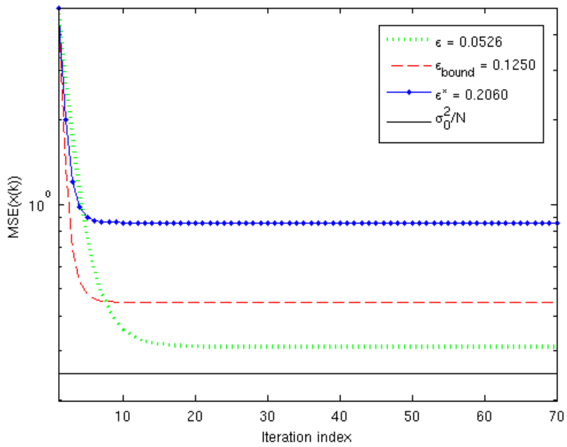


Figure 6.11: Empirical MSE$(x(k))$ as a function of k in log-linear scale for a random geometric graph with randomized transmission power and different values of ϵ.

6.9 Conclusions of the Chapter

Almost sure convergence to a consensus in random WSNs has been studied, assuming both instantaneous directed topologies and instantaneous undirected topologies, and allowing the links to be spatially correlated. Convergence in expectation to the average consensus has been shown and a useful criterion for the minimization of the convergence time has been adopted. This criterion states a sufficient condition for almost sure convergence and is based on the spectral radius of a positive semidefinite matrix, for which we have derived closed-form expressions assuming uniform link weights. The closed-form expressions derived are useful for the computation of the optimum link weights minimizing the convergence time of an upper bound for the error vector.

The general expressions derived in this chapter subsume existing protocols found in literature as it has been shown, and greatly simplify the derivation of the optimum link weights, not only in networks with Erdős-Rényi topologies but also in networks with generic expected topologies where the communication links among pairs of links exist with equal probability.

The analytical results obtained for particular cases of networks with random topologies and equally probable communication links show that spatial correlation is detrimental to the convergence rate of consensus algorithms and reduces the dynamic range of the link weights.

The analytical results are further validated with computer simulations of a general case with different probabilities of connection for the links and different correlations among pairs of links. Simulations of a small-world network and simulations of a randomized transmission power network using the optimum link weights derived in the chapter are also provided, as well as using the link weights minimizing the MSE upper bound derived in Chapter 5. A reduction on the total number of iterations until convergence can be observed with the optimization criterion adopted in all cases, as well as an improved performance with respect to the link weights minimizing the MSE upper bound from Chapter 5.

6.A Appendix: Proof of Theorem 6.2

We start the proof replacing the expression (6.3) in (6.24). After some matrix manipulations we obtain

$$\mathcal{W} = \mathbb{E}\left[\mathbf{L}(k)^T\mathbf{L}(k)\right]\epsilon^2 - 2\bar{\mathbf{L}}\epsilon + \mathbf{I} - \mathbf{J}_N. \tag{6.72}$$

In order to compute the expected matrix to the right-hand side of the expression above, remark that at time k we have diagonal and non-diagonal elements given by

$$[\mathbf{L}(k)^T\mathbf{L}(k)]_{mm} = \left(\sum_i a_{mi}\right)^2 + \sum_i a_{im}^2 = \sum_i a_{mi}^2 + \sum_i\sum_{j\neq i} a_{mi}a_{mj} + \sum_i a_{im}^2 \tag{6.73}$$

$$\begin{aligned}[\mathbf{L}(k)^T\mathbf{L}(k)]_{mn} &= -\sum_i a_{mi}a_{mn} - \sum_i a_{ni}a_{nm} + \sum_i a_{im}a_{in} \\ &= -a_{mn}^2 - \sum_{i\neq n} a_{mi}a_{mn} - a_{nm}^2 - \sum_{i\neq m} a_{ni}a_{nm} + \sum_i a_{im}a_{in}\end{aligned} \tag{6.74}$$

where $a_{mn} = [\mathbf{A}(k)]_{mn}$, $a_{mm} = 0$ for all m, and the sums are from $\{i,j\} = 1$ to N. Consider the $\mathbf{C}$ matrix in (6.1), and for notation clarity, let $\mathbf{C}_{\substack{ij\\qr}}$ denote the entry $\mathbf{C}_{st}$ with $s = i + (j-1)N$ and $t = q + (r-1)N$ as before. Taking the expectation of the expressions above we obtain

$$\mathbb{E}[\mathbf{L}(k)^T\mathbf{L}(k)]_{mm} = \sum_i p_{mi} + \sum_i\sum_{j\neq i} p_{mi}p_{mj} + \sum_i p_{im} + \sum_i\sum_j \mathbf{C}_{\substack{mi\\mj}}$$

$$\mathbb{E}[\mathbf{L}(k)^T\mathbf{L}(k)]_{mn} = -p_{mn} - \sum_{i\neq n} p_{mi}p_{mn} - p_{nm} - \sum_{i\neq m} p_{ni}p_{nm} + \sum_i p_{im}p_{in} - \sum_i \mathbf{C}_{\substack{mi\\mn}} - \sum_i \mathbf{C}_{\substack{ni\\nm}} + \sum_i \mathbf{C}_{\substack{im\\in}}$$

where we have assumed that $\mathbf{C}_{\substack{ii\\qr}} = \mathbf{C}_{\substack{ij\\qq}} = \mathbf{C}_{\substack{ij\\ij}} = 0$, for all $\{i,j,q,r\} \in \{1,\cdots,N\}$ and considered that $\mathbf{P}$ is symmetric with zero diagonal entries. Rearranging terms we have

$$\begin{aligned}\mathbb{E}[\mathbf{L}(k)^T\mathbf{L}(k)]_{mm} &= 2\sum_i p_{mi} + \left(\sum_i p_{mi}\right)^2 - \sum_i p_{mi}^2 + \sum_i\sum_j \mathbf{C}_{\substack{mi\\mj}} \\ &= 2\bar{\mathbf{D}}_{mm} + \bar{\mathbf{D}}_{mm}^2 - \tilde{\mathbf{D}}_{mm} + \sum_i\sum_j \mathbf{C}_{\substack{mi\\mj}}\end{aligned} \tag{6.75}$$

$$\mathbb{E}[\mathbf{L}(k)^T\mathbf{L}(k)]_{mn} = -2p_{mn} - p_{mn}\left(\sum_i p_{mi} + \sum_i p_{ni}\right) + \sum_i p_{im}p_{in} + 2p_{mn}^2 - \sum_i \mathbf{C}_{\substack{mi\\mn}} - \sum_i \mathbf{C}_{\substack{ni\\nm}} + \sum_i \mathbf{C}_{\substack{im\\in}}$$

$$= -2\mathbf{P}_{mn} - [\bar{\mathbf{D}}\mathbf{P}]_{mn} - [\mathbf{P}\bar{\mathbf{D}}]_{mn} + [\mathbf{P}^2 - \tilde{\mathbf{D}}]_{mn} + 2[\mathbf{P}\odot\mathbf{P}]_{mn} - \sum_i \mathbf{C}_{\substack{mi\\mn}} - \sum_i \mathbf{C}_{\substack{ni\\nm}} + \sum_i \mathbf{C}_{\substack{im\\in}} \tag{6.76}$$

where $\tilde{\mathbf{D}}$ is a diagonal matrix with $\{ii\}^{th}$ entry $\tilde{\mathbf{D}}_{ii} = \left[(\mathbf{P}\odot\mathbf{P})\mathbf{1}\right]_i$ subtracted also in (6.76) to compensate for the contribution of $\mathbf{P}^2$ in the main diagonal. Combining the results from (6.75) and (6.76) we obtain

$$\mathbb{E}\left[\mathbf{L}(k)^T\mathbf{L}(k)\right] = (\bar{\mathbf{D}}-\mathbf{P})^2 + 2\left(\bar{\mathbf{D}}-\mathbf{P}-\tilde{\mathbf{D}}+\mathbf{P}\odot\mathbf{P}\right)+\mathbf{R}' \tag{6.77}$$

where we have arranged the correlation terms in the matrix $\mathbf{R}'$ as follows

$$\begin{aligned}\mathbf{R}'_{mm} &= \mathbf{g}_m^T\mathbf{C}\,\mathbf{g}_m\\ \mathbf{R}'_{mn} &= \sum_{i=1}^{N}\mathbf{e}_{in}^T\mathbf{C}\,\mathbf{e}_{im} - \mathbf{g}_n^T\mathbf{C}\,\mathbf{e}_{nm} - \mathbf{g}_m^T\mathbf{C}\,\mathbf{e}_{mn}\end{aligned}$$

with $\mathbf{g}_m$ and $\mathbf{e}_{mn}$ defined in (6.28). Finally, defining $\tilde{\mathbf{L}} = \tilde{\mathbf{D}} - \mathbf{P}\odot\mathbf{P}$ and replacing (6.77) in (6.72) we obtain the closed-for expression

$$\mathcal{W} = \left(\bar{\mathbf{L}}^2 + 2(\bar{\mathbf{L}}-\tilde{\mathbf{L}}) + \mathbf{R}'\right)\epsilon^2 - 2\bar{\mathbf{L}}\epsilon + \mathbf{I} - \mathbf{J}_N$$

which completes the proof.

6.B Appendix: Proof of Theorem 6.4

For the directed case, recall that replacing (6.3) in (6.12) we obtain

$$\mathcal{W} = \mathbb{E}\left[\mathbf{L}(k)^T(\mathbf{I}-\mathbf{J}_N)\mathbf{L}(k)\right]\epsilon^2 - 2\bar{\mathbf{L}}\epsilon + \mathbf{I} - \mathbf{J}_N \tag{6.78}$$

and note that

$$\mathbb{E}\left[\mathbf{L}(k)^T(\mathbf{I}-\mathbf{J}_N)\mathbf{L}(k)\right] = \mathbb{E}\left[\mathbf{L}(k)^T\mathbf{L}(k)\right] - \mathbb{E}\left[\mathbf{L}(k)^T\mathbf{J}_N\mathbf{L}(k)\right] \tag{6.79}$$

Although the matrix $\mathbf{L}(k)$ is non-symmetric for all k, we can use the computations for the diagonal and the non-diagonal entries in (6.73) and (6.74), obtaining (6.75) and (6.76) respectively. In other words, we can compute the second matrix to the right-hand side of the expression in (6.79) and combine it with the results of Theorem 6.2. Since $\mathbf{J}_N = \mathbf{J}/N$, we use $\mathbf{J}$ instead for simplicity such that

$$\begin{aligned}[\mathbf{L}(k)^T\mathbf{J}\mathbf{L}(k)]_{mm} &= \left(\sum_i a_{mi}\right)^2 + \left(\sum_i a_{im}\right)^2 - \sum_i\sum_j a_{mi}a_{jm} - \sum_i\sum_j a_{im}a_{mj}\\ &= \sum_i a_{mi}^2 + \sum_i a_{im}^2 + \sum_i\sum_{j\neq i} a_{mi}a_{mj} + \sum_i\sum_{j\neq i} a_{im}a_{jm} - \sum_i\sum_j a_{mi}a_{jm} - \sum_i\sum_j a_{im}a_{mj}\end{aligned}$$

$$[\mathbf{L}(k)^T\mathbf{J}\mathbf{L}(k)]_{mn} = \sum_i\sum_j a_{mi}a_{nj} - \sum_i\sum_{j\neq m} a_{im}a_{nj} - \sum_i a_{im}a_{nm} - \sum_{i\neq n} a_{mi}a_{mn} - a_{mn}^2 + \sum_i a_{im}a_{mn}$$
$$- \sum_i\sum_{j\neq m} a_{mi}a_{jn} + \sum_i\sum_{j\neq m} a_{im}a_{jn}$$
$$= \sum_i\sum_j a_{mi}a_{nj} - \sum_i\sum_{j\neq m} a_{im}a_{nj} - \sum_{i\neq n} a_{im}a_{nm} - a_{nm}^2 - \sum_{i\neq n} a_{mi}a_{mn} - a_{mn}^2$$
$$+ \sum_i a_{im}a_{mn} - \sum_i\sum_j a_{mi}a_{jn} + \sum_i a_{mi}a_{mn} + \sum_i\sum_j a_{im}a_{jn} - \sum_i a_{im}a_{mn}$$

where again $a_{mn} = [\mathbf{A}(k)]_{mn}$, $a_{mm} = 0$ for all m, and the sums are from $\{i,j\} = 1$ to N. Taking the expectation of the expressions above we obtain

$$\mathbb{E}[\mathbf{L}(k)^T\mathbf{J}\mathbf{L}(k)]_{mm} = 2\sum_i p_{mi} + \sum_i\sum_{j\neq i} p_{mi}p_{mj} + \sum_i\sum_{j\neq i} p_{im}p_{jm} - \sum_i\sum_j p_{mi}p_{jm} - \sum_i\sum_j p_{im}p_{mj}$$
$$+ \sum_i\sum_j \mathbf{C}_{\substack{mi\\mj}} + \sum_i\sum_j \mathbf{C}_{\substack{im\\jm}} - \sum_i\sum_j \mathbf{C}_{\substack{mi\\jm}} - \sum_i\sum_j \mathbf{C}_{\substack{im\\mj}}$$
$$\mathbb{E}[\mathbf{L}(k)^T\mathbf{J}\mathbf{L}(k)]_{mn} = \sum_i\sum_j p_{mi}p_{nj} - \sum_i\sum_j p_{mi}p_{jn} - \sum_i\sum_{j\neq m} p_{im}p_{nj} + \sum_i\sum_j p_{im}p_{jn} - \sum_{i\neq n} p_{im}p_{nm}$$
$$- \sum_{i\neq n} p_{mi}p_{mn} - p_{mn} - p_{nm} + \sum_i p_{mi}p_{mn} + \sum_i\sum_j \mathbf{C}_{\substack{mi\\nj}} + \sum_i\sum_j \mathbf{C}_{\substack{im\\jn}} - \sum_i\sum_j \mathbf{C}_{\substack{im\\nj}} - \sum_i\sum_j \mathbf{C}_{\substack{mi\\jn}}$$

where as before we have assumed that $\mathbf{C}_{\substack{ii\\qr}} = \mathbf{C}_{\substack{ij\\qq}} = \mathbf{C}_{\substack{ij\\ij}} = 0$, for all $\{i,j,q,r\} \in \{1,\cdots,N\}$ and considered that $\mathbf{P}$ is symmetric with zero diagonal entries. Rearranging terms yields

$$\mathbb{E}[\mathbf{L}(k)^T\mathbf{J}\mathbf{L}(k)]_{mm} = 2\sum_i p_{mi} - 2\sum_i p_{mi}^2 + \sum_i\sum_j \mathbf{C}_{\substack{mi\\mj}} + \sum_i\sum_j \mathbf{C}_{\substack{im\\jm}} - \sum_i\sum_j \mathbf{C}_{\substack{mi\\jm}} - \sum_i\sum_j \mathbf{C}_{\substack{im\\mj}}$$
$$= 2\bar{\mathbf{D}}_{mm} - 2\tilde{\mathbf{D}}_{mm} + \sum_i\sum_j \mathbf{C}_{\substack{mi\\mj}} + \sum_i\sum_j \mathbf{C}_{\substack{im\\jm}} - \sum_i\sum_j \mathbf{C}_{\substack{mi\\jm}} - \sum_i\sum_j \mathbf{C}_{\substack{im\\mj}} \quad (6.80)$$

$$\mathbb{E}[\mathbf{L}(k)^T\mathbf{J}\mathbf{L}(k)]_{mn} = -2p_{mn} + 2p_{mn}^2 + \sum_i\sum_j \mathbf{C}_{\substack{mi\\nj}} + \sum_i\sum_j \mathbf{C}_{\substack{im\\jn}} - \sum_i\sum_j \mathbf{C}_{\substack{im\\nj}} - \sum_i\sum_j \mathbf{C}_{\substack{mi\\jn}}$$
$$= -2\mathbf{P}_{mn} - [\mathbf{P}\odot\mathbf{P}]_{mn} + \sum_i\sum_j \mathbf{C}_{\substack{mi\\nj}} + \sum_i\sum_j \mathbf{C}_{\substack{im\\jn}} - \sum_i\sum_j \mathbf{C}_{\substack{im\\nj}} - \sum_i\sum_j \mathbf{C}_{\substack{mi\\jn}} \quad (6.81)$$

Finally, combining the results from (6.75), (6.76), (6.80) and (6.81) according to (6.79) and replacing $\mathbf{J}$ with $\mathbf{J}_N$ we obtain

$$\mathbb{E}\left[\mathbf{L}(k)^T(\mathbf{I}-\mathbf{J}_N)\mathbf{L}(k)\right] = (\bar{\mathbf{D}}-\mathbf{P})^2 + \frac{2(N-1)}{N}\left(\bar{\mathbf{D}}-\mathbf{P}-\tilde{\mathbf{D}}+\mathbf{P}\odot\mathbf{P}\right) + \mathbf{R} \quad (6.82)$$

where we have arranged the correlation terms in the matrix $\mathbf{R}$ as follows

$$\mathbf{R}_{mm} = \mathbf{g}_m^T\mathbf{C}\,\mathbf{g}_m + \frac{1}{N}\left(\mathbf{g}_m^T\mathbf{C}\,\mathbf{q}_m + \mathbf{q}_m^T\mathbf{C}\,\mathbf{g}_m - \mathbf{g}_m^T\mathbf{C}\,\mathbf{g}_m - \mathbf{q}_m^T\mathbf{C}\,\mathbf{q}_m\right)$$
$$= \mathbf{g}_m^T\mathbf{C}\,\mathbf{g}_m + \frac{1}{N}(\mathbf{g}_m^T - \mathbf{q}_m^T)\mathbf{C}\,(\mathbf{q}_m - \mathbf{g}_m)$$
$$\mathbf{R}_{mn} = \sum_{i=1}^N \mathbf{e}_{in}^T\mathbf{C}\,\mathbf{e}_{im} - \mathbf{g}_n^T\mathbf{C}\,\mathbf{e}_{nm} - \mathbf{g}_m^T\mathbf{C}\,\mathbf{e}_{mn} + \frac{1}{N}(\mathbf{g}_n^T - \mathbf{q}_n^T)\mathbf{C}\,(\mathbf{q}_m - \mathbf{g}_m)$$

with $\mathbf{g}_m$ and $\mathbf{e}_{mn}$ defined in (6.28), and $\mathbf{q}_m$ defined in (6.32). Replacing (6.82) in (6.78) and using the already defined $\tilde{\mathbf{L}}$ we obtain the closed-for expression

$$\mathcal{W} = \left(\bar{\mathbf{L}}^2 + \frac{2(N-1)}{N}(\bar{\mathbf{L}} - \tilde{\mathbf{L}}) + \mathbf{R}\right)\epsilon^2 - 2\bar{\mathbf{L}}\epsilon + \mathbf{I} - \mathbf{J}_N$$

which completes the proof.

6.C Appendix: Autoregressive Correlation Model

Let s denote the index for the communication link between node i and node j with probability p_{ij}, and t denote the index for the link between nodes m and n with probability p_{mn}. The correlation scheme considered for this simulation model has an autoregressive structure where ϱ_{st} denotes the correlation coefficient between links e_{ij} and e_{mn}. Let $\{Y_i\}$, $\{U_i\}$ and $\{Z_i\}$ be sets of independent r.v.'s with binomial distributions $\mathcal{B}(1,\psi)$, $\mathcal{B}(1,\upsilon)$ and $\mathcal{B}(1,\zeta_i)$ respectively, with $1 \leq i \leq |\mathcal{E}(k)|$, where $|\mathcal{E}(k)|$ is the number of directed links at time k. The r.v. X_i is defined as

$$X_1 = Y_1, \quad X_i = (1 - U_i)Y_i + U_i X_{i-1}, \quad (2 \leq i \leq |\mathcal{E}(k)|)$$

and $W_i = Z_i X_i$, where Z_i is distributed binomially $\mathcal{B}(1,\zeta_i)$ independently of all the other variables. Then, $\{W_i\}$ are correlated r.v.'s with mean

$$\mathbb{E}[W_i] = p_i = \zeta_i \psi$$

and variance

$$\mathrm{var}(W_i) = \zeta_i \psi (1 - \zeta_i \psi)$$

where

$$\psi = \max\{\mathbf{P}\}$$

is the maximum entry of the connection probability matrix. The correlation coefficients are given by

$$\varrho_{st} = \upsilon^{|t-s|}\zeta_s \zeta_t$$

and the entries of the $\mathbf{C}$ matrix for this model are therefore computed as in (6.71).

7

Conclusions and Open Problems

This PhD thesis has addressed the problem of distributed estimation in wireless sensor networks using consensus algorithms. The convergence to the average consensus has been studied for networks with fixed topology assuming quantization noise, whereas probabilistic convergence to a consensus has been studied for networks with random topologies where the information exchanged might exhibit spatial correlation.

A simple model to achieve the average consensus in a WSN with quantized information exchange among neighboring nodes has been proposed in Chapter 4, where the impact of quantization noise on the performance of the algorithm has been characterized studying the MSE of the state computed with respect to the average of the initial measurements. The analysis has been carried out considering both temporally uncorrelated and temporally correlated quantization error, while assumed independent among the nodes. Conversely to other models found in literature studying consensus with quantized data, the MSE of the state for the proposed model converges as time evolves. Closed-form expressions have been derived for the limit of the MSE and for an upper bound assuming temporally uncorrelated quantization error, as well as for the limit of the MSE assuming temporally correlated quantization error. These closed-form expressions depend on general network parameters and can be computed offline. Particularly, the upper bound might be useful in the design of the quantizer implemented by the nodes.

Some open problems regarding consensus with quantized information exchange includes considering more accurate models for the quantization error vector exploiting for instance spatial correlation, since as the states of the nodes approach a consensus, the error becomes correlated among the nodes. Moreover, the analysis in Chapter 4 can be extended for networks with other characteristics, like for instance directed communication links or random communication failures. Another interesting problem consists in finding a means to compute analytically the optimum link weight minimizing the convergence time of the MSE of the state, which would result in a reduction of the convergence time of the algorithm.

Convergence in the mean square sense to a consensus in networks with random topology and directed communication links has been studied in Chapter 5, where convergence in expectation to the mean average consensus has been also shown. The MSE of the state computed with respect to the statistical mean of the initial measurements is characterized analytically with knowledge of the probability of connection of the links and the statistics of the measurements.

From the analysis considering equally probable communication links, closed-form expressions for the MSE, for the asymptotic MSE, for the dynamic range of the link weights and for its optimum value have been derived. The impact of the number of nodes and the impact of the probability of connection on the asymptotic MSE have been studied as well.

From the analysis considering links with different probabilities of connection, an upper bound for the MSE which can be computed for any time instant, as well as its asymptotic behavior and its convergence conditions have been derived. A criterion to minimize the convergence time of the upper bound has been proposed, and a sufficient condition guaranteeing convergence of the upper bound has been provided. Although the upper bound differs from the empirical MSE, it can be used for the computation of a link weight that reduces the convergence time of the algorithm under these connectivity conditions.

A practical transmission scheme applying randomized transmission power to reduce the overall energy consumption of the network until convergence to a common value has been proposed. The nodes are programmed to transmit at each iteration using a different power level selected at random from a predefined set of values and independently of the other nodes. The set of possible transmission power levels may be discrete or continuous. The randomization of the instantaneous power results in a reduction of the consumption with respect to a fixed topology with the same average power consumption, such that the total energy required to reach a consensus is reduced and the network lifetime can be therefore lengthened.

The impact of additive noise and the impact of time-varying link weights in the convergence of the consensus algorithm, as well as the assumption of different link weights applied at every node in networks with random topologies have not been addressed in this PhD thesis and remain as open problems.

Almost sure convergence of the consensus algorithm in random networks with spatially correlated links is considered in Chapter 6, where convergence in expectation to the average consensus is also shown. A criterion establishing a sufficient condition for almost sure convergence to the agreement space is applied. This criterion, valid also for topologies with spatially uncorrelated links, is based on the spectral radius of a positive semidefinite matrix for which closed-form expressions are derived assuming uniform link weights. Expressions for directed topologies with spatial correlation, directed topologies with no spatial correlation, undirected topologies with spatial correlation and undirected topologies with no spatial correlation are provided. The closed-form expressions allow the computation of the optimum link weights minimizing the convergence time of an upper bound for the squared norm of the error vector. Convexity of the optimization problem is shown, as well as the existence of a dynamic range for the link weights. The general expressions derived subsume existing protocols found in literature, and are particularized for networks with equally probable links. The analytical results obtained for these particular cases show that spatial correlation is detrimental to the convergence rate of consensus algorithms in random topologies and reduces the dynamic range of the link weights.

The simulations show that the optimum link weight derived in Chapter 6 outperforms the one obtained in the mean square convergence analysis of Chapter 5. Nevertheless, the MSE expressions derived in Chapter 5 are still useful to characterize the deviation of the state with respect to the average consensus when the instantaneous links are directed.

An open problem regarding consensus in correlated random networks is the assumption of non-symmetric connection probability matrix. The expressions derived in Chapter 6 assume the uniform weights model with a symmetric expected Laplacian matrix. Recall that these assumptions are sufficient to ensure that the expected weight matrix is double-stochastic, a property that has been used in the derivation of the closed-form expressions. However, the minimization of the spectral radius of the positive semidefinite matrix considered does not require a symmetric expected Laplacian. Additionally, the closed-form expressions derived in Chapter 6 could be extended to more general models of expected weight matrices. The effect of time-varying link weights is also an interesting issue not covered by the work conducted in this PhD thesis.

References

[Aba10] N. Abaid, and M. Porfiri, "Consensus over numerosity-constrained random networks", *IEEE Trans. on Automatic Control*, vol. PP, no. 99, pp. 649–654 , March 2010.

[Aky02] I.F. Akyildiz, Weilian Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks", *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[Ald05] S.A. Aldosari, and J.M.F. Moura, "Distributed Detection in Sensor Networks: Connectivity Graph and Small World Networks", *Record of the Thirty-Ninth Asilomar Conf. on Signals, Systems and Computers*, pp. 230–234, October 28 - November 1, 2005.

[Ald06] S.A. Aldosari, and J.M.F. Moura, "Topology of Sensor Networks in Distributed Detection", *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP'06)*, vol. 5, May 2006.

[Ays07a] T. C. Aysal, M. Coates, and M. Rabbat, "Distributed average consensus using probabilistic quantization", *IEEE 14th Workshop on Statistical Signal Processing, SSP '07*, pp. 640–644, 26-29 Aug. 2007.

[Ays07b] T. C. Aysal, M. Coates, and M. Rabbat, "Rates of convergence for distributed average consensus with probabilistic quantization", *45th Annual Allerton Conference, September 26-28*, 2007.

[Ays09] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast Gossip Algorithms for Consensus", *IEEE Trans. on Signal Processing*, vol. 57, no. 7, pp. 2748–2761, 2009.

[Ays10] T.C. Aysal, and K.E. Barner, "Convergence of consensus models with stochastic disturbances", *IEEE Trans. on Information Theory*, vol. 56, no. 8, pp. 4101 –4113, Aug. 2010.

[Bar05a] S. Barbarossa, "Self-organizing sensor networks with information propagation based on mutual coupling of dynamic systems", *Proc. of IWWAN, London, UK*, May 2005.

[Bar05b] S. Barbarossa, and F. Celano, "Self-Organizing sensor networks designed as a population of mutually coupled oscillators", *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC'05), New York, US*, June 2005.

[Bar06] S. Barbarossa, G. Scutari, and L. Pescosolido, "Global Stability of a Population of Mutually Coupled Oscillators Reaching Global ML Estimate Through a Decentralized Approach", *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP'06)*, vol. 4, May 2006.

[Bar07a] S. Barbarossa, and G. Scutari, "Decentralized Maximum-Likelihood Estimation for Sensor Networks Composed of Nonlinearly Coupled Dynamical Systems", *IEEE Trans. on Signal Processing*, vol. 55, no. 7, pp. 3456–3470, July 2007.

[Bar07b] S. Barbarossa, G. Scutari, and A. Swami, "Achieving Consensus in Self-Organizing Wireless Sensor Networks: The Impact of Network Topology on Energy Consumption", *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP'07)*, vol. 2, pp. II–841–II–844, 15-20 April 2007.

[Bel04] V. N. Belykh, I. V. Belykh, and M. Hasler, "Connection graph stability method for synchronized coupled chaotic systems", *Physica D: Nonlinear Phenomena*, vol. 195, no. 1-2, pp. 159–187, 2004.

[Ben97] K. Benjelloun, E. K. Boukas, and P. Shi, "Robust stochastic stability of discrete-time linear systems with markovian jumping parameters", *Proc. 36th IEEE Conf. Decision and Control*, vol. 1, pp. 559–564, 1997.

[Ber97] D. P. Bertsekas, and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice Hall, 1997.

[Blo05] V.D. Blondel, J.M. Hendrickx, A. Olshevsky, and J.N. Tsitsiklis, "Convergence in multiagent coordination, consens us, and flocking", *CDC-ECC Proc. 44th IEEE Conf. on Decision and Control and European Control Conference*, pp. 2996 – 3000, Dec. 2005.

[Bol01] Béla Bollobás, *Random graphs*, Cambidge Univ. Press, 2nd ed., 2001.

[Bor82] V. Borkar, and P. Varaiya, "Asymptotic agreement in distributed estimation", *IEEE Trans. on Automatic Control*, vol. 27, no. 3, pp. 650–655, jun. 1982.

[Boy04] Stephen Boyd, and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[Boy05] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Gossip algorithms: design, analysis and applications”, *Proc. 24th Annual Joint Conf. of the IEEE Computer and Communications Societies INFOCOM*, vol. 3, pp. 1653–1664, 2005.

[Boy06] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms”, *IEEE Trans. on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[Cal07] Guido Caldarelli, *Scale-Free Networks*, Oxford University Press, 2007.

[Cho03] Chee-Yee Chong, and S.P. Kumar, “Sensor networks: evolution, opportunities, and challenges”, *Proc. of the IEEE*, vol. 91, no. 8, pp. 1247–1256, Aug. 2003.

[Chu97] F.R.K. Chung, *Spectral graph theory*, American Mathematical Society, 1997.

[Clo07] P. Closas, E. Calvo, J.A. Fern·ndez-Rubio, and A. Pagès-Zamora, “Coupling noise effect in self-synchronizing wireless sensor networks”, *Proc. IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC’07)*, June 2007.

[Cyb89] G. Cybenko, “Dynamic load balancing for distributed memory multiprocessors”, *J. Parallel Distrib. Comput.*, vol. 7, pp. 279–301, October 1989.

[Den08] P. Denantes, F. Benezit, P. Thiran, and M. Vetterli, “Which distributed averaging algorithm should i choose for my sensor network?”, *Proc. 27th IEEE Conf. on Computer Communications INFOCOM*, pp. 986–994, 2008.

[Die97] R. Diekmann, S. Muthukrishnan, and M. V. Nayakkankuppam, “Engineering diffusive load balancing algorithms using experiments”, *Proc. of the 4th International Symposium on Solving Irregularly Structured Problems in Parallel*, vol. 1253 of *Lecture Notes in Computer Science*, pp. 111–122, Springer Berlin / Heidelberg, 1997.

[Dim08] A. D. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, “Geographic gossip: Efficient averaging for sensor networks”, *IEEE Trans. on Signal Processing*, vol. 56, no. 3, pp. 1205–1216, 2008.

[Erd60] P. Erdős, and A. Rényi, “On the evolution of random graphs”, *Acta Mathematica Academiae Scientiarum*, vol. 5, pp. 17–61, 1960.

[Fag08] F. Fagnani, and S. Zampieri, “Randomized consensus algorithms over large scale networks”, *IEEE Journal on Selected Areas In Communications*, vol. 26, no. 4, pp. 634–649, 2008.

[Fan94] Y. Fang, K.A. Loparo, and X. Feng, “Inequalities for the trace of matrix product”, *IEEE Trans. on Automatic Control*, vol. 39, no. 12, pp. 2489–2490, Dec. 1994.

[Fan05] L. Fang, and P.J. Antsaklis, “Information consensus of asynchronous discrete-time multi-agent systems”, *Proc. of the American Control Conference*, vol. 1-7, pp. 1883–1888, 2005.

[Fax04] J.A. Fax, and R.M. Murray, "Information flow and cooperative control of vehicle formations", *IEEE Trans. on Automatic Control*, vol. 49, no. 9, pp. 1465–1476, Sept. 2004.

[Fen92] X. Feng, K. A. Loparo, Y. Ji, and H. J. Chizeck, "Stochastic stability properties of jump linear systems", *IEEE Trans. on Automatic Control*, vol. 37, no. 1, pp. 38–53, 1992.

[Fie73] M. Fiedler, *Algebraic connectivity of graphs*, vol. 23, Czechoslovak Mathematical Journal, 1973.

[Fra08] P. Frasca, R. Carli, F. Fagnani, and S. Zampieri, "Average consensus by gossip algorithms with quantized communication", *Proc. 47th IEEE Conference on Decision and Control CDC 2008*, pp. 4831–4836, 2008.

[Gir05] A. Giridhar, and P.R. Kumar, "Computing and communicating functions over sensor networks", *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 755–764, April 2005.

[Gir06] A. Giridhar, and P.R. Kumar, "Toward a theory of in-network computation in wireless sensor networks", *IEEE Communications Magazine*, vol. 44, no. 4, pp. 98–107, April 2006.

[God01] C. Godsil, and G. Royle, *Algebraic graph theory*, vol. 207, Graduate Texts in Mathematics. Berlin, Germany: Springer-Verlag, 2001.

[Hat04] Y. Hatano, and M. Mesbahi, "Agreement over random networks", *CDC. 43rd IEEE Conf. on Decision and Control*, vol. 2, pp. 2010–2015, Dec. 2004.

[Hat05] Y. Hatano, A.K. Das, and M. Mesbahi, "Agreement in presence of noise: pseudogradients on random geometric networks", *CDC-ECC Proc. 44th IEEE Conf. on Decision and Control and European Control Conference*, pp. 6382–6387, 2005.

[Hib96] J. L. Hibey, "Stochastic stability theory for systems containing interval matrices", *IEEE Trans. on Aerospace and Electronic Systems*, vol. 32, no. 4, pp. 1385–1391, 1996.

[Hon05] Yao-Win Hong, and A. Scaglione, "A scalable synchronization protocol for large scale sensor networks and its applications", *IEEE Journal on Selected Areas in Comm.*, vol. 23, no. 5, pp. 1085–1099, 2005.

[Hoo06] S. Hoory, N. Linial, and A. Widgerson, "Expander graphs and their applications", *Bulletin of the American Mathematical Society*, vol. 43, no. 4, pp. 439–531, Oct. 2006.

[Hor06] R.A. Horn, and C.R. Johnson, *Matrix analysis*, Cambridge University Press, 2006.

[Jad03] A. Jadbabaie, Jie Lin, and A.S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules", *IEEE Trans. on Automatic Control*, vol. 48, no. 6, pp. 988–1001, June 2003.

[Jak10] D. Jakovetić, J. Xavier, and J. M. F. Moura, "Weight optimization for consensus algorithms with correlated switching topology", *IEEE Trans. on Signal Processing*, vol. 58, no. 7, pp. 3788–3801, July 2010.

[Kar06] Soummya Kar, and Jose M. F. Moura, "Topology for Global Average Consensus", *ACSSC '06. Fortieth Asilomar Conf. on Signals, Systems and Computers*, pp. 276–280, Oct.-Nov. 2006.

[Kar07] S. Kar, and J.M.F. Moura, "Distributed Average Consensus in Sensor Networks with Random Link Failures", *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP'07)*, vol. 2, pp. II–1013–II–1016, 15-20 April 2007.

[Kar08a] S. Kar, and J. M. F. Moura, "Distributed average consensus in sensor networks with quantized inter-sensor communication", *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP'08)*, pp. 2281–2284, April 2008.

[Kar08b] S. Kar, and J.M.F. Moura, "Sensor Networks With Random Links: Topology Design for Distributed Consensus", *IEEE Trans. on Signal Processing*, vol. 56, no. 7, pp. 3315–3326, 2008.

[Kar09] S. Kar, and J.M.F. Moura, "Distributed Consensus Algorithms in Sensor Networks With Imperfect Communication: Link Failures and Channel Noise", *IEEE Trans. on Signal Processing*, vol. 57, no. 1, pp. 355–369, 2009.

[Kar10] S. Kar, and J.M.F. Moura, "Distributed Consensus Algorithms in Sensor Networks: Quantized Data and Random Link Failures", *IEEE Trans. on Signal Processing*, vol. 58, no. 3, pp. 1383–1400, 2010.

[Kas06] A. Kashyap, T. Basar, and R. Srikant, "Quantized consensus", *IEEE International Symposium on Information Theory*, pp. 635–639, July 2006.

[Kur03] Y. Kuramoto, *Chemical oscillations, waves, and turbulence*, Chemistry Series, Dover Publications, 2003.

[Kus71] Harold Kushner, *Introduction to Stochastic Control*, New York, 1971.

[Lin05] Z. Lin, B. Francis, and M. Maggiore, "Necessary and sufficient graphical conditions for formation control of unicycles", *IEEE Trans. on Automatic Control*, vol. 50, no. 1, pp. 121–127, 2005.

[Lin09] P. Lin, and Y. Jia, "Consensus of second-order discrete-time multi-agent systems with nonuniform time-delays and dynamically changing topologies", *Automatica*, vol. 45, no. 9, pp. 2154 – 2158, 2009.

[Lun98] A.D. Lunn, and S. J. Davies, "A note on generating correlated binary variables", *Biometrika*, vol. 85, no. 2, pp. 487–490, 1998.

[Lyn96] Nancy A. Lynch, *Distributed Algorithms*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1996.

[Mao94] Xuerong Mao, *Exponential Stability of Stochastic Differential Equations*, Marcel Dekker, New York, 1994.

[Mir90] R. E. Mirollo, and S. H. Strogatz, "Synchronization of pulse-coupled biological oscillators", *SIAM Journal on Applied Mathematics*, vol. 50, no. 6, pp. 1645–1662, December 1990.

[Moa06] C.C. Moallemi, and B. Van Roy, "Consensus propagation", *IEEE Trans. on Information Theory*, vol. 52, no. 11, pp. 4753–4766, Nov. 2006.

[Mor05] L. Moreau, "Stability of multiagent systems with time-dependent communication links", *IEEE Trans. on Automatic Control*, vol. 50, no. 2, pp. 169–182, Feb. 2005.

[Nob88] B. Noble, and J.W. Daniel, *Applied linear algebra*, Prentice-Hall, 3rd ed., 1988.

[Ols06] A. Olshevsky, and J. N. Tsitsiklis, "Convergence rates in distributed consensus and averaging", *Proc. 45th IEEE Conf. Decision and Control*, pp. 3387–3392, 2006.

[OS04] R. Olfati-Saber, and R.M. Murray, "Consensus problems in networks of agents with switching topology and time-delays", *IEEE Trans. on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, Sept. 2004.

[OS05] R. Olfati-Saber, "Ultrafast consensus in small-world networks", *Proc. of the American Control Conference*, vol. 4, pp. 2371–2378, 8-10 June 2005.

[OS06] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: algorithms and theory", *IEEE Trans. on Automatic Control*, vol. 51, no. 3, pp. 401 – 420, March 2006.

[OS07] R. Olfati-Saber, J.A. Fax, and R.M. Murray, "Consensus and Cooperation in Networked Multi-Agent Systems", *Proc. of the IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[Pap65] Athanasios Papoulis, *Probability, random variables, and stochastic processes*, Part I. McGraw-Hill, 1965.

[Pat07] S. Patterson, B. Bamieh, and A. El Abbadi, "Distributed average consensus with stochastic communication failures", *Proc. 46th IEEE Conf. Decision and Control*, pp. 4215–4220, 2007.

[Pen03] M. Penrose, *Random Geometric Graphs*, Oxford University Press, June 2003.

[Pes06] L. Pescosolido, S. Barbarossa, and G. Scutari, "Decentralized Detection and Localization Through Sensor Networks Designed As a Population of Self-Synchronizing Oscillators", *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP'06)*, vol. 4, May 2006.

[Pic07] G. Picci, and T. Taylor, "Almost sure convergence of random gossip algorithms", *Proc. 46th IEEE Conf. Decision and Control*, pp. 282–287, 2007.

[Por07] M. Porfiri, and D.J. Stilwell, "Consensus Seeking Over Random Weighted Directed Graphs", *IEEE Trans. on Automatic Control*, vol. 52, no. 9, pp. 1767–1773, 2007.

[Por08] M. Porfiri, D. J. Stilwell, and E. M. Bollt, "Synchronization in Random Weighted Directed Networks", *IEEE Trans. on Circuits and Systems*, vol. 55, no. 10, pp. 3170–3177, Nov. 2008.

[Pre10] V. M. Preciado, A. Tahbaz-Salehi, and A. Jadbabaie, "Variance analysis of randomized consensus in switching directed networks", *Proc. of the American Control Conference*, pp. 6330–6335, 2010.

[Rab04] M. Rabbat, and R. Nowak, "Distributed optimization in sensor networks", *Third International Symposium on Information Processing in Sensor Networks, (IPSN'04).*, pp. 20–27, 26-27 April 2004.

[Rab05] M.G. Rabbat, R.D. Nowak, and J.A. Bucklew, "Generalized consensus computation in networked systems with erasure links", *Proc. IEEE 6th Workshop on Signal Processing Advances in Wireless Communications (SPAWC'05)*, pp. 1088–1092, June 2005.

[Ren05a] W. Ren, and R.W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies", *IEEE Trans. on Automatic Control*, vol. 50, no. 5, pp. 655–661, May 2005.

[Ren05b] W. Ren, R.W. Beard, and E.M. Atkins, "A survey of consensus problems in multiagent coordination", *Proc. of the American Control Conference*, vol. 3, pp. 1859–1864, June 2005.

[Sch04] D.S. Scherber, and H.C. Papadopoulos, "Locally constructed algorithms for distributed computations in ad-hoc networks", *Third International Symposium on Information Processing in Sensor Networks (IPSN'04)*, pp. 11–19, 26-27 April 2004.

[Sch08] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in Ad Hoc WSNs With Noisy Links—Part I: Distributed Estimation of Deterministic Signals", *IEEE Trans. on Signal Processing*, vol. 56, no. 1, pp. 350–364, 2008.

[Scu06] G. Scutari, S. Barbarossa, and L. Pescosolido, "Optimal decentralized estimation through self-synchronizing networks in the presence of propagation delays", *IEEE 7th Workshop on Signal Processing Advances in Wireless Communications(SPAWC'06)*, pp. 1–5, July 2006.

[Sil08] S. Silva Pereira, and A. Pagès-Zamora, "Distributed consensus in wireless sensor networks with quantized information exchange", *Proc. IEEE 9th Workshop on Signal Processing Advances in Wireless Communications (SPAWC'08)*, pp. 241–245, July 2008.

[Sil09a] S. Silva Pereira, and A. Pagès-Zamora, "Fast mean square convergence of consensus algorithms in WSNs with random topologies", *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP'09)*, pp. 2213–2216, April 2009.

[Sil09b] S. Silva Pereira, and A. Pagès-Zamora, "Randomized transmission power for accelerated consensus in asymmetric WSNs", *Proc. 3rd IEEE International Workshop* *on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP'09)*, pp. 348–351, Dec. 2009.

[Sil10a] S. Silva Pereira, and A. Pagès-Zamora, "Mean square convergence of consensus algorithms in random WSNs", *IEEE Trans. on Signal Processing*, vol. 58, no. 5, pp. 2866–2874, May 2010.

[Sil10b] S. Silva Pereira, S. Barbarossa, and A. Pagès-Zamora, "Consensus for distributed EM-based clustering in WSNs", *6th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM'10)*, pp. 45 –48, Oct. 2010.

[Sil11a] S. Silva Pereira, and A. Pagès-Zamora, "Consensus in random WSNs with correlated symmetric links", *Proc. 12th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC'11)*, pp. 136 –140, June 2011.

[Sil11b] S. Silva Pereira, and A. Pagès-Zamora, "When gossip meets consensus: convergence in correlated random WSNs", *International Conference on Wireless Technologies for Humanitarian Relief (ACWR2011)*, December 2011.

[Sil11c] S. Silva Pereira, and A. Pagès-Zamora, "Consensus in correlated random wireless sensor networks", *IEEE Trans. on Signal Processing*, vol. 59, no. 12, Dec. 2011.

[Spa05] D. P. Spanos, R. Olfati-Saber, and R. Murray, "Dynamic consensus for mobile networks", *Proc. of the 16th IFAC World Congress*, Elsevier, July 2005.

[TS08] A. Tahbaz-Salehi, and A. Jadbabaie, "A Necessary and Sufficient Condition for Consensus Over Random Networks", *IEEE Trans. on Automatic Control*, vol. 53, no. 3, pp. 791–795, 2008.

[Tsi84] J. N. Tsitsiklis, *Problems in decentralized decision making and computation*, PhD Thesis, Massachusetts Institute of Technology, Cambrigde, MA, 1984.

[Tsi86] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms", *IEEE Trans. on Automatic Control*, vol. 31, no. 9, pp. 803–812, 1986.

[Vic95] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel type of phase transition in a system of self-driven particles", *Phys. Rev. Lett.*, vol. 75, no. 6, pp. 1226–1229, Aug 1995.

[Wat98] D. J. Watts, and S. H. Strogatz, "Collective dynamics of 'small-world' networks", *Nature*, vol. 393, no. 6684, pp. 440–442, June 1998.

[Wol63] J. Wolfowitz, "Products of indecomposable, aperiodic, stochastic matrices", *Proc. Am. Math. Soc.*, vol. 15, pp. 733–736, 1963.

[Wu05] C. W. Wu, "Agreement and consensus problems in groups of autonomous agents with linear dynamics", *IEEE International Symposium on Circuits and Systems, ISCAS.*, pp. 292–295Vol.1, 23-26 May 2005.

[Wu06] C. W. Wu, "Synchronization and convergence of linear dynamics in random directed networks", *IEEE Trans. on Automatic Control*, vol. 51, no. 7, pp. 1207–1210, 2006.

[Xia03] L. Xiao, and S. Boyd, "Fast linear iterations for distributed averaging", *Proc. 42nd IEEE Conf. on Decision and Control*, vol. 5, pp. 4997–5002, Dec. 2003.

[Xia05] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus", *Fourth International Symposium on Information Processing in Sensor Networks*, pp. 63–70, 15 April 2005.

[Xia06a] L. Xiao, S. Boyd, and S. Lai, "A space-time diffusion scheme for peer-to-peer least-squares estimation", *The Fifth International Conf. on Information Processing in Sensor Networks, IPSN.*, pp. 168 –176, July 2006.

[Xia06b] L. Xiao, S. Boyd, and S.J. Kim, "Distributed average consensus with least-mean-square deviation", *Proc. of the 17th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, pp. 2768–2776, July 2006.

[Xia06c] L. Xiao, S. Boyd, and S. Lall, "Distributed Average Consensus with Time-Varying Metropolis Weights, [online] Available: http://www.stanford.edu/∼boyd/papers/avg_metropolis.html", 2006.

[Yil07] M.E. Yildiz, and A. Scaglione, "Differential nested lattice encoding for consensus problems", *6th International Symposium on Information Processing in Sensor Networks, IPSN*, pp. 89–98, 25-27 April 2007.

[Zha03] J. Zhao, R. Govindan, and D. Estrin, "Computing aggregates for monitoring wireless sensor networks", *Proc. of the First IEEE International Workshop on Sensor Network Protocols and Applications*, pp. 139–148, May 2003.

[Zho09] J. Zhou, and Q. Wang, "Convergence speed in distributed consensus over dynamically switching random networks", *Automatica*, vol. 45 Issue 6, pp. 1455–1461, June 2009.