THE INSTITUTE FOR SYSTEMS RESEARCH

ISR TECHNICAL REPORT 2015-2

A General Framework for Consenses Networks

Christoforos Somarakis

The Institute for Systems Research

A. JAMES CLARK SCHOOL OF ENGINEERING

ISR develops, applies and teaches advanced methodologies of design and analysis to solve complex, hierarchical, heterogeneous and dynamic problems of engineering technology and systems for industry and government.

ISR is a permanent institute of the University of Maryland, within the A. James Clark School of Engineering. It is a graduated National Science Foundation Engineering Research Center.

www.isr.umd.edu

Technical Report submitted to
INSTITUTE FOR SYSTEMS RESEARCH
Digital Repository at the University of Maryland
Copyrighted Material **doi:10.13016**, Number **0M2WK77**, March **2015**

Website: http://drum.lib.umd.edu/

pp. **1–28**

A GENERAL FRAMEWORK FOR CONSENSUS NETWORKS

Christoforos Somarakis[1] and John S. Baras[2]
The Institute For Systems Research
Univerisity Of Maryland
College Park, MD 20740, USA

Abstract. A novel approach for the study of the long-term behavior of solutions of continuous time consensus networks is developed. We establish sufficient conditions for convergence to a common equilibrium, under the mildest possible connectivity conditions, with emphasis on explicit estimates on the rate of convergence. The discussion ranges from linear to non-linear, decentralized to leader-follower as well as ordinary and time-delayed versions. Our work extends and unifies past works in the literature. Examples and simulations are presented to support the theoretical results.

1. **Introduction.** Dynamics of autonomous agents that exchange information over an abstract communication network is a central topic in the Applied Science. From Mathematical Biology & Pharmacokinetics to Engineering & Robotics, Social Sciences & Economics the scientific community is persistently interested in the evolution of interconnected systems and the global patterns that emerge out of local interactions (see [1, 2, 5, 6, 8, 14, 15, 16, 17, 26, 28, 30, 34, 37, 41, 42] and references therein).

The common denominator of all the aforementioned works is the study of a certain co-operative dynamic algorithm and the standard objective is for the agents' states to converge to a common equilibrium state under some global connectivity conditions. The classic framework considers a finite number of agents labeled as $\{1, \dots, N\}$, each of which possesses a value of interest say $x_i \in \mathbb{R}$, $i = 1, \dots, N$ that it is updated under the following scheme:

$$\dot{x}_i(t) = \sum_j a_{ij}\big(x_j(t) - x_i(t)\big). \tag{1}$$

Here $a_{ij} \geq 0$ model the coupling weights with the non-negativeness to characterize the co-operative nature of the dynamics and the magnitude to characterize the effect of agent j on agent i.

2. **Related Literature & Contribution.** The research on systems like (1) has been particularly active over the past decade. Consensus systems are proved to serve as an appropriate mathematical abstraction of co-operative networks of individuals that seek to coordinate a state of interest out of purely local interaction. Such systems have been identified as abstract models in disparate scientific disciplines: One can come across consensus-based dynamic models in biological networks [8, 14, 42], teams of robots [26], synchronization of oscillators [3, 19, 22], opinion or other gossip social networks [17, 25]. It recently came to the authors' attention

Key words and phrases. Distributed Consensus, Linear and Nonlinear Networks, Time-Varying Delays, Switching Connectivity, Rate of Convergence.

[1]csomarak@umd.edu
[2]baras@umd.edu

the striking similarity of (1) to models studied in Mathematical Biology, known as compartmental systems [2, 8, 14, 16].

The subject of distributed consensus networks is very well documented. It varies from continuous to discrete time settings, linear to nonlinear, deterministic to stochastic, ordinary to functional [5, 9, 10, 15, 24, 28, 29, 30, 32, 33, 34]. The underlying theory distinguishes between symmetric connectivity weights, where spectral graph theory is preferred [20] or the asymmetric case where nonnegative matrix theory applies [36]. The main drawback with the latter framework is that it is incompatible with continuous time dynamics. On the other hand the main objective is to prove convergence of the proposed control consensus algorithms to a common value. The literature although extensive, it lacks explicit estimates of rates of convergence. To the best of our knowledge there are no such rates even for the general linear case under mild connectivity conditions [30]. This may not be enough in real-world applications. It is certainly not enough in types of systems, where the contraction rate is essential in proving convergence [5, 9, 10, 30].

On the other hand, the nonlinar case and the effect of delays in the rate of convergence for general linear and nonlinear systems is another issue yet to be investigated [28, 33]. In [27, 32], linear time invariant networks with constant delays are analyzed with frequency methods. The drawback of this approach is that it does not apply in time varying or nonlinear systems. In [31] discrete time linear systems with delays are considered. The discretization method downgrades the problem to finite dimensions as the delays are multiples of the step time, hence the state-space can be extended and the problem is essentially reformulated as an ordinary discrete time problem. This approach can fail to approximate certain nonlinear networks where the step size must depend on the system's parameters [5].

The contribution of this paper is three-fold. Firstly, we develop a rigorous theory for the continuous time distributed consensus systems with the use of standard tools from the Non-Negative Matrix Theory. We analyze the continuous time setting and derive asymptotic stability results of the general linear case with emphasis on the rate of convergence. Our model investigates types of non-uniform asymptotic stability conditions that depend on the connectivity regime, a special case of which is uniform (i.e. exponential) asymptotic stability.

Secondly, we consider the case of arbitrary time-dependent delays and we prove a general exponential stability result under the same mild communication scheme. Furthermore, we demonstrate, in the example section, that the framework can be stretched to incorporate unbounded delays. Next, via a stability in variation argument, the derived results cope with leader-follower communication schemes.

Third, we explain how increased connectivity among the agents leads to a simple proof of the consensus problem in the spirit of the framework developed in the preceding sections. Next, we extend the linear theory to three non-linear models. The first class is this of passive systems, used in synchronization problems [1, 19, 33] and the second is this of monotonic consensus networks introduced in [34] from the control community but have also been previously considered in the field of mathematical biology [2, 8, 14, 16]. We will show via elementary direct linearization arguments that the systems can generate solutions indistinguishable from linear models. The third class is this of flocking networks of Cucker-Smale type and we freely draw results from [38, 39] so as to highlight the importance of the developed framework to such networks.

Our ultimate objective is to demonstrate how a large variety of networks can be treated under a theory that can provide solid convergence results in a unified perspective.

2.1. **Organization of the paper.** In §3 the basic notation and preliminary theory is introduced. In §4 the linear model (6), together with the fundamental hypotheses is stated. In §5 we make a number of observations concerning the connectivity properties of graphs that correspond to non-negative matrices, we prove a few preliminary results on a vital adaptation of the coefficient of ergodicity and on the behavior of the solutions of our model. The analysis of the ordinary version of (6) is held in §6. We provide convergence results in the non-uniform sense as well as we show that uniform coupling weights imply exponential convergence. In §7, we elevate on the full version of (6) and provide exponential stability estimates. In §8, we illustrate the main results by examples and simulations and conclude by demonstrating how the theory could include unbounded delays. In §9 we take a digression and provide a very simple proof of the undelayed problem under increased connectivity assumption with the use of the contraction coefficient. Furthermore, we explain through a variational stability argument that the theory can be adapted to leader-follower scenarios. We conclude with applications to non-linear systems. A concluding discussion on a number of points that were touched and need clarification is held in §10. For the sake of readability, the most important proofs are placed in the Appendix section at the end. For the full version and the detailed proofs the reader is directed to [40].

3. **Notations & Definitions.** For any natural number $N \in \mathbb{N}$, V will denote the set $\{1, 2, \ldots, N\}$. The upper integer part of $c \in \mathbb{R}$ is denoted by $[c]$. The N-dimensional Euclidean space is denoted by $\mathbb{R}^N$ and any $\mathbf{x} \in \mathbb{R}^N$ is considered to be a column vector, unless otherwise stated. The spread of $\mathbf{x} \in \mathbb{R}^N$ is

$$S(\mathbf{x}) = \max_i x_i - \min_j x_j \tag{2}$$

This quantity is a pseudo-norm and will be used in the stability analysis. Denote by $\mathbb{1}$ the N-dimensional column vector with all entries equal to 1. Then $S(\mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbb{1}k$ for some $k \in \mathbb{R}$. The *agreement* or *consensus* space Δ is defined as the subset of $\mathbb{R}^N$ such as

$$\Delta = \big\{\mathbf{x} \in \mathbb{R}^N : x_1 = x_2 = \cdots = x_N\big\}.$$

Next, $||\cdot||_p$ defines the p-norm in $\mathbb{R}^N$ whereas we shall use the notation $||\cdot||_\infty$ for the maximum norm. By $\mathbb{I}$ we understand the $N \times N$ identity matrix. For $I \subset \mathbb{R}$, $L^1(I, \mathbb{R}^N)$ denotes the space of integrable functions defined on I and taking values in $\mathbb{R}^N$. Similarly $C^l(I, \mathbb{R}^N)$ denotes the space of functions with $l \geq 0$ continuous derivatives defined accordingly. For $\mathbf{x} \in L^1$ or $\mathbf{x} \in C^l$ we define the spread

$$S_I(\mathbf{x}) = \max_i \max_{s \in I} x_i(s) - \min_i \min_{s \in I} x_i(s) \tag{3}$$

which is a straightforward generalization of (2) as it serves as a pseudonorm with respect to the agreement functional subspace

$$\Delta_I = \big\{\mathbf{x} \in L^1(I, \mathbb{R}^N) : x_i(t) = x_j(t), t \in I\big\}. \tag{4}$$

For these linear spaces the norm is defined to be $|\boldsymbol{x}| = \sup_{t \in I} |\mathbf{x}(t)|$. Due to potential discontinuities in the systems' parameters, $\frac{d}{dt}$ or "$\cdot$" denote the right Dini derivative [11]. Also, $\delta(\cdot)$ will denote the delta function.

The delay with which agent i observes the state of j is $\tau_{ij} \in C^1([t_0,\infty),\mathbb{R}_+)$. For $t \geq t_0$ set $\tau(t) := \max_{ij} \tau_{ij}(t)$. Also a subset of interest in the real line is $I_t := [t-\tau(t), t]$. We will occasionally use the convenient notation $\lambda_{ij}(t) := t-\tau_{ij}(t)$ and $\lambda(t) = t-\tau(t)$ with the n^{th} composition as $\lambda^{(n)}(t) := \lambda\big(\lambda^{(n-1)}(t)\big)$.

3.1. **Elements of Non-Negative Matrix Theory** [12, 36]**.** By a *topological directed graph* $\mathbb{G}$ we understand the pair (V,E) where V is the set of vertices, $E = \{(i,j) : i,j \in V : i \neq j\}$ is the set of edges. The *degree* N_i of a vertex i is the number of adjacent edges to i. The graph $\mathbb{G}$ is *routed-out branching* if there exists a vertex $i \in V$ such that for any $j \neq i \in V$ there is a *path* of edges $(l_k, l_{k-1})|_{k=0}^m$ such that $l_0 = i$ and $l_m = j$. For two graphs $\mathbb{G}_1 = (V,E_1)$ and $\mathbb{G}_2 = (V,E_2)$, we say that $\mathbb{G}_1$ is a sub-graph of $\mathbb{G}_2$ if $E_1 \subset E_2$. Two vertices $i,j \in V$ *communicate* if there is a path form i to j and a path from j to i. A vertex is *essential* if whenever there is a path from i to j then there is a path from j to i. A vertex is called *inessential* if it is not essential. All essential vertices are divided into communication classes and all inessential vertices that communicate with at least one vertex may be divided into inessential classes such that all vertices within a class communicate. All such classes are *self-communicating*. Each remaining inessential vertex communicates with no vertices and individually forms an inessential class called *non self-communicating*. By $\mathcal{S}$ we denote the family of topological graphs with fixed N vertices and self-edges on every vertex and by $\mathcal{T} \subset \mathcal{S}$ the set of graphs each of which is routed-out branching.

A *non-negative* matrix $P = [p_{ij}]$ is such that $p_{ij} \geq 0$ for all i,j[3] and P is called *generalized stochastic*, or m-stochastic, if $\sum_j p_{ij} = m$ for all $i \in V$. The family of $m \geq 0$ stochastic matrices is denoted by $\mathcal{M}$. For $m = 1$ we have the standard stochastic matrix.

The properties of stochastic matrices and their products play a crucial role in the analysis to follow and the standard approach is through graph theory: Any non-negative (and in particular stochastic) matrix P can be represented as a graph $\mathbb{G}_P$ with its adjacency matrix A_P the elements of which satisfy the property $A_{ij} = 1 \Leftrightarrow P_{ij} \neq 0$. For two stochastic matrices P_1 and P_2, we write $P_1 \sim P_2$ if $\mathbb{G}_{P_1} = \mathbb{G}_{P_2}$. The main tool to study the contraction rate of stochastic matrices with respect to Δ is the coefficient of ergodicity which we will state for generalized stochastic matrices. Given an m-stochastic matrix $P = [p_{ij}]$ the non-negative number

$$\rho(P) = \frac{1}{2}\max_{i,j\in V}\sum_{s=1}^{N}|p_{is}-p_{js}| = m - \min_{i,j\in V}\sum_{s=1}^{N}\min\{p_{is},p_{js}\} \tag{5}$$

is *the generalized coefficient of ergodicity* of P.

Theorem 3.1. [12] *Let P be an m-stochastic matrix. Then for any $\mathbf{z} \in \mathbb{R}^N$,*

$$S(P\mathbf{z}) \leq \rho(P)S(\mathbf{z})$$

with equality for some $\mathbf{z}$.

The coefficient of ergodicity measures the averaging effect of stochastic matrices and it is the central concept behind many convergence results in linear consensus algorithms. Its history dates back to one of Markov's first papers [23] and in the literature there exist an abundance of similar tools (for a recent review we refer to [13]).

[3]Unless otherwise specified each matrix is supposed to be square and of dimension $N \times N$.

Traditionally, ρ applies to discrete time dynamics of the type $\mathbf{z}_{n+1} = P_n \mathbf{z}_n$ where P_n is stochastic for any $n \in \mathbb{Z}_+$. We will show how ρ can be extended to apply in general linear operators that act on $L^1(I, \mathbb{R}^N)$.

A non-negative matrix P is called *irreducible* if $\mathbb{G}_P$ consists of a single essential class and a stochastic matrix P is called *regular* if $\mathbb{G}_P$ is routed-out branching. Now, the classical theory studies products of stochastic matrices; a setting that is readily applicable to the discrete time case. In continuous time dynamics we will encounter the following type of products of matrices:

$$P_{p,h}(t) = \int_{I_{p+h}} C(t, s_1) \int_{I_{p+h-1}(s_1)} C(s_1, s_2) \cdots \int_{I_{p+1}(s_{h-1})} C(s_{h-1}, s_h)\, ds_h \dots ds_1$$

for some appropriate matrix functions $C(t_1, t_2)$ so that $P_{p,h}(t)$ is stochastic for every $p \geq 0, h \geq 1$.

Regardless the time setting, a crucial point is to ask for what elements p_{ij} of an m-stochastic matrix it holds that $\min_{i,j} \sum_s \min\{p_{is}, p_{js}\} > 0$. It can be easily verified that $\rho(P) < m$ if and only if P possesses at least one strictly positive column. An m-stochastic matrix P with the property that $\rho(P) < m$ is called *scrambling*. A standard result in the theory of products of stochastic matrices is that for a regular matrix P there is a power of it that makes it scrambling: i.e. $\exists\ \gamma \geq 1 : \ \rho(P^\gamma) < 1$ and from the sub-multiplicative property $P^t \to \mathbb{1}\mathbb{1}^T k$ for some $k \in \mathbb{R}$, as $t \to \infty$. The power of P that makes it scrambling is known as the *scrambling index* for which the symbols γ and σ are reserved.

4. The Model.

For fixed $t_0 \in \mathbb{R}$, we consider the initial value problem

$$i \in V : \begin{cases} \dot{x}_i(t) = \sum_j a_{ij}(t)\big(x_j(\lambda_{ij}(t)) - x_i(t)\big), & t \geq t_0 \\ x_i(t) = \phi_i(t), & t \in I_{t_0} \end{cases} \tag{6}$$

where $\boldsymbol{\phi} = \big(\phi_1(t), \dots, \phi_N(t)\big) \in L^1(I_{t_0}, \mathbb{R}^N)$ are given initial data a_{ij} are the connectivity weight at which vertex j affects i. By $d_i(t) := \sum_j a_{ij}(t)$ we understand the cumulative effect weight on i from the rest of the network at time $t \geq t_0$. Next, $\tau_{ij}(t)$ is the imposed delay with which i receives the state of j at time t. The solution of (6) $\mathbf{x}(t, t_0, \boldsymbol{\phi}) = \big(x_1(t), \dots x_N(t)\big)$ is an absolutely continuous vector valued function that is defined in $[\lambda(t_0), \infty)$ and it takes values in $\mathbb{R}^N$. This relaxation on the smoothness of $\mathbf{x}$ is considered in order to incorporate possible switching couplings and although it deviates from the classical theory it, however, does not affect either of the fundamental properties of existence or uniqueness, or the integral representation of the solutions [35].

The delays τ_{ij} are known as *propagation* delays. Another type is this of the *processing* delays and they regard the time it takes for agent i to process their own information. This feature is modeled with x_i on the right hand side of (6) to be time delayed as well. The dynamics in the event of processing delays are fundamentally different and may lead to instability [28]. This point will be revisited in §10.

4.1. Hypotheses.

Let us now state the assumptions to accompany (6).

Assumption 4.1. *$\forall\ i \neq j \in V$ and $I \subset [t_0, \infty)$ bounded, $a_{ij} \in L^1\big(I, [0, \infty)\big)$. Also $a_{ii} = 0$.*

Additionally, the weight functions a_{ij} are assumed to satisfy the dwelling time condition [15]:

Assumption 4.2. *For any $t \geq t_0$ there exists $\epsilon > 0$ independent of t such that $a_{ij}(t) \neq 0$ implies that there exist a neighborhood of t, $N(t) \subset [t_0, \infty)$ of length ϵ such that $a_{ij}(s) \geq f(s) > 0$ for any $s \in N(t)$ and $f \in L^1\big([t_0,\infty),(0,\infty)\big)$ is non-increasing.*

The above assumption allows a_{ij} to asymptotically vanish, extending the majority of current results, to non-uniform type of convergence to consensus. With this simple setup, we are allowed to establish explicit contraction estimates of the solution with respect to Δ that depend on the type of the switching signal.

The final assumption concerns the nature of the propagation delays:

Assumption 4.3. *For all $i \neq j$ it holds that $\tau_{ij} \in C^1([t_0,\infty),\mathbb{R}_+)$ so that $1 - \dot{\tau}_{ij}(t) > 0$ for $t \geq t_0$ and*

$$\sup_{t \geq t_0} \int_{\lambda(t)}^{t} a_{ij}(s)\, ds < \infty.$$

The last condition on Assumption 4.3 implies that whenever a_{ij} is bounded from below then τ is necessarily bounded from above.

5. **Preliminaries.**

In this section, we review a number of underlying results most of which are already known in the literature, yet a thorough discussion is essential for a consistent presentation of the general theory. The first result concerns a classification of graphs with respect to scrambling index and the crucial lower bound on the edges one needs to add on one graph to decrease its scrambling index. The second result is a novel extension of ρ and a generalization of Theorem 3.1. These tools set the ground for the study of the long-term behavior solutions of (6), an important property of which concludes this section.

5.1. **Graphs and Non-Negative Matrices.**

We recall the set $\mathcal{S}$ and its subset $\mathcal{T}$. Let $R = R(N)$ denote the cardinality of $\mathcal{T}$. Each member $\mathbb{G}_i$ of it, has a scrambling index γ_i. In fact $\mathcal{T}$ can be partitioned in such mutually disjoint subsets: $\mathcal{T} = \bigsqcup_v \mathcal{Y}_v$ so that for $\mathbb{G}_1 \in \mathcal{Y}_{z_1}$, $\mathbb{G}_2 \in \mathcal{Y}_{z_2}$, $z_1 \neq z_2$ if and only if $\gamma_{z_1} \neq \gamma_{z_2}$. Consequently, we can enumerate

$$1 = \gamma_0 < \gamma_1 < \cdots < \gamma_{\max} \leq \left[\frac{N}{2}\right]$$

For instance, $\mathcal{Y}_0$ is the subclass of routed-out branching graphs, each member $\mathbb{G}_{\mathcal{Y}_0}$ of which has scrambling index, $\gamma_0 = 1$, i.e. there exist i such that $[\mathbb{G}_{\mathcal{Y}_0}]_{ji} \in E_{\mathbb{G}_{\mathcal{Y}_0}}$ for all $j \in V$. Next we note that for any $\mathbb{G}_1, \mathbb{G}_2 \in \mathcal{T}$ with $\mathbb{G}_2$ being a sub-graph of $\mathbb{G}_1$, it holds that $\gamma_1 \leq \gamma_2$ and thus we deduce that by adding an edge to any graph, the scrambling index may only decrease. In particular, there exists a sufficient number of new edges that will decrease the scrambling index. Fix $j < i$. Then for any $\mathbb{G}_i \in \mathcal{Y}_i$ there exists a positive number $q_{i,j}$ such that the graph $\mathbb{G}_j$ formed out of $\mathbb{G}_i$ with $q_{i,j}$ additional edges will be a member of $\bigcup_{v=0}^{j} \mathcal{Y}_v$, in which case $\gamma_j \leq \gamma_i - 1$.

Remark 5.1. The minimum number of edges needed to be added on an arbitrary member of $\mathcal{Y}_i$ so that the resulting graph is a member of $\bigcup_{v=0}^{i-1} \mathcal{Y}_v$, is denoted by $l^* := \max_i \{q_{i,i-1}\}$.

5.2. **An extension of the coefficient of ergodicity.** Let us, now, extend Theorem 3.1 to the case where P acts as an abstract linear operator on members of L^1. For $B \geq 0$, $t \geq t_0 + B$ and $s \in [t-B, t]$ consider the matrix

$$C(t,s) = e^{-mB}\delta(s-(t-B))\mathbb{I} + e^{-m(t-s)}W(s)$$

with $W(s) = m\mathbb{I} - D(s) + A(s)$.

Proposition 5.2. *Let $m > \sup_{s\geq t_0} \max_{i\in V} d_i(s)$. Then for any $B > 0$, $l \geq 1$, $t \geq t_0 + B$ the matrix*

$$P_B^{(l)}(t) := \begin{cases} \int_{t-B}^{t} C(t,s)\, ds, & l = 1 \\ \int_{t-B}^{t} C(t,s) P_B^{(l-1)}(s)\, ds, & l > 1 \end{cases}$$

whenever defined, is stochastic.

Theorem 5.3. *Let I be a compact subset of $\mathbb{R}$ and assume that for any compact $I' \subset I$, $W_{I'} = \int_{s\in I'} P(s)\, ds \in \mathcal{M}$ and W_I is m-stochastic. If $\mathbf{w} = \int_{s\in I} P(s)\mathbf{z}(s)\, ds$, then*

$$S(\mathbf{w}) = \rho(W_I) S(\mathbf{z}^*)$$

for some $\mathbf{z}^* = \big(z_1(s_1), \dots z_N(s_N)\big)$ for $s_i \in I$ and

$$\begin{aligned} \rho(W_I) &= \frac{1}{2}\max_{h,h'} \sum_{k=1}^{N} \int_{s\in I} |p_{hk}(s) - p_{h'k}(s)|\, ds \\ &= m - \min_{h,h'} \sum_{k=1}^{N} \min\left\{ \int_{s\in I} p_{hk}(s)\, ds, \int_{s\in I} p_{h'k}(s)\, ds \right\} \end{aligned}$$

Remark 5.4. Similarly, for the expression

$$\mathbf{w} = \int_{s\in I_1} P_1(s) \int_{q\in I_2(s)} P_2(q)\mathbf{z}(q)\, dq ds$$

one can show, along the lines of the proof of Theorem 5.3 that if

$$W_I^{(2)} = \int_{s\in I_1} P_1(s) \int_{q\in I_2(s)} P_2(q)\, dq ds$$

is stochastic, then

$$S(\mathbf{w}) \leq \rho(W_I^{(2)}) S(\mathbf{z}^*)$$

for some $\mathbf{z}^* = \big(z_1(s_{(ij)}^{(1)}), z_2(s_{(ij)}^{(2)}), \dots, z_N(s_{(ij)}^{(N)})\big)$ all $s_{(ij)}^{(l)}$ of which are in $I_1 \cup I_2$.

5.3. **Bounds on $\mathbf{x}(t, t_0, \boldsymbol{\phi})$.**

Lemma 5.5. *Under Assumption 4.1 the solution $\mathbf{x} = \mathbf{x}\big(t, t_0, \boldsymbol{\phi}\big)$, $t \geq t_0$ of (6) satisfies*

$$\min_{j\in V} \min_{s\in I_{t_0}} \phi_j(s) \leq x_i(t) \leq \max_{j\in V} \max_{s\in I_{t_0}} \phi_j(s)$$

$\forall\ t \geq t_0, i \in V$.

Remark 5.6. It holds that $S_{I_{t_1}}(\mathbf{x}) \leq S_{I_{t_2}}(\mathbf{x})$ for any $t_1 \geq t_2 \geq t_0$

Remark 5.7. If $\tau(t) \equiv 0$ it holds that $S(\mathbf{x}(t_1)) \leq S(\mathbf{x}(t_2))$ for any $t_1 \geq t_2 \geq t_0$.

Another fact from the boundedness of the solutions is stated in the following result:

Lemma 5.8. *If $\mathbf{x} = \mathbf{x}(t, t_0, \boldsymbol{\phi})$, $t \geq t_0$ is the solution of (6) such that $S_{I_t}\big(\mathbf{x}\big) \to 0$ as $t \to \infty$, then the forward limit set $\omega(\boldsymbol{\phi})$ is a singleton in Δ_I.*

6. **Convergence rates of the undelayed version.** The un-delayed version of (6) is this with $\tau(t) \equiv 0$ and it reads

$$i \in V : \begin{cases} \dot{x}_i(t) = \sum_j a_{ij}(t)\big(x_j(t) - x_i(t)\big), & t \geq t_0 \\ x_i(t) = x_i^0, & t = t_0 \end{cases}$$

and in vector form

$$\begin{cases} \dot{\mathbf{x}}(t) = -L(t)\mathbf{x}(t), & t \geq t_0 \\ \mathbf{x}(t) = \mathbf{x}^0, & t = t_0 \end{cases} \tag{7}$$

where $L(t) = D(t) - A(t)$ is the time-varying laplacian matrix [26].

The first result is based on the additional property of f, as defined in Assumption 4.2 that it asymptotically vanishes. Then a (non-uniform) asymptotic stability result with respect to Δ can be provided, based on the rate at which f is allowed to vanish. The importance of this result is to be discussed in §10.

Theorem 6.1. *Let Assumptions 4.1 and 4.2 to hold with f as defined in Assumption 4.2 to satisfy $f(t) \to 0$ as $t \to \infty$. If there exists $B > 0$ and $M \geq t_0$ so that for any $t \geq M$ the graph $\mathbb{G}_{P_B(t)}$ that corresponds to $P_B(t)$ is routed-out branching, then unconditional asymptotic consensus for the solution of the system (7) is achieved if there exists a sequence $t_n \geq M$ with $t_{n+1} - t_n \geq \sigma B$, such that*

$$\sum_n f^\sigma(t_n) = \infty.$$

with $\sigma = l^*([N/2] + 1)$ and l^* with the meaning of Remark 5.1.

The rate of convergence is dictated by the non-summability of $\sum_n f^\sigma(t_n)$ and in the general case proves convergence to consensus in a non-uniform sense.

Remark 6.2. If, in addition, the connectivity is static (but with time-varying weights), then from the discussion in §5, we have $\sigma = \gamma_j$ for some $j = 0, \ldots, \max$, where γ_j is the scrambling index of $\mathbb{G}_{P_B(t)}$ and $B > 0$ can be chosen arbitrarily small.

A significantly more elegant (exponential) result is obtained if we take f to be uniformly lower bounded.

Corollary 6.3. *Let the conditions of Theorem 6.1 hold with $M = t_0$ and B, ϵ, σ the same parameters defined in its statement. If f as defined in Assumption 4.2 satisfies $f(t) \geq \underline{f} > 0$, $t \geq t_0$ then*

$$S(\mathbf{x}(t)) \leq \frac{S(\mathbf{x}^0)}{1-\rho} e^{-\theta(t-t_0)},$$

for $\theta = -\frac{\ln(1-\rho)}{\sigma B}$, $\rho = \min\left\{e^{-\sigma m B}, \left(\frac{(1-e^{-m\epsilon})\underline{f}}{m}\right)^\sigma\right\} \in (0,1)$, $m > \sup_{s \geq t_0} \max_{i \in V} d_i(s)$.

Corollary 6.3 is a direct application of Theorem 6.1. It unifies and extends previous results [15, 28, 30] by providing explicit rate estimates.

6.1. **Special Case: Simple Proof under Increased Connectivity.** We consider (7) and discuss a quick and elegant proof under the assumption of strong connectivity. We set

$$\rho(P) = m - \min_{i,j} \sum_{s=1}^{N} \min\{p_{is}, p_{js}\} =: m - \theta.$$

By Assumption 4.1, we pick $m > \sup_{t\geq t_0} \max_{i\in V} d_i(t)$ and we rewrite (7) as

$$\dot{\mathbf{x}} = -m\mathbb{I}\mathbf{x} + \big(m\mathbb{I} - D(t) + A(t)\big)\mathbf{x} \Rightarrow e^{-mt}\frac{d}{dt}\big(e^{mt}\mathbf{x}\big) = \big(m\mathbb{I} - D(t) + A(t)\big)\mathbf{x}$$

Now it is easy to check that $m\mathbb{I} - D(t) + A(t)$ is $m-$stochastic and it follows that

$$\begin{aligned}\frac{d}{dt}S\big(\mathbf{x}(t)\big) &= \frac{d}{dt}\Big(e^{-mt}S\big(e^{mt}\mathbf{x}(t)\big)\Big) = -mS\big(\mathbf{x}(t)\big) + e^{-mt}\frac{d}{dt}S\big(e^{mt}\mathbf{x}(t)\big)\\ &\leq -mS\big(\mathbf{x}(t)\big) + S\Big(e^{-mt}\frac{d}{dt}\big(e^{mt}\mathbf{x}(t)\big)\Big)\\ &\leq -mS\big(\mathbf{x}(t)\big) + \rho S\big(\mathbf{x}(t)\big)\\ &= -mS\big(\mathbf{x}(t)\big) + (m-\theta(t))S\big(\mathbf{x}(t)\big)\\ &= -\theta(t)S\big(\mathbf{x}(t)\big).\end{aligned}$$

Consequently,

$$S\big(\mathbf{x}(t)\big) \leq e^{-\int_{t_0}^{t}\theta(s)\,ds}S\big(\mathbf{x}^0\big)$$

so that $\int^{\infty}\theta(s)\,ds = \infty$ implies $S(\mathbf{x}(t)) \to 0$ and by Lemma 5.8 we have $\mathbf{x}(t) \to \Delta$.

This simple calculation generalizes recent results in the literature. Here not only one needs not to assume d_i to be upper bounded by one as in [30] but also one needs not assuming lower bounds on a_{ij} so that the convergence can be non-uniform.

The non-integrability of $\theta(t)$ requires as a minimum type of connectivity, this of a uniformly recurrent, transmission of signals at least from one agent to the rest of the agents, at the same time. This means that there must exist at least one agent to affect the rest of the group at the same time. This is however a hardly decentralized architecture.

6.2. **Necessary conditions.** Sufficient conditions for stability are usually much easier to obtain as opposed to necessary conditions that are very rare and consensus systems are not an exception. The related results focus on stochastic systems where the driving signal is especially designed to either enforce asymptotic consensus to a common state or asymptotic anti-consensus, that is, under certain initial data, convergence to at least two different states. [24]. At this point, we take a small digression to discuss necessary conditions for asymptotic consensus.

Theorem 6.4. *Consider the initial value problem (6) with $\tau \equiv 0$ and its solution $\mathbf{x}$. Let Assumption 4.1 for this system to hold and the communication graph to be routed-out branching. Assume also that over a population of N autonomous agents there is a partition $V_1, V_2 \subset V$ with $V = V_1 \sqcup V_2$ so that $(i,j) \in V_1 \times V_2$ implies $\int^{\infty} a_{ij}(s)\,ds < \infty$. If for any $l_1, l_2 \in V$, $x_{l_1}(t) - x_{l_2}(t) \to 0$ implies $|x_{l_1}(t) - x_{l_2}(t)| \leq \Gamma e^{-\gamma t}$ for some $\gamma, \Gamma > 0$ then there exist initial conditions such that $S\big(\mathbf{x}(t)\big) > 0$ for any $t \geq t_0$.*

The reader might observe a strong discrepancy between the necessary condition (i.e. the divergence of $\int^{\infty} a_{ij}(s)\,ds$) and the sufficient conditions discussed in §6 (i.e. the divergence of $\int^{\infty} a_{ij}^{\sigma}(s)\,ds$). This occurs due to the fact that ρ is a weak contraction estimate tool. This remark will be discussed in §10.

The exponential convergence assumption taken in the theorem above is a moderate condition that it can be dropped if the corresponding coupling weights are uniformly bounded away from zero as we explained above. In such case, one is allowed to combine the uniform convergence of solutions and the linearity of the system to conclude on the exponential rate.

7. **Convergence rates of the delayed version.** In this section we will investigate the stability of (6) in full and for the sake of simplicity we will assume that any non-zero coupling weight is uniformly bounded away from zero and that the delay magnitudes are arbitrary but bounded. The idea here, relies on the elementary observation that the solution $\mathbf{x} = \mathbf{x}(t, t_0, \boldsymbol{\phi})$, $t \geq t_0$ satisfies

$$x_i(t) = e^{-\int_{t'}^{t} d_i(s)ds} x_i(t') + \int_{t'}^{t} e^{-\int_{s}^{t} d_i(w)dw} \sum_j a_{ij}(s) x_j(\lambda_{ij}(s))\, ds. \tag{8}$$

We will investigate the rate at which $S_{I_t}(\mathbf{x})$ contracts by combining (8) with Theorem 5.3 and Lemma 5.5 assuming that there is no leader in the graph, i.e. all vertices are affected by at least one other vertex. The leader follower case is to be discussed in §9. Estimating the way the solution contracts over I_t, we conclude that the limit point must lie in $\Delta_{I_{t_0}}$ i.e. the solutions contract to a constant similar to the un-delayed case.

Theorem 7.1. *Let Assumptions 4.1, 4.2 and 4.3 hold such that $a_{ij}(t) \neq 0$ implies $a_{ij}(t) \geq f(t) > \underline{f}$, $t \geq t_0$ for some $\underline{f} > 0$. If there exists $B > 0$ so that for any $t \geq t_0$ the graph $\mathbb{G}_{P_B(t)}$ is routed-out branching, then unconditional asymptotic consensus for the solution of the system (6) is achieved. In particular, there exists $k \in [\min_{i\in V, s\in I_{t_0}} \phi_i(s), \max_{i\in V, s\in I_{t_0}} \phi_i(s)]$ such that*

$$||\mathbf{x}(t) - \mathbb{1}k||_\infty \leq \frac{S_{I_{t_0}}(\boldsymbol{\phi})}{1 - \mu e^{-\bar{N}a\tau}} e^{-\theta(t-t_0)}$$

where $\theta = -\frac{\ln(1-\mu e^{-\bar{N}a\tau})}{(\sigma(B+\tau)+\tau)}$, $\mu := \inf_{t\geq t_0} \min_{i,j} \sum_l \min\{p_{il}^\sigma(t), p_{jl}^\sigma(t)\} \in (0,1)$ with $p_{ij}^\sigma(t)$ the elements of $P_{B+\tau}^{(\sigma)}(t)$, $\sigma = l^*([N/2]+1)$, $\bar{N} = \max_{j\in V} N_j$ is the maximum degree over V, l^* has the meaning of Remark 5.1 and a is the upper bound of a_{ij} by virtue of Assumption 4.1.

Just like the ordinary algorithm and Corollary 6.3, Theorem 7.1 provides explicit estimates on the rate of the exponential convergence as a function of the parameters, the connectivity signal and the imposed delays. This is a delay-independent result and in the next section we will show that under Assumption 4.3 it can be extended to unbounded delays.

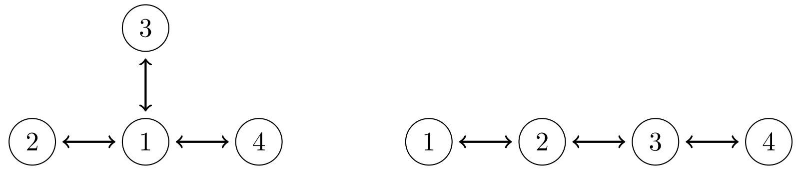
8. **Examples & Simulations.** In this section, we discuss a couple of illustrative examples. The first is a 4×4 linear network with linear switching coupling and bounded delays. The second is a 2×2 time-varying network with static connectivity and unbounded delays. All simulations were carried out in MATLAB with the `ddesd` routine.

8.1. **A 4×4 graph.** Let a network of $N = 4$ agents with communication weights $a_{ij}(t)$ where $i, j = 1, \ldots, 4$. We classify two communication schemes:

8.1.1. *Star topology.* Depicted in Figure 1(a), in this scheme the connectivity matrix reads

$$A_{\text{star}}(t) = \begin{bmatrix} 0 & a_{12}(t) & a_{13}(t) & a_{14}(t) \\ a_{21}(t) & 0 & 0 & 0 \\ a_{31}(t) & 0 & 0 & 0 \\ a_{41}(t) & 0 & 0 & 0 \end{bmatrix}$$

We assume here the switching (on/off) transmission signal to be defined as follows: There exist $B, \epsilon, \alpha > 0$ such that for any B interval of time, there exists a ϵ-subset



(A) Ex. 7.1.1: The star graph.

(B) Ex. 7.1.2: The path graph.

FIGURE 1. The communication topology in Example 7.1

so that $a_{ij}(s), a_{ji}(s) \geq \alpha > 0$ for $s \in [t, t+\epsilon]$. In this scenario, the analysis is very simple because the results of §?? applies: Take $d_i(t) = \sum_j a_{ij}(t)$ and $m > \sup_t \max_{i \in V} d_i(t)$. Then $W(t) := m\mathbb{I} - D(t) + A(t)$ is

$$W(t) = \begin{bmatrix} m - d_1(t) & a_{12}(t) & a_{13}(t) & a_{14}(t) \\ a_{21}(t) & m - d_2(t) & 0 & 0 \\ a_{31}(t) & 0 & m - d_3(t) & 0 \\ a_{41}(t) & 0 & 0 & m - d_4(t) \end{bmatrix}$$

and it is obviously scrambling during some ϵ interval over any B-interval of time. For m large enough the coefficient of ergodicity ρ is lower bounded by α for $s \in [t, t+\epsilon]$. This implies that for any $t > t_0 + B$, there exists $k \leq 0$ such that $t_0 + kB \leq t \leq t_0 + (k+1)B$ and by Lemma 5.5

$$S(\mathbf{x}(t)) \leq S(\mathbf{x}_0) e^{\alpha\epsilon} e^{-(k+1)\alpha\epsilon} \leq S(\mathbf{x}_0) e^{-\frac{\alpha\epsilon}{B}(t-t_0)}.$$

In the presence of delays, $\tau_{ij}(t) \leq \tau < \infty$ the discussion in §?? applies. We can easily calculate $\rho = \min\{e^{-mB}, \frac{1-e^{-m\epsilon}}{m}\alpha\}$ and by Theorem 7.1 exponential convergence is guaranteed with rate $\theta = -\frac{\ln(1-\rho e^{-3\alpha\tau})}{B+2\tau}$.

As a numerical example take $a_{12}(t) = 0.02u(t)$, $a_{13}(t) = 0.05u(t)$, $a_{14}(t) = 0.03u(t)$, $a_{21}(t) = 0.2(0.01 + e^{-t})u(t)$, $a_{31}(t) = 0.07u(t)$,$a_{41}(t) = 0.06u(t)$ for $u(t) = 2, t \in [n + 1/2, n + 1]$, $n \in \mathbb{N}$ and 0 otherwise. Then $\frac{\alpha\epsilon}{B} = 0.002$ and this is the estimated rate of convergence for the un-delayed system. See Fig. 2(a). In the case of delays we set $\tau_{12}(t) = 8 - 0.5\cos(t)$, $\tau_{13}(t) = 3 - 0.2\sin(2t)$, $\tau_{14}(t) = 9$, $\tau_{21}(t) = 10$ $\tau_{31}(t) = 5 - 0.9\sin(t/4)$, $\tau_{41}(t) = 2$ so that $\tau = 10$, $\rho = 0.0019$ and $\theta = 0.00008031$. See Fig. 2(b).

8.1.2. *Path topology.* Here the connectivity matrix is

$$A_{\text{path}}(t) = \begin{bmatrix} 0 & a_{12}(t) & 0 & 0 \\ a_{21}(t) & 0 & a_{23}(t) & 0 \\ 0 & a_{32}(t) & 0 & a_{34}(t) \\ 0 & 0 & a_{43}(t) & 0 \end{bmatrix}$$

Now we consider the switching signal to be defined as follows: For all $t \geq 0$ it holds that $a_{ij}(t) \neq 0 \Rightarrow 0 < \alpha \leq a_{ij}(t) < \frac{1}{2}$ and also

$$\begin{cases} a_{23}(t) = a_{32}(t) = a_{34}(t) = a_{43}(t) = 0 \ \& \ a_{12}(t), a_{21}(t) \neq 0, t \in [3l\epsilon, (3l+1)\epsilon) \\ a_{12}(t) = a_{21}(t) = a_{34}(t) = a_{43}(t) = 0 \ \& \ a_{23}(t), a_{32}(t) \neq 0, t \in [(3l+1)\epsilon, (3l+2)\epsilon) \\ a_{23}(t) = a_{32}(t) = a_{12}(t) = a_{21}(t) = 0 \ \& \ a_{34}(t), a_{43}(t) \neq 0, t \in [(3l+2)\epsilon, (3l+3)\epsilon) \end{cases}$$

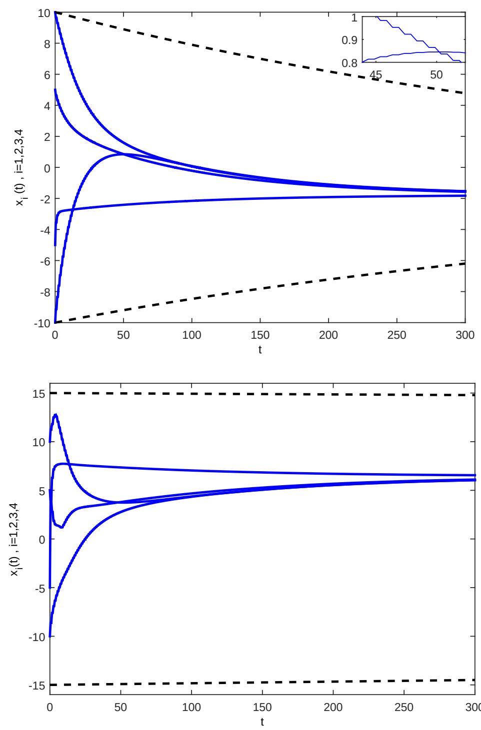


FIGURE 2. Example 7.1.1. (a): The convergence of 4 agents with a star graph topology. The detail on the upper right part is to denote the effect of the switching signal. (b): Convergence under the effect of delays. In both figures the dashed lines depict the theoretical rate estimates. It is remarked that in (b) the estimate is significantly weak.

for some fixed $\epsilon > 0$ and $l \in \mathbb{Z}_+$. Here $B = 3\epsilon$, $m = 1$ and

$$C(t,s) = \begin{bmatrix} \bar{d}_1(t,s) & e^{-(t-s)}a_{12}(s) & 0 & 0 \\ e^{-(t-s)}a_{21}(s) & \bar{d}_2(t,s) & e^{-(t-s)}a_{23}(s) & 0 \\ 0 & e^{-(t-s)}a_{32}(s) & \bar{d}_3(t,s) & e^{-(t-s)}a_{34}(s) \\ 0 & 0 & a_{43}(s)e^{-(t-s)} & \bar{d}_4(t,s) \end{bmatrix}$$

where $\bar{d}_i(t,s) = e^{-3\epsilon}\delta(s-(t-3\epsilon)) + e^{-(t-s)}(1-d_i(s))$. This is a non-scrambling matrix so the discussion in §**??** is of no use and we need to escalate to the general result using Corollary 6.3. Indeed applying this result we obtain

$$S(\mathbf{x}(t)) \le S(\mathbf{x}(3(l-1)\epsilon)) \le (1-2\alpha^2(1-e^{-\epsilon})^2)^{l-1}S(\mathbf{x}(0)) \le \frac{S(\mathbf{x}(0))}{1-2\alpha^2(1-e^{-\epsilon})^2}e^{-\theta t}$$

where $\theta := \frac{\ln(1-2\alpha^2(1-e^{-\epsilon})^2)}{3\epsilon}$.

Together with the switching signal, we now consider a common bounded propagation delay $0 \le \tau(t) \le \tau < \infty$. We apply Theorem 7.1 to estimate the rate of convergence as follows: $\rho = 2\alpha^2(1-e^{-\epsilon})^2$, $\sup_t \int_{\lambda(t)}^t d_i(s)\,ds \le \tau$ so that the rate of

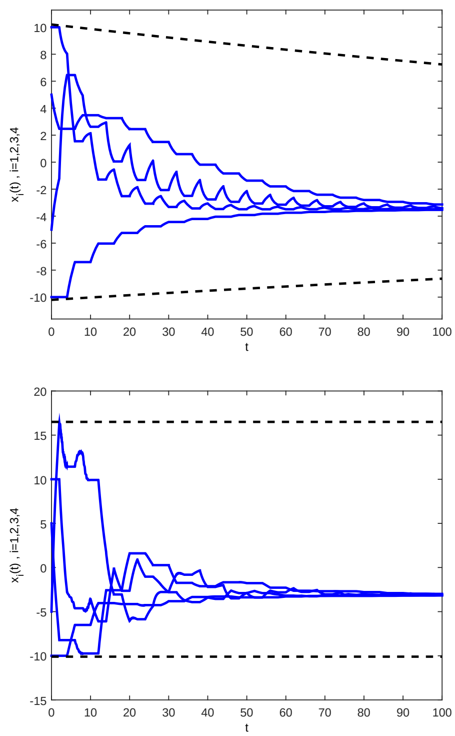


FIGURE 3. Example 7.1.2. (a): The convergence of 4 agents with a path graph topology. (b): Convergence under the effect of delays. In both figures, the dashed lines depict the theoretical rate estimate. It is remarked that in (b) the estimate is significantly weak.

convergence with the delay is

$$\theta = \frac{\ln(1 - 2\alpha^2(1 - e^{-\epsilon})^2 e^{-\tau})}{6\epsilon + 3\tau}.$$

As a numerical example, take $\epsilon = 2$, $B = 6$, $\alpha = 0.1$ and $a_{12}(t) = 0.2u(t)$, $a_{21}(t) = 0.3u(t)$, $a_{23}(t) = 0.21$, $a_{32}(t) = 0.2u(t)$, $a_{34}(t) = 0.25u(t)$ $a_{43}(t) = 0.1u(t)$ by an appropriate switching function $u(t)$. The rate of convergence is $\theta = 0.00251$, see Fig. 3(a). In the presence of the common delay with $\sup_{t\geq 0}\tau(t) = 10$ the rate is $\theta = 2.61 \cdot 10^{-8}$, see Fig. 3(b).

8.2. **A 2×2 network with unbounded delays.** Fix $t_0 > 0$ and consider the network of two agents to satisfy

$$\begin{cases} \dot{x}_1(t) = \frac{1}{\alpha t}\big(x_2(\beta t) - x_1(t)\big) \\ \dot{x}_2(t) = \frac{1}{\gamma t}\big(x_1(\varepsilon t) - x_2(t)\big), t \geq t_0 \\ \big(x_1(t), x_2(t)\big) = \big(\phi_1(t), \phi_2(t)\big), t \in [\beta t_0, t_0) \end{cases}$$

for some $\alpha, \gamma > 0$ and $\beta, \varepsilon \in (0, 1)$. This system lies beyond the theory developed in the preceding sections. In fact, it only takes few elementary, yet tedious, modifications to include systems with unbounded delays. These are, in fact, easily

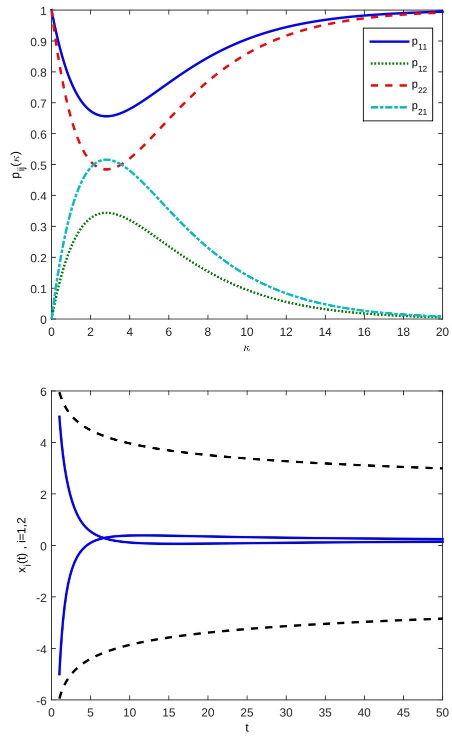


FIGURE 4. Example 7.2 (a): Graphs of the elements $p_{ij}(\kappa)$ for selected values of α, β, γ demonstrating the dependence of ρ in κ. (b) The convergence of the 2×2 static, time-varying network with unbounded delays.

illustrated for $N = 2$. Without loss of generality assume $\beta < \varepsilon$ and $\alpha > \gamma$. Now $\tau_1(t) = (1-\beta)t$ and $\tau_2(t) = (1-\varepsilon)(t)$ so that $\tau(t) = (1-\beta)t$. We work as follows: Firstly, we introduce the "rate" function $h(t) = t^\kappa$ for $t > t_0$ and $\kappa > 0$. It is easy to see that x_1, x_2 satisfy the system of integral equations:

$$\begin{cases} x_1(t) = t^{-\kappa} \int_{\lambda(t)}^{t} \left[\kappa s^{\kappa-1} - \frac{s^{\kappa-1}}{\alpha} + s^\kappa \delta(s - \lambda(t))\right] x_1(s)\, ds + t^{-\kappa} \int_{\lambda(t)}^{t} \frac{s^{\kappa-1}}{\alpha} x_2(\beta s)\, ds \\ x_2(t) = t^{-\kappa} \int_{\lambda(t)}^{t} \left[\kappa s^{\kappa-1} - \frac{s^{\kappa-1}}{\gamma} + s^\kappa \delta(s - \lambda(t))\right] x_2(s)\, ds + t^{-\kappa} \int_{\lambda(t)}^{t} \frac{s^{\kappa-1}}{\gamma} x_1(\varepsilon s)\, ds \end{cases}$$

It is easy to check that for small κ the matrix

$$P = \begin{bmatrix} 1 - \frac{\kappa}{\alpha}(1-\beta)^\kappa & \frac{\kappa}{\alpha}(1-\beta)^\kappa \\ \frac{\kappa}{\gamma}(1-\beta)^\kappa & 1 - \frac{\kappa}{\gamma}(1-\beta)^\kappa \end{bmatrix}$$

is stochastic and obviously scrambling. Then

$$\rho = \min\left\{1 - \frac{\kappa}{\alpha}(1-\beta)^\kappa, \frac{\kappa}{\gamma}(1-\beta)^\kappa\right\} + \min\left\{\frac{\kappa}{\alpha}(1-\beta)^\kappa, 1 - \frac{\kappa}{\gamma}(1-\beta)^\kappa\right\} > 0$$

Then similar analysis as in Theorem 5.3, Lemma 5.5 and Remark 5.6

$$S_{I_t}(\mathbf{x}) \leq (1 - \rho e^{\frac{\ln \beta}{\alpha}}) S_{I_{\lambda^{(2)}(t)}}(\mathbf{x}) \leq (1 - \rho \beta^{1/\alpha}) S_{I_{\lambda^{(2)}(t)}}(\mathbf{x})$$

Now as for any t there exists $l \in \mathbb{Z}_+$ such that $\lambda^{(2l)}(t) \leq \frac{t_0}{\beta^2}$ or equivalently $l \geq \frac{\ln(t_0/(t\beta^2))}{2\ln\beta}$. So

$$\max_i |x_i(t) - x_\infty| \leq Z S_{t_0}(\phi) \left(\frac{t_0}{t}\right)^\zeta$$

where $Z = e^{\kappa\beta^{1/\alpha}}$ and $\zeta = -\frac{\kappa\beta^{1/\alpha}}{2\ln\beta} > 0$. The rate is sub-exponential due to unbounded delays.

As a numerical example we take $t_0 = 1$, $\alpha = 3$, $\gamma = 2$, $\varepsilon = 0.5$, $\beta = 0.3$. Fig. 4(a) depicts the dependence of the elements of P as κ varies. The selection of κ determines the estimates of ρ. If we take $\kappa = 1.22$ we obtain $\rho = 0.655$ and we calculate $Z = 1.953$ and $\zeta = 0.1821$. The simulation of the solution x_1, x_2 is presented in Fig. 4(b).

9. Special Cases & Applications.

The aim of this section is twofold. In the case of leader-follower dynamics an easy modification and a stability in variation argument guarantee convergence. Secondly, we explain how the linear theory applies to three types of non-linear networks.

9.1. Leader-follower dynamics.

In a communication network, a leader $i \in V$ is defined to be an agent that only affects the rest of the group, yet it cannot be affected by it, i.e. $d_i \equiv 0$. Non-negative matrix theory assures that the corresponding graph $\mathbb{G}$ of such a network can be routed-out branching if there is at most one leader [36] (i.e. the root of the graph). The aforementioned framework allows both in the ordinary and the delayed case allows for leader dynamics if the state of the leader is constant. Here we will assume that there is an agent with individual behavior that nevertheless converges asymptotically to some constant. Take, without loss of generality, the leader to be agent number 1. Then the network dynamics can be written as

$$i \in V : \begin{cases} \dot{z}_1(t) = g\big(t, z_1(t)\big), & t \geq t_0 \\ \dot{z}_i(t) = \sum_j a_{ij}(t)\big(z_j(\lambda_{ij}(t)) - z_i(t)\big), & t \geq t_0, i \neq 1 \\ z_i(t) = \phi_i(t), & t \in I_{t_0} \end{cases} \tag{9}$$

The dynamics of the leader's state $z_1(t)$ are assumed to evolve free of any interaction with the rest of the group so that it satisfies

$$|z_1(t) - k| \leq \frac{1}{h(t)} \tag{10}$$

for some $h \in C^0\big([t_0, \infty), (0, \infty)\big)$ with the property that $h(t) \to \infty$ as $t \to \infty$. Then the long-run behavior of (9) is associated with this of (6) via a stability in variation argument. The idea is for all $i \in V$ such that $a_{i1} \neq 0$ to write

$$\begin{aligned} \dot{z}_i(t) &= \sum_{j\neq 1} a_{ij}(t)\big(z_j(\lambda_{ij}(t)) - z_i(t)\big) + a_{i1}(t)\big(z_1(\lambda_{i1}(t)) - z_i(t)\big) \\ &= \sum_{j\neq 1} a_{ij}(t)\big(z_j(\lambda_{ij}(t)) - z_i(t)\big) + a_{i1}(t)\big(k - z_i(t)\big) + a_{i1}(t)\big(z_1(\lambda_{i1}(t)) - k\big) \end{aligned}$$

so that we can introduce a leaderless consensus system with an external, state-independent, disturbance that converges to a constant value, k with some prescribed rate. Then the limit point of the leader-follower network is necessarily k as it is the only constant solution to satisfy the system of differential equations.

Theorem 9.1. *Let the solution* $\mathbf{z} = \mathbf{z}(t, t_0, \boldsymbol{\phi})$, $t \geq t_0$ *of* (9) *and the dynamics of the leader together with condition* (10) *to hold. Assume the uniformity and connectivity conditions of Theorem 7.1. Then if* θ *is the rate of convergence of a delayed consensus with leader and if there exists a function* $c \in C^0\big([t_0,\infty),(0,\infty)\big)$ *with the property that* $c(t) \to \infty$ *as* $t \to \infty$ *and*

$$\sup_{t\geq t_0} e^{-\theta(t-t_0)}c(t) < \infty \qquad \& \qquad \sup_{t\geq t_0} c(t)\int_{t_0}^{t}\frac{e^{-\theta(t-s)}}{h(s)}ds < \infty \tag{11}$$

then there exists a constant $K > 0$ *such that*

$$||\mathbf{z}(t) - \mathbb{1}k||_\infty \leq \frac{K}{c(t)}.$$

9.2. **Non-Linear Networks.** We will see now that §4 is readily applicable to two important non-linear networks extensively discussed in the literature [1, 2, 8, 17, 19, 33, 34]. We will show that an elementary direct linearization argument suffices to apply Theorem 7.1 and obtain solid convergence results. In fact, the solution of the nonlinear networks are indistinguishable from the solutions of certain systems of the type (6).

9.2.1. *Passive Coupling.* A network of $N < \infty$ agents exchanges information according to the following algorithm:

$$i \in V : \begin{cases} \dot{x}_i(t) = \sum_j g_{ij}\big(t, x_j\big(t - \tau_{ij}(t)\big) - x_i(t)\big), & t \geq t_0 \\ x_i(t) = \phi_i(t), & t \in I_{t_0} \end{cases} \tag{12}$$

For any $t \geq t_0$ there may or may not exist a connection between j and i. This defines a connectivity regime that can be described by a graph $\mathbb{G}_g(t) = (V, E(t))$ with $(i,j) \in E(t)$ if and only if $g_{ij}(t,\cdot) \neq 0$ with the convention that $g_{ii} \equiv 0$. The passivity conditions for g_{ij} is summarized in the next statement:

Assumption 9.2. *For an open, connected subset of* $\mathbb{R}$, U *that contains the origin and for any* $i,j \in V$, *the functions* $g_{ij}(t,x) : [t_0,\infty) \times U \to \mathbb{R}$ *are continuous in* x *and right-continuous in* t *and they satisfy the following properties:*

1. $g_{ij}(\cdot,x) : [t_0,\infty) \to [0,g)$ *uniformly in* x *for some* $g < \infty$,
2. $g_{ij}(t,0) = 0$, $t \geq t_0$,
3. $g_{ij}(t,\cdot) \neq 0 \Rightarrow \frac{g_{ij}(t,x)}{x} > 0$ *for* $x \neq 0$ *so that* $\lim_{x\to 0}\frac{g_{ij}(t,x)}{x} \in \mathbb{R}_+$ *independent of* t.

The form of g_{ij} incorporates two crucial features of the consensus algorithms: The first is that g_{ij} are compatible with the previously discussed connectivity regimes (switching connectivity) and the second is the passivity property which makes the solutions to behave in a qualitative similar way to the linear case. The next Corollary is a direct application of Theorem 7.1

Corollary 9.3. *Consider* (12) *with the solution* $\mathbf{x} = \mathbf{x}(t,t_0,\boldsymbol{\phi})$, $t \geq t_0$ *so that* $S_{I_{t_0}}(\boldsymbol{\phi}) \in U$ *and let Assumption 9.2 hold. If for any* $(i,j) \in E(t)$, $\frac{g_{ij}(t,x)}{x}$ *is* t-*uniformly bounded away from zero and* $\mathbb{G}_g(t)$ *satisfies the connectivity conditions of Theorem 7.1, its solution satisfies* $\mathbf{x}(t) \to \Delta_I$ *as* $t \to \infty$. *In particular there is a constant* $k \in \big[\min_{j\in V}\min_{s\in I_{t_0}}\phi_j(s), \max_{j\in V}\max_{s\in I_{t_0}}\phi_j(s)\big]$ *to which* $x_i(t)$, $i \in V$ *converge exponentially fast.*

A standard example of passive functions for $g_{ij}(x) = \sin(x)$. The function is passive in $(-\pi,\pi)$ and the corollary applies whenever $\phi_i(t) \in (-\pi/2,\pi/2)$, $t \in I_{t_0}$, $i \in V$.

9.2.2. *Monotone Networks.* Another type of non-linear networks occurs in the following initial value problem

$$i \in V : \begin{cases} \dot{x}_i(t) = \sum_j g_{ij}\big(t, x_j(t-\tau_{ij}(t))\big) - \sum_j g_{ij}\big(t, x_i(t)\big), & t \geq t_0 \\ x_i(t) = \phi_i(t), & t \in I_{t_0} \end{cases} \tag{13}$$

where $g_{ij}(\cdot,\cdot) \in C^1\big([t_0,\infty) \times U, \mathbb{R}\big)$ for an open connected $U \subset R$, are appropriate smooth functions with the monotone condition

$$(i,j) \in E(t) \Rightarrow \frac{\partial}{\partial \xi} g_{ij}(t,x) > 0 \text{ uniformly in } t. \tag{14}$$

Apart from a non-linear extension of the original algorithm [34] considered in the control community, models of the type (13) and condition (14) have been introduced to model compartmental systems with lags [2, 8]. The next statement establishes the close connection between the systems (6) and (13). The sufficient connectivity conditions are similar with (12) and (6) and are therefore omitted.

Corollary 9.4. *Consider (13) and its solution $\mathbf{x} = \mathbf{x}(t, t_0, \boldsymbol{\phi})$, $t \geq t_0$. Under condition (14), if $\phi_i(t) \in U$ for $t \in I_{t_0}$ and $i \in V$, then $\mathbf{x}(t) \to \Delta_I$, as $t \to \infty$. In particular, there is a constant $k \in \big[\min_{j\in V} \min_{s \in I_{t_0}} \phi_j(s), \max_{j \in V} \max_{s\in I_{t_0}} \phi_j(s)\big]$ to which $x_i(t)$, $i \in V$ converge exponentially fast.*

9.2.3. *Flocking Networks of Cucker-Smale type.* Consider a finite population of N birds, each of which is characterized by the pair of position and velocity (x_i, u_i). The second order consensus (flocking) scheme

$$i \in V : \begin{cases} \dot{x}_i = u_i \\ \dot{u}_i = \sum_j a_{ij}(\mathbf{x})\big(u_j - u_i\big) \end{cases} \tag{15}$$

with initial data $\mathbf{x}^0, \mathbf{u}^0$, was proposed by Cucker and Smale in [5] as a flocking model. Ever since it has attracted the interest of many researchers [9, 10, 30] who extended and improved convergence results. The underlying scenario is that the two birds communicate at a rate that depends on their relative distance. The more distant two birds are, the smaller their interraction. The coupling weight is essentially non bounded from below and the objective is to derive sufficient initial conditions so that the flock achieve asymptotic speed alignment without being dissolved. Mathematically we are looking for initial data so that the solution $(\mathbf{x}, \mathbf{u})$ satisfies $S\big(\mathbf{u}(t)\big) \to 0$ as $t \to \infty$ and $\sup_t S\big(\mathbf{x}(t)\big) < \infty$. In case of symmetric communication weights, i.e. $a_{ij}(\mathbf{x}) = a(|x_i - x_j|)$ spectral graph theory methods are implemented [5, 10], while in the asymmetric case only increased connectivity results exist [30]. In the latter work, under the increased connectivity condition it was showed that if $a_{ij}(\mathbf{x}) \geq f\big(S(\mathbf{x})\big)$ then it suffices for the initial conditions to satisfy

$$S(\mathbf{u}^0) < \int_{S(\mathbf{x}^0)}^{\infty} f(s)\, ds \tag{16}$$

Our theory applies directly both in the ordinary and the delayed case and provides generalized convergence sufficient conditions. In the event of the switching connectivity we have the following result

Theorem 9.5. [39] *Consider the initial value problem (15) and the Assumptions 4.1, 4.2 with respect to the connectivity regime $\{a_{ij}\}$ to hold. If there exists a*

$B > 0$ to satisfy the conditions of Theorem 6.1 and $a_{ij}\big(\mathbf{x}(t)\big) \geq f\big(S(\mathbf{x}(t))\big)$. Then a sufficient condition for $(\mathbf{x}, \mathbf{u})$ to exhibit asymptotic flocking is:

$$S(\mathbf{u}^0) < \frac{(1-e^{-m\epsilon})^\sigma}{m^\sigma \sigma B} \int_{P^{\sigma,B}_{\mathbf{x}^0,\mathbf{u}^0}}^{\infty} f^\sigma(s)\, ds$$

where $P^{\sigma,B}_{\mathbf{x}^0,\mathbf{v}^0} = \max\{S(\mathbf{x}^0), |S(\mathbf{x}^0) - S(\mathbf{u}^0)\sigma B\}$ and σ, B, ϵ to have the meaning of Theorem 6.1.

It is remarked that in the case of increased static connectivity we have $\sigma = 1$, $\epsilon = B$ and we can take $B > 0$ arbitrarily small to see that (16) is recovered as a special case. Additionally, if $\int^\infty f^\sigma(s)\, ds$ diverges then we have unconditional flocking for (15). For the case of propagation delays a similar result is presented in [38].

10. Discussion & Concluding Remarks.

In this work we revisited the consensus problem from a unification perspective. We proved convergence result under the mildest connectivity assumptions and provided explicit estimates on the rate of convergence for ordinary and functional versions of the distributed algorithm. The novelty of this work is the development of a theoretical framework that was primarily used for discrete time dynamics. Our analysis extends to unbounded delays as well as various type of non-linearities.

However a number of points that occur out of the analysis are due for clarification.

A first issue with the contraction coefficient ρ is the seemingly mystic role of $B > 0$. Inverting the system of differential equations to a system of integral equations, we asked for a positive number $B > 0$, that will allow the solution to evolve over $[t - B, t]$ so that the analysis can apply. Interestingly enough, B has its roots in the classification of states in finite state continuous time Markov Chains. Then as $B > 0$ is the necessary length interval of time for the contraction coefficient to act, so is the necessary time needed for the state classification into communication classes of the Markov Chain: Indeed, a pair of states i, j of the chain belongs to the same communication class if the probability starting from i to arrive in j is strictly positive for positive times, i.e. $\mathbb{P}(X_{t+B} = j | X_t = i) > 0$ for $B > 0$ (see p. 260 of [7]). In other words, $B > 0$ is the necessary time the system needs to identify the communication classes in the network.

A second issue arises in the delayed case where we essentially ask for λ_{ij} to be invertible so that the aforementioned framework applies with mild modifications. We conjecture that this technical assumption may be dropped with an appropriate extension of the contraction coefficient to appropriate functional spaces. The results are delay-independent but the presence of delays significantly debilitates the rate of convergence. In fact the stronger the coupling weights a_{ij} the weaker the rate of convergence becomes. In the example section we showed how the theory can be extended to systems with unbounded delays. In this case, of course, the rate of convergence is sub-exponential.

The delay-independent convergence is due to propagation delays and it generalized scalar versions of this model that were studied in [18] in a total different vein. If, in addition, to propagation, one considers processing delays, i.e. systems of the form

$$\dot{x}_i(t) = \sum_j a_{ij}(t)\big(x_j(t - \tau_i^j(t)) - x_i(t - \sigma_i^j(t))\big)$$

they may destabilize the system [27, 28, 32] . The main feature in the presence of processing delays is that the solution loses the critical property of Remark 5.6 so that the methodology developed here does not apply.

All in all, the utilization of the contraction coefficient for the study of continuous time consensus dynamic yields simple and concise convergence results without strong assumptions on the connectivity regime for both ordinary and delayed versions of the problem. The rate estimates depend on the coupling weights, the switching signal, the parameter $B > 0$ and the magnitude of the delays.

On the negative side, the rate estimates are very weak, as the simulations clearly suggest. In case of symmetric networks ($a_{ij} = a_{ji}$) the use of a metric derived from spectral graph theory is preferred [30]. In this case another approximation of the second eigenvalue that controls the convergence rate, the Fiedler number yields better estimates than the contraction coefficient [4]. On the positive side, we have a rigorous framework that provides a deep understanding of the long-term behavior of this category of distributed cooperative systems ranging from ordinary and linear to functional and non-linear. Future research problems involve an improvement of the contraction coefficient at least to match the performance of the symmetric estimator with or without the effect of delays.

10.1. **Conclusions.** The present paper develops a framework for a large family of distributed consensus networks and establishes generalized necessary and sufficient conditions for asymptotic stability. The emphasis is put on the rate of convergence. This work aspires to serve as a first step towards a unified theory of consensus.

11. **Appendix.** In this section we have put all the proofs of this paper.
Proof of Proposition 5.2. The matrix

$$P_B(t) := \int_{t-B}^{t} \left(e^{-mB}\delta(s_1 - (t-B))\mathbb{I} + e^{-m(t-s_1)}W(s_1) \right) ds_1$$

is stochastic. Indeed, the i^{th} row of $P_B(t)$ consists of the positive diagonal element

$$e^{-mB} + \int_{t-B}^{t} e^{-m(t-s_1)}\big(m - d_i(s_1)\big)\, ds_1 = 1 - \int_{t-B}^{t} e^{-m(t-s_1)} d_i(s_1)\, ds_1$$

and the non-negative off-diagonal elements

$$\int_{t-B}^{t} e^{-m(t-s_1)} a_{ij}(s_1)\, ds_1.$$

since $d_i(s_1) = \sum_j a_{ij}(s_1)$, $P_B(t)$ is stochastic. We proceed with induction: For $l = 2$,

$$P_B^{(2)}(t) =$$
$$= \int_{t-B}^{t} \int_{s_1-B}^{s_1} \left(e^{-mB}\delta(s_1 - (t-B))\mathbb{I} + e^{-m(t-s_1)}W(s_1) \right) \cdot$$
$$\cdot \left(e^{-mB}\delta(s_2 - (s_1 - B))\mathbb{I} + e^{-m(s_1-s_2)}W(s_2) \right) ds_2 ds_1$$
$$= \int_{t-B}^{t} \int_{s_1-B}^{s_1} e^{-2mR}\delta(s_1 - (t-B))\delta(s_2 - (s_1 - B))\, ds_2 ds_1 \mathbb{I} +$$
$$+ \int_{t-B}^{t} \int_{s_1-B}^{s_1} e^{-mR}\delta(s_1 - (t-B))e^{-m(s_1-s_2)}W(s_2)\, ds_2 ds_1 +$$
$$+ \int_{t-B}^{t} \int_{s_1-B}^{s_1} e^{-m(t-s_1)}W(s_1)e^{-mR}\delta(s_2 - (s_1 - B))\, ds_2 ds_1 +$$
$$+ \int_{t-B}^{t} \int_{s_1-B}^{s_1} e^{-m(t-s_1)}W(s_1)e^{-m(s_1-s_2)}W(s_2)\, ds_2 ds_1$$

and straightforward calculations yield

$$P_B^{(2)}(t) = e^{-2mB}\mathbb{I} + \int_{t-2B}^{t-B} e^{-m(t-s_2)}W(s_2)\, ds_2 + e^{-mB}\int_{t-B}^{t} e^{-m(t-s_1)}W(s_1)\, ds_1$$
$$+ \int_{t-B}^{t} \int_{s_1-B}^{s_1} e^{-m(t-s_2)}W(s_1)W(s_2)\, ds_2 ds_1$$

Now, every element of $P_B^{(2)}(t)$ is non-negative as a sum of non-negative matrices. It is only left to verify that $\sum_j [P_B^{(2)}(t)]_{ij} = 1$ for any i. Indeed, the first matrix contributes e^{-2mB}, the second and the third $e^{-mB} - e^{-2mB}$ and the fourth $(1 - e^{-mB})^2$, so eventually

$$e^{-2mB} + 2(e^{-mB} - e^{-2mB}) + (1 - 2e^{-mB} + e^{-2mB}) = 1$$

Let $P_B^{(l)}(t)$ be stochastic. Then the elements of $P_B^{(l+1)}(t)$ are non-negative by the same reasoning as above and finally, since

$$P_B^{(l+1)}(t)$$
$$= \int_{t-B}^{t} C(t, s_1) \cdots \int_{s_l-B}^{s_l} C(s_l, s_{l-1})\, ds_{l+1} \dots d_{s_1}$$
$$= e^{-mB}P_B^{(l)}(t) + (1 - e^{-mB})P_B^{(l)}(t) -$$
$$- \int_{t-B}^{t} \int_{s_1-B}^{s_1} \cdots \int_{s_l-B}^{s_l} C(t, s_1)C(s_1, s_2) \cdots \big(D(s_{l+1}) - A(s_{l+1})\big)\, ds_{l+1} \dots d_{s_1}$$
$$= P_B^{(l)}(t) - \int_{t-B}^{t} \int_{s_1-B}^{s_1} \cdots \int_{s_l-B}^{s_l} C(t, s_1)C(s_1, s_2) \cdots \big(D(s_{l+1}) - A(s_{l+1})\big)\, ds_{l+1} \dots d_{s_1}$$

the sum of the i^{th} row of $P_B^{(l+1)}(t)$ equals 1 because the corresponding sum in the final integrand is zero (as it is a left multiplication of a matrix with a Laplacian matrix). □

Proof of Theorem 5.3. The proof relies on the second mean value theorem and a technical lemma, both of which are cited below for quick reference:

Lemma 11.1. *If $G \in C^0[J, \mathbb{R}]$ and ϕ is integrable that does not change sign on J then there exists $x \in J$ such that*

$$G(x)\int_J \phi(t)\,dt = \int_J G(t)\phi(t)\,dt.$$

We recall that two vectors $\mathbf{x}, \mathbf{y}$ are sign compatible if $x_i y_i \geq 0$ for all i.

Lemma 11.2 (Lemma 1.1 of [12]). *Suppose $\boldsymbol{\delta} \in \mathbb{R}^N$ such that $\boldsymbol{\delta}^T \mathbb{1} = 0$ and $\boldsymbol{\delta} \neq 0$. Then there is an index $\mathcal{I} = \mathcal{I}(\delta)$ of ordered pairs (i, j) with $i, j \in V$ such that*

$$\boldsymbol{\delta}^T = \sum_{(i,j)\in\mathcal{I}} \frac{T_{ij}}{2}(\mathbf{e}_i - \mathbf{e}_j)$$

where $T_{ij} > 0$, $\mathbf{e}_i$ is the row vector $(0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 at the i^{th} position, $\mathbf{e}_i - \mathbf{e}_j$ is sign compatible to δ for all i, j. Thus $||\boldsymbol{\delta}||_1 = \sum_{(i,i)\in\mathcal{I}} T_{ij}$.

Pick $h, h^{'} \in V$. Then for $\mathbf{p}_h, \mathbf{p}_{h'}$ the h^{th} and $h^{'th}$ rows of P respectively, we have

$$\int_{s\in I} \big(\mathbf{p}_h(s) - \mathbf{p}_{h'}(s)\big)\mathbf{z}(s)\,ds$$

Now, since $N < \infty$ there is a partition $\{I_l\}_{l=1}^m$ of I which depends on h, h' such that for any I_l, $p_{hk}(s) - p_{h'k}(s)$ does not change sign in for $s \in I_l$, $k \in V$ and it is not identically zero. Then for fixed I_l we apply Lemma 11.1 to obtain

$$\sum_k \int_{s\in I_l} \big(p_{hk}(s) - p_{h'k}(s)\big) z_k(s)\,ds = \sum_k \int_{s\in I_l} \big(p_{hk}(s) - p_{h'k}(s)\big)\,ds z_k(s_k^*) = \boldsymbol{\delta}_l^T \mathbf{z}_l^*$$

for some $s_k^* = s(I_l, h, h')$, $\boldsymbol{\delta}_l^T = \int_{I_l} \big(\mathbf{p}_h(s) - \mathbf{p}_h'(s)\big)\,ds \neq 0$ and $\mathbf{z}_l^* = (z_1(s_1^*), \ldots z_N(s_N^*))^T$. By Assumption $\int_{I_l} P(s)\,ds$ is m-stochastic and therefore $\boldsymbol{\delta}_l^T \mathbb{1} = 0$. Hence, Lemma 11.2 is applied and together with the triangle inequality

$$|\boldsymbol{\delta}_l^T \mathbf{z}_l^*| \leq \frac{1}{2}||\boldsymbol{\delta}_l||_1 S(\mathbf{z}_l^*)$$

(see also [12]). Then if we let $S(\mathbf{z}*) = \max_l S(\mathbf{z}_l^*)$, we obtain the bound

$$S(\mathbf{w}) = \max_{h,h'} \left| \int_{s\in I} \big(\mathbf{p}_h(s) - \mathbf{p}_{h'}(s)\big)\mathbf{z}(s)\,ds \right|$$
$$= \sum_l |\boldsymbol{\delta}_l^T \mathbf{z}_l^*| \leq \max_{h,h'} \frac{1}{2} \int_I ||\mathbf{p}_h(s) - \mathbf{p}_{h'}(s)||_1\,ds S(\mathbf{z}^*).$$

Finally, from the identity $|x-y| = x+y-2\min\{x,y\}$ for any $x, y \in \mathbb{R}$ and the fact that $\forall h, h' \in V$, $\sum_k \int_{s\in I} p_{hk}(s)\,ds = \sum_k \int_{s\in I} p_{h'k}(s)\,ds = m$ we get

$$\frac{1}{2}\max_{h,h'} \sum_k \int_{s\in I} |p_{hk}(s) - p_{h'k}(s)| ds = m - \min_{h,h'} \sum_k \min\left\{ \int_{s\in I} p_{hk}(s)\,ds, \int_{s\in I} p_{h'k}(s)\,ds \right\}.$$

□

Proof of Lemma 5.5. Let $t^* \geq t_0$ and $i \in V$ be the first time and agent that the solution $x_i(t)$ escapes $\big[\min_{j\in V, s\in I_{t_0}} \phi_j(s), \max_{j\in V, s\in I_{t_0}} \phi_j(s)\big]$ say, to the right. Then it must hold both that $x_i(t^*) = \min_{j\in V, s\in I_{t_0}} \phi_j(s)$ and $\dot{x}_i(t^*) > 0$, which is a contradiction in view of the dynamics in (6). The same argument can be made for escaping to the left. □

Proof of Lemma 5.8. From Lemma 5.5 we have that $\omega(\boldsymbol{\phi})$ is non-empty, compact and connected and any element of which must lie in Δ. Since $\lambda(t_0) < \infty$ any point $\boldsymbol{\phi}^\omega \in \omega(\boldsymbol{\phi})$ is actually a vector valued function with the property that $\phi_i^\omega(s) = \phi_j^\omega(s)$ $\forall i, j \in V$. It is obvious however that $\mathbf{x}(t, t_0, \boldsymbol{\phi}^\omega) \equiv \boldsymbol{\phi}^\omega(t_0)$ and at the same time a member of $\omega(\boldsymbol{\phi})$. By the uniqueness of solutions it follows that $\boldsymbol{\phi}^\omega$ must be a constant vector valued function in $\mathbb{R}^N \cap \Delta$ and the result follows. ◻

Proof of Theorem 6.1. The solution $\mathbf{x}$ of (7) satisfies

$$\begin{aligned}
&\dot{\mathbf{x}}(t) = -m\mathbf{x}(t) + \big(m\mathbb{I} - D(t) + A(t)\big)\mathbf{x}(t)\\
&\Rightarrow \frac{d}{dt}\big(e^{mt}\mathbf{x}(t)\big) = e^{mt}\big(m\mathbb{I} - D(t) + A(t)\big)\mathbf{x}(t)\\
&\Rightarrow e^{mt}\mathbf{x}(t) - e^{m(t-B)}\mathbf{x}(t-B) = \int_{t-B}^{t} e^{ms}\big(m\mathbb{I} - D(s) + A(s)\big)\mathbf{x}(s)\, ds
\end{aligned}$$

Set for simplicity $C(t,s) := \big(e^{-mB}\delta(s-(t-B))\mathbb{I} + e^{-m(t-s)}(m\mathbb{I} - D(s) + A(s))\big)$. By the imposed condition in the statement of the theorem it holds that $P_B(t) = \int_{t-B}^{t} C(t, s_0)\, ds_0 \in \mathcal{T}$, for any $t \geq t_0 + B$. Then

$$\begin{aligned}
\mathbf{x}(t) &= \int_{t-B}^{t} C(t, s_1)\mathbf{x}(s_1)\, ds_1\\
&= \int_{t-B}^{t}\int_{s_1-B}^{s_1}\cdots\int_{s_{\sigma-1}-B}^{s_{\sigma-1}} C(t,s_1)C(s_1,s_2)\cdots C(s_{\sigma-1},s_\sigma)\mathbf{x}(s_\sigma)\, ds_\sigma\cdots ds_1
\end{aligned}$$

for $t \geq t_0 + \sigma B$. Choosing σ large enough, $P_B^{(\sigma)}(t)$ should be scrambling. We will show this by estimating the value of σ. Since for any t and $s \in [t-B, t]$, $\gamma_{P_B(t)}, \gamma_{P_B(s)} \geq 1$ are the scrambling indexes of $P_B(t)$ and $P_B(s)$ respectively, and $\gamma_{P_B(s)} \geq 1$ for $s \in [t-B, t]$ so that $\gamma_{P_{B(s^*(t))}} = \max_{s\in[t-B,t]} \gamma_{P_B(s)}$, then $P_B^{(2)}(t) = \int_{t-B}^{t} C(t,s)P_B(s)\, ds \in \mathcal{T}$ as well and for its scrambling index $\gamma_{P_B^{(2)}(t)}$ it holds that

$$\gamma_{P_B^{(2)}}(t) \leq \begin{cases} \max\{\gamma_{P_B(t)}, \gamma_{P_{B(s^*(t))}}\} - 1, & \mathbb{G}_{P_B(t)} \subset \mathbb{G}_{P_B(s^*)} \text{ or } \mathbb{G}_{P_B(s^*)} \subset \mathbb{G}_{P_B(t)}\\ \max\{\gamma_{P_B(t)}, \gamma_{P_{B(s^*(t))}}\}, & \text{o.w.}\end{cases}$$

Now, in the second case, there is no strict decrease of the scrambling index over $[t-2B, t]$. In this case, however, it holds that $E^C_{P_B(t)} \cap E_{P_B(s^*)} \neq \emptyset$[4], i.e. there exists $(i,j) \in V \times V$ such that $[\mathbb{G}_{P_B(t)}]_{ij} > 0$ and $[\mathbb{G}_{P_B(s^*)}]_{ij} = 0$ or vice versa. This element, however, will be a member of $E_{P_B^{(2)}(t)}$ exactly because P_B have strictly positive diagonal elements. From the discussion on the partitioning of $\mathcal{T}$ with respect to the scrambling indexes,

$$\gamma_{P_B^{l^*}(t)} \leq \max\{\gamma_{P_B(t)}, \gamma_{P_{B(u^*(t))}}\} - 1,$$

$\gamma_{P_B(u^*(t))} := \max_{s\in[t-(l^*-1)B,t]} \gamma_{P_B(s)}$. Consequently for $\sigma = l^*([N/2]+1)$ the matrix $P_B^{(\sigma)}(t)$ is scrambling.

A direct calculation reveals that for any $I' \subset [t-B, t]$, $\int_{I'} C(t,s)\, ds$ (and consequently $\int_{s_1\in I'}\int_{s_1-B}^{s_1}\cdots\int_{s_{\sigma-1}-B}^{s_{\sigma-1}} C(t,s_1)C(s_1,s_2)\cdots C(s_{\sigma-1},s_\sigma)\mathbf{x}(s_\sigma)\, ds_\sigma\cdots ds_1$) belong to $\mathcal{M}$. So that Theorem 5.3 (by Remark 5.4) applies to yield together with Lemma 5.5 (which obviously holds for $\tau(t) \equiv 0$) and with Proposition 5.2 the estimate

$$S\big(\mathbf{x}(t)\big) \leq \rho\big(P_B^{(\sigma)}(t)\big)S\big(\mathbf{x}(t-\sigma B)\big)$$

[4] E^C denotes the complement of E.

for $\sigma = l^*([N/2]+1)$ and $\rho\big(P_B^{(\sigma)}(t)\big) < 1$ on the assumption of static connectivity. The next step is to estimate $\rho\big(P_B^{(\sigma)}(t)\big)$. Since $P_B^{(\sigma)}(t)$ is scrambling, there exists $j^* \in V$ such that $[P_B^{(\sigma)}(t)]_{j^*i} > 0$. By direct calculations we have:

$$
\begin{aligned}
&[P_B^{(\sigma)}(t)]_{j^*j^*} \geq \\
&\int_{t-B}^{t}\int_{s_1-B}^{s_1}\cdots\int_{s_{\sigma-1}-B}^{s_{\sigma-1}}\prod_{k=1}^{\sigma}\Big(e^{-mB}\delta\big(s_k-(s_{k-1}-B)\big)+ \\
&\qquad\qquad + e^{-m(s_{k-1}-s_k)}\big(m-d_i(s_k)\big)\Big)\,ds_\sigma\cdots ds_1 \\
&> e^{-\sigma mB}
\end{aligned}
$$

and for $i \neq j^*$

$$
\begin{aligned}
&[P_B^{(\sigma)}(t)]_{j^*i} \geq \\
&\geq \int_{t-B}^{t}\int_{s_1-B}^{s_1}\cdots\int_{s_{\sigma-1}-B}^{s_{\sigma-1}}\sum_{l_0,\dots,l_{\sigma-1}} e^{-m(t-s_\sigma)}a_{il_0}(s_1)a_{l_0l_1}(s_2)\dots a_{l_{\sigma-1}j^*}(s_\sigma)\,ds_\sigma\cdots ds_1 \\
&> \int_{t-B}^{t}\int_{s_1-B}^{s_1}\cdots\int_{s_{\sigma-1}-B}^{s_{\sigma-1}} e^{-m(t-s_\sigma)}\,ds_\sigma\cdots ds_1 f^\sigma(t) = \frac{(1-e^{-m\epsilon})^\sigma}{m^\sigma}f^\sigma(t)
\end{aligned}
$$

where f and $\epsilon > 0$ have the meaning of Assumption 4.2. For $t' \geq M$ large enough so that $f(t) \leq \frac{me^{-mB}}{1-e^{-m\epsilon}}$ whenever $t \geq t'$, from the definition of ρ we obtain the estimate:

$$
\rho\big(P_B^{(\sigma)}(t)\big) \leq 1 - \frac{(1-e^{-m\epsilon})^\sigma}{m^\sigma}f^\sigma(t) \tag{17}
$$

Then, for the aforementioned sequence $\{t_n\}$, for any $t \geq t'$, there exists $\bar{n} \in \mathbb{N}$ such that $t \in [t_{\bar{n}}, t_{\bar{n}+1}]$ so that

$$
\begin{aligned}
S\big(\mathbf{x}(t)\big) \leq S\big(\mathbf{x}(t_{\bar{n}})\big) &\leq \Big(1-\frac{(1-e^{-m\epsilon})^\sigma}{m^\sigma}f^\sigma(t_{\bar{n}})\Big)S\big(\mathbf{x}(t_{\bar{n}}-\sigma B)\big) \\
&\leq \Big(1-\frac{(1-e^{-m\epsilon})^\sigma}{m^\sigma}f^\sigma(t_{\bar{n}})\Big)S\big(\mathbf{x}(t_{\bar{n}-1})\big)
\end{aligned}
$$

For any $\varepsilon > 0$, pick n_1 and n_2 large enough so that $t_{n_1} \geq t'$ and $\sum_{j=n_1}^{n_2} f(t_j) \geq \big[\frac{(1-e^{-m\epsilon})^\sigma}{m^\sigma}\big]^{-1}\log(\frac{\varepsilon}{S(\mathbf{x}^0)})$. Then for $t \geq t_{n_1}$

$$
\begin{aligned}
S\big(\mathbf{x}(t)\big) &\leq \prod_{k=i_1}^{i_2}\Big(1-\frac{(1-e^{-m\epsilon})^\sigma}{m^\sigma}f^\sigma(t_k)\Big)S\big(\mathbf{x}^0\big) \\
&\leq e^{-\frac{(1-e^{-m\epsilon})^\sigma}{m^\sigma}\sum_{k=i_1}^{i_2} f^\sigma(t_k)}S\big(\mathbf{x}^0\big) \leq \varepsilon.
\end{aligned}
$$

□

Proof of Corollary 6.3. This is a direct application of Theorem 6.1. If $f \geq \underline{f} > 0$, then from the proof or Theorem 6.1 it holds that

$$
S\big(\mathbf{x}(t)\big) < (1-\rho)S\big(\mathbf{x}(t-\sigma B)\big)
$$

where $\rho := \min\{e^{-\sigma mB}, \frac{((1-e^{-m\epsilon})\underline{f})^\sigma}{m^\sigma}\}$. Then there exists $l \in \mathbb{Z}_+$ such that $t-(l+1)\sigma B \leq t_0 \leq t - l\sigma B$ and hence

$$
S\big(\mathbf{x}(t)\big) \leq (1-\rho)^{l+1}S(\mathbf{x}^0) \leq \frac{(1-\rho)^{\frac{t-t_0}{\sigma B}}}{1-\rho}S(\mathbf{x}^0) = \frac{S(\mathbf{x}^0)}{1-\rho}e^{-\theta(t-t_0)}
$$

where $\theta = -\frac{\ln(1-\rho)}{\sigma B}$ and the proof is concluded. □

Proof of Theorem 6.4. Let the initial conditions be set such $x_l(0) < x_n(0)$ for $l \in V_1$ and $n \in V_2$. Consider then the subset V_{11} of V_1 and accordingly the subset V_{22} of V_2 which by assumption they must have connections between them. Let $i \in V_{11}$ such that $x_i(t) \leq x_{i^*}(t)$ for any $i^* \in V_{11}$ and $j \in V_{22}$ such that $x_j(t) \geq x_{j^*}(t)$ for any $j^* \in V_{22}$. Then by the imposed initial conditions we have:

$$\dot{x}_i(t) \leq d_{ij}\big(x_j(t) - x_i(t)\big) + z_i(t)$$
$$\dot{x}_j(t) \geq d_{ji}\big(x_i(t) - x_j(t)\big) + z_j(t)$$

where $z_i(t) = \sum_{l \in V_1} a_{il}\big(x_l(t) - x_i(t)\big)$, $z_j(t) = \sum_{l \in V_2} a_{jl}\big(x_l(t) - x_j(t)\big)$ are functions that signify the interconnections among agents on the separated subsets. Taking the difference

$$\frac{d}{dt}\big(x_j(t) - x_i(t)\big) \geq -\big(d_{ij}(t) + d_{ji}(t)\big)\big(x_j(t) - x_i(t)\big) + z_j(t) - z_i(t)$$

and if either $z_i(t)$ or $z_j(t)$ do not vanish then $S(\mathbf{x}(t))$ will not converge to zero and there is nothing to prove. On the other hand, we have by assumption that $|z_i(t) - z_j(t)| \leq 2(N-1)C\Gamma e^{-\gamma t}$ for $(N-1)C$ to play the role of the uniform upper bound of $a_{ij}(t)$ according to Assumption 4.1. We set for simplicity $Q(s) = \big(d_{ij}(s) + d_{ji}(s)\big)$. Now, $\int_{t_0}^{\infty} Q(s)\,ds < \infty$ and this means that there is a sequence $\{t_n\}_{n \geq 1}$ and a constant $M_1 > 0$ such that

$$\int_0^{t_n} Q(s)\,ds \geq M_1.$$

Since

$$x_j(t) - x_i(t) \geq e^{-\int_0^t Q(s)\,ds}(x_j^0 - x_i^0) + \int_0^t e^{-\int_w^t Q(s)\,ds}\big(f_j(w) - f_i(w)\big)\,dw,$$

we have that

$$|x_j(t_n) - x_i(t_n)| \geq \left| e^{-M_1}|x_{ji}^0| - \int_0^{t_n} e^{-\int_w^{t_n} Q(s)\,ds} 2(N-1)Ce^{-\gamma w}\,dw S(\mathbf{x}^0) \right|.$$

Choosing $\left|e^{-M_1}|x_{ji}^0| - \frac{2(N-1)C\Gamma}{\gamma}\right| > \epsilon$ we obtain $|x_{ij}(t)| > \epsilon$ for infinitely many t. □

Proof of Theorem 7.1. Fit $t \geq \sigma(B+\tau)+\tau+t_0$ and $B > 0$. The first step is to write the solution $\mathbf{x}$ as an integral equation in the form of $\int_{t-B}^t C(t,s)\mathbf{x}(s)\,ds$ in order to use Theorem 5.3. In view of Assumption 4.3 we see that the functions $t - \tau_{ij}(t)$ are invertible. Denote by $\kappa_{ij}(t)$ their inverse. Define $M(t,s) = [m_{ij}(t,s)]$ with elements

$$m_{ij}(t,s) = \begin{cases} e^{-m(t-\kappa_{ij}(s))} \frac{a_{ij}(\kappa_{ij}(s))}{1-\dot{\tau}_{ij}(\kappa_{ij}(s))} \mathbf{1}_{s \in [\lambda_{ij}(t-B), \lambda_{ij}(t)]}, & i \neq j \\ e^{-mB}\delta\big(s-(t-B)\big) + e^{-m(t-s)}\big(m - d_i(s)\big)\mathbf{1}_{s \in [(t-B),t]}, & i = j \end{cases}$$

Following the steps of the proof of Theorem 6.1 the solution $\mathbf{x}$ in vector form reads

$$\mathbf{x}(t) = \int_{t-(B+\tau)}^t M(t,s)\mathbf{x}(s)\,ds$$

so that again $P_{B+\tau}(t) = \int_{t-(B+\tau)}^t M(t,s)\,ds$ is stochastic. Under the imposed connectivity conditions, Proposition 5.2 implies that the stochastic matrix $P_{B+\tau}^{(\sigma)}(t)$ is scrambling. Then by Theorem 5.3 and Remark 5.6 we have the estimate:

$$S(\mathbf{x}(t)) \leq (1-\mu)S_{I_{t-\sigma(B+\tau)-\tau}}(\mathbf{x}) \tag{18}$$

for $\mu := \inf_{t \geq t_0} \min_{i,j} \sum_l \min\{p_{il}^t, p_{jl}^t\} \in (0,1)$ with p_{ij}^t the elements of $P_B^{(\sigma)}(t)$.

Consider the moments $t_1, t_2 \in I_t$ and the agents $i^*, j^* \in V$ such that $S_{I_t}(\mathbf{x}) = x_{i^*}(t_1) - x_{j^*}(t_2)$. Assume, without loss of generality, that $t_1 \geq t_2$. Then from (8), Lemma 5.5 and Remark 5.6 we have

$$
\begin{aligned}
&S_{I_t}(\mathbf{x}) = x_{i^*}(t_1) - x_{j^*}(t_2)\\
&= e^{-\int_{t_2}^{t_1} d_{i^*}(s)\,ds}\big(x_{i^*}(t_2) - x_{j^*}(t_2)\big) +\\
&\qquad\qquad + \int_{t_2}^{t_1} e^{-\int_s^{t_1} d_{i^*}(w)dw} \sum_j a_{i^*j}(s)\big(x_j(\lambda_{i^*j}(s)) - x_{j^*}(t_2)\big)\,ds\\
&\leq e^{-\int_{t_2}^{t_1} d_{i^*}(s)\,ds} S(\mathbf{x}(t_2)) + \big(1 - e^{-\int_{t_2}^{t_1} d_{i^*}(s)\,ds}\big) S_{I_{t-2\tau}}(\mathbf{x})\\
&\leq e^{-\int_{t_2}^{t_1} d_{i^*}(s)\,ds}(1-\mu) S_{I_{t-\sigma(B+\tau)-\tau}}(\mathbf{x}) + \big(1 - e^{-\int_{t_2}^{t_1} d_{i^*}(s)\,ds}\big) S_{I_{t-2\tau}}(\mathbf{x})\\
&\leq (1 - \mu e^{-\int_{t_2}^{t_1} d_{i^*}(s)\,ds}) S_{I_{t-\sigma(B+\tau)-\tau}}(\mathbf{x})\\
&\leq (1 - \mu e^{-\bar{N}a\tau}) S_{I_{t-\sigma(B+\tau)-\tau}}(\mathbf{x})
\end{aligned}
$$

Similar estimates occur for $t_1 < t_2$. Consequently there exists integer $l \geq 1$ such that $t_0 + l(\sigma(B+\tau)+\tau) \leq t \leq t_0 + (l+1)(\sigma(B+\tau)+\tau)$ and so a recursive argument implies that

$$
S_{I_t}(\mathbf{x}) \leq \frac{S_{I_{t_0}}(\boldsymbol{\phi})}{1 - \mu e^{-\bar{N}a\tau}} e^{\frac{\ln(1-\mu e^{-\bar{N}a\tau})}{(\sigma(B+\tau)+\tau)}(t-t_0)}
$$

and the proof is concluded in view of Lemma 5.8. □

Proof of Theorem 9.1. The proof relies on an elementary variation argument. Eq. (6) can be written in compact vector form

$$
\dot{\mathbf{x}} = -\mathcal{L}(t, \mathbf{x}_t)
$$

where $\mathbf{x}_t = \mathbf{x}_t(t_0, \boldsymbol{\phi})$ stands for the function segment $\mathbf{x}(t+s, t_0, \boldsymbol{\phi})$, $s \in [-\tau, 0]$, $t \geq t_0$ and $\mathcal{L}(t, \boldsymbol{\psi})$ is a functional linear in $\boldsymbol{\psi}$. At this stage we can define a linear and continuous operator T such that $T(t, t_0)\boldsymbol{\phi} = \mathbf{x}_t(t_0, \boldsymbol{\phi})$ so that $(T(t, t_0)\boldsymbol{\phi})(s) = \mathbf{x}(t+s, t_0, \boldsymbol{\phi})$, $s \in [-\tau, 0]$. See also [11, 21]. Theorem 7.1 and the existence of leader implies that $T(t, t_0)$ is also endowed with the property:

$$
||(T(t, t_0)\boldsymbol{\phi})(t_0) - \mathbb{1}(1, 0, \ldots, 0)\boldsymbol{\phi}(t_0)||_\infty \leq \Theta e^{-\theta(t-t_0)} \tag{19}
$$

for some constant Θ that depends on the initial data and the norm. Based on these remarks we will study the dynamics of

$$
\dot{\mathbf{z}} = -\mathcal{L}(t, \mathbf{z}_t) + \boldsymbol{\eta}(t) \tag{20}
$$

where $\boldsymbol{\eta}(t) = \big(\eta_1(t), \ldots, \eta_N(t)\big)$ with

$$
\eta_i(t) = \begin{cases} a_{i1}(t)\big(z_1(\lambda_{i1}(t)) - k\big), & 1 \text{ affects } i\\ 0, & o.w. \end{cases}
$$

Note that the way the system is defined, $\boldsymbol{\eta}$ is a state independent perturbation that vanishes as fast as $\frac{1}{h(t)}$. Using the linear variation of constants formula [21], $\mathbf{z}$ satisfies

$$
\mathbf{z}_t(t_0, \boldsymbol{\phi}) = \mathbf{x}_t(t_0, \boldsymbol{\phi}) + \int_{t_0}^t T(t, s)(\boldsymbol{x}_{t_0}\boldsymbol{\eta}(s))\,ds
$$

where $\boldsymbol{x}_{t_0}(q) = \mathbf{0}$ if $q \in [-\tau, 0)$ and $\boldsymbol{x}(0) = \mathbb{I}$ is an auxiliary function.

Since $T(t,s)$ projects any vector to Δ, exponentially fast with the common element and the first component of the vector $\boldsymbol{\eta}(s)$ is 0, then $\boldsymbol{\eta}$ is by construction orthogonal to $(1,0,\dots,0)$, we see that

$$\int_{t_0}^{t} T(t,s)(\boldsymbol{x}_{t_0}\boldsymbol{\eta}(s))\,ds = \int_{t_0}^{t}\Big(T(t,s)-\mathbb{1}(1,0,\dots,0)\Big)(\boldsymbol{x}_{t_0}\boldsymbol{\eta}(s))\,ds$$

also $||\mathbf{x}(t,t_0,\phi)-\mathbb{1}k||_\infty \le e^{-\theta(t-t_0)}$ we conclude that $\mathbf{z}$ satisfies

$$\mathbf{z}(t,t_0,\boldsymbol{\phi})-\mathbb{1}k = \big(\mathbf{x}(t,t_0,\boldsymbol{\phi})-\mathbb{1}k\big)+\int_{t_0}^{t}\Big(T(t,s)-\mathbb{1}(1,0,\dots,0)\Big)(\boldsymbol{x}_{t_0}\boldsymbol{\eta}(s))\,ds$$

and this implies

$$||\mathbf{z}(t,t_0,\boldsymbol{\phi})-\mathbb{1}k||_\infty \le e^{-\theta(t-t_0)}+\int_{t_0}^{t}\bar{\Theta}e^{-\theta(t-s)}\frac{1}{h(s)}ds$$

for some $\bar{\Theta}$ that depends on Θ and the parameters a_{i1}, τ_{i1} which is well-defined and finite. Condition (11) then applies to conclude the proof.∎

Proof of Corollary 9.3. The proof follows the steps of the argumentation presented for the linear case and especially Lemma 5.5. That proof is applied here as well to show that with initial data to satisfy $S_{I_{t_0}}(\boldsymbol{\phi})\subset W$, the solutions never escape $[\min_{i\in V}\min_{s\in I_{t_0}}\phi_i(s), \max_{i\in V}\max_{s\in I_{t_0}}\phi_i(s)]$ and hence the passivity property is preserved throughout $\mathbf{x}$. For an arbitrary but fixed $\mathbf{x}$ of (12) we define

$$a_{ij}(t) := \frac{g_{ij}\big(t, x_j(\lambda_{ij}(t))-x_i(t)\big)}{x_j(\lambda_{ij}(t))-x_i(t)}$$

and consequently the initial value problem

$$i\in V:\begin{cases}\dot{y}_i(t)=\sum_j a_{ij}(t)\big(y_j(\lambda_{ij}(t))-y_i(t)\big), & t\ge t_0\\ y_i(t)=\phi_i(t), & t\in I_{t_0}\end{cases}$$

Under Assumption 9.2 we see that $a_{ij}(t)$ are well defined taking strictly positive values. Then certain connectivity condition can make them satisfy the assumptions of Theorem 7.1. In view of the fact that $\mathbf{y}$ is indistinguishable of $\mathbf{x}$ the result follows.
∎

Proof of Corollary 9.4. The proof is again based on a direct linearization as in Corollary 9.3 with a minor modification. At first it is essential to show that whenever starting in $\phi_i(t)\in W$ for $t\in I_{t_0}$ the solution $\mathbf{x}=\mathbf{x}(t,t_0,\boldsymbol{\phi})$, $t\ge t_0$ of (13) never escapes W, i.e. it is bounded and defined for all times. This can be shown with the same argument of Lemma 5.5. System (13) can be written as

$$\dot{x}_i(t)=\sum_j a_{ij}(t,\mathbf{x})\big(x_j(\lambda_{ij}(t))-x_i(t)\big)$$

for

$$a_{ij}(t) := \int_0^1 g'_{ij}\big(t, sx_j(\lambda_{ij}(t))+(1-s)x_i(t)\big)\,ds.$$

and $g'_{ij}(t,x)=\frac{\partial g_{ij}(t,x)}{\partial x}$. Since W is open and connected subset of $\mathbb{R}$ and solutions never escape it, the function a_{ij} are well-defined and they take strictly positive values for all $t\ge t_0$. The argumentation again follows Corollary 9.3 and Theorem 7.1 to prove convergence.
∎

REFERENCES

[1] M. Arcak. Passivity as a design tool for group coordination. *IEEE Transactions on Automatic Control*, 52(8):1380–1390, 2007.

[2] O. Arino and E. Haourigui. On the asymptotic behavior of solutions of some delay differential systems which have a first integral. *Journal of Mathematical Analsysis and Applications*, 122:36–46, 1987.

[3] L. O. Chua, T. Yang, G. Q. Zhong, and C. W. Wu. Adaptive synchronization of chua's oscillators. *International Journal of Bifurcation and Chaos*, 6(1):189–201, 1996.

[4] F. R. K. Chung. *Spectral Graph Theory*, volume 92. American Mathematical Society, 1997.

[5] F. Cucker and S. Smale. Emergent behavior in flocks. *IEEE Transactions on Automatic Control*, 52(5):852–862, 2007.

[6] M. H. DeGroot. Reaching a consensus. *Journal of American Statistical Association*, 69:118–121, 1974.

[7] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford Univ. Press, 3 edition, 2003.

[8] I. Györi. Connections between compartmental systems with pipes and integro-differential equations. *Mathematical Modeling*, 7:1215–1238, 1986.

[9] S-Y. Ha and Tadmor E. From particle to kinetic and hydrodynamic description of flocking. *Kinetic and Related Models*, 1(3):415–435, 2008.

[10] S-Y. Ha and J. G. Liu. A simple proof of the cucker-smale flocking dynamics and mean-field limit. *Commun. Math. Sci.*, 7(2):297–325, 2009.

[11] J. K. Hale and S.M.V. Lunel. *Introduction to Functional Differential Equations*, volume 99. Springer-Verlag, 1993.

[12] D. J. Hartfiel. *Markov Set-Chains*, volume 1695. Springer, 1998.

[13] I. C. F. Ipsen and T. M. Selee. Ergodicity coefficients defined by vector norms. *SIAM J. Matrix Anal. Appl.*, 32:153–200, March 2011.

[14] A. J. Jacquez and P. C. Simon. Qualitative theory of compartmental systems with lags. *Mathematical Biosciences*, 180:329–362, 2002.

[15] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.

[16] J. Jiang. Asymptotic behavior for systems comparable to quasimonotone systems. *Fields Institute Communications*, 48:201–211, 2006.

[17] U. Krause. A discrete nonlinear and non-autonomous model of consensus formation. *Communications in difference equations*, pages 227–236, 2000.

[18] T. Krisztin. On the rate of convergence of solutions of functional differential equations. *Funkcialaj Ekvacioj*, 29:1–10, 1986.

[19] Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Springer-Verlag, 1984.

[20] V. Lakshminkantham and Deo S.G. *Method of Variation of Parameters for Dynamic Systems*, volume 1. Gordon and Breach Science, 1998.

[21] W. Lu, F. M. Atay, and J. Jost. Synchronization of discrete-time dynamical networks with time-varying couplings. *SIAM J. Math. Anal.*, 39(4):1231–1259, 2007.

[22] A. A. Markov. Extensions of the law of large numbers to dependent quantities (in russian). *Izn. Fiz. - Matern. Obsch. Kazan Univ.*, 15, 1906.

[23] I. Matei, J. S. Baras, and C. Somarakis. Convergence results for the linear consensus problem under markovian random graphs. *SIAM J. Control Optim.*, pages 1574–1591, 2013.

[24] I. Matei, C. Somarakis, and J. S. Baras. A generalized gossip algorithm on convex metric spaces. *IEEE Trans. Automat. Contr.*, 2015. to appear.

[25] M. Mesbahi and M. Egerstedt. *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press, 2010.

[26] U. Mnz, A. Papachristodoulou, and F. Allgower. Delay robustness in consensus problems. *Automatica*, 46(8):1252 – 1265, 2010.

[27] L. Moreau. Stability of continuous-time distributed consensus algorithms. In *43rd IEEE Conference on Decision and Control*. IEEE, 2004.

[28] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(02):169–182, 2005.

[29] S. Motsch and E. Tadmor. Heterophilious dynamics enhances consensus. *SIAM: Review*, 56(4):577–621, 2014.

[30] A. Nediċ and A. Ozdaglar. Convergence rate for consensus with delays. *Journal of Global Optimization*, 47(3):437–456, 2010.

[31] R. Olfati-Saber and R. M. Murray. Consensus problems in networks with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(09), 2004.

[32] A. Papachristodoulou, A. Jadbabaie, and U. Mnz. Effects of delay in multi-agent consensus and oscillator synchronization. *IEEE Trans. Automat. Contr.*, 55(6):1471–1477, 2010.

[33] H. Qing and W. M. Haddad. Distributed nonlinear control algorithms for network consensus. *Automatica*, 44(9):2375 – 2381, 2008.

[34] F. Riesz and B. Szőkefalvi-Nagy. *Functional Analysis*. Frederick Ungar Publishing, New York, 1972.

[35] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer, 2006.

[36] H. L. Smith. *Monotone Dynamical Systems, An Introduction to the Theory of Competitive and Cooperative Systems*, volume 41. AMS, 1995.

[37] C. Somarakis and J. S. Baras. Delay-independent convergence for linear consensus networks with applications to non-linear flocking systems. In *The 12th IFAC Workshop on Time Delay Systems*, Ann Arbor,MI,USA, June 2015.

[38] C. Somarakis and J. S. Baras. A simple proof of the continuous time linear consensus problem with applications in non-linear flocking networks. In *The 14th European Control Conference*, Linz,Austria, July 2015.

[39] C. Somarakis and J.S. Baras. A general framework for consensus networks. Technical Report doi:10.13016/M2WK77, Digital Repository at UMD, 2015.

[40] J. Tsitsiklis, D. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, 1986.

[41] 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.*, 75(6):1226–1229, 1995.

[42] C. W. Wu, T. Yang, and L. O. Chua. On adaptive synchronization and control of nonlinear dynamical systems. *International Journal of Bifurcation and Chaos*, 6(2), 1996.