A distributed methodology for approximate uniform global minimum sharing[☆]

Michelangelo Bin [a,*], Thomas Parisini [a,b,c]

[a] *Department of Electrical and Electronic Engineering, Imperial College London, London, UK*
[b] *KIOS Research and Innovation Center of Excellence, University of Cyprus, Cyprus*
[c] *Department of Engineering and Architecture at University of Trieste, Trieste, Italy*

A R T I C L E   I N F O

A B S T R A C T

The paper deals with the distributed minimum sharing problem: a set of decision-makers compute the minimum of some local quantities of interest in a distributed and decentralized way by exchanging information through a communication network. We propose an adjustable approximate solution which enjoys several properties of crucial importance in applications. In particular, the proposed solution has good decentralization properties and it is scalable in that the number of local variables does not grow with the size or topology of the communication network. Moreover, a global and uniform (both in the initial time and in the initial conditions) asymptotic stability result is provided towards a steady state which can be made arbitrarily close to the sought minimum. Exact asymptotic convergence can be recovered at the price of losing uniformity with respect to the initial time.

1. Introduction

1.1. Problem description, objectives and context

We consider the problem of computing the minimum of a set of numbers over a network, and we propose a distributed, iterative solution achieving *global* and *uniform*, albeit *approximate*, asymptotic stability. We are given a set $\mathcal{N}$ of N decision makers (or *agents*), where each agent $i \in \mathcal{N}$ is provided with a number $M_i \in \mathbb{R}_{\geq 0}$ not known a priori by the others. The agents exchange information over a communication network with only a subset of other agents (called their *neighborhood*). The approximate minimum sharing problem consists in the design of an algorithm guaranteeing that each agent asymptotically obtains a "sufficiently good" estimate of the quantity

$$M^{\star} := \min_{i \in \mathcal{N}} M_i. \tag{1}$$

Clearly, "$x_i = M^{\star}, \ \forall i \in \mathcal{N}$" is also the *unique* solution to every constrained optimization problem of the form

$$
\begin{aligned}
\max \ & \sum_{i \in \mathcal{N}} \psi_i(x_i) \\
& x_i \leq M_i, && \forall i \in \mathcal{N} \\
& x_i = x_j, && \forall i, j \in \mathcal{N}
\end{aligned}
\tag{2}
$$

obtained with ψ_i, $i \in \mathcal{N}$, continuous and strictly increasing functions. Therefore, the minimum sharing problem is equivalent to the constrained distributed optimization problem (2), thus intersecting the wide research field of distributed optimization (Notarstefano, Notarnicola, & Camisa, 2019).

The problem of computing a minimum (or, equivalently, a maximum) over a network of decision makers is a classical problem in multi-agent control, with applications in distributed estimation and filtering, synchronization, leader election, and computation of network size and connectivity (see, e.g., Bullo, Cortes, and Martinez (2009), Golfar and Ghaisari (2019), Iutzeler, Ciblat, and Jakubowicz (2012), Nejad, Attia, and Raisch (2009), Santoro (2006) and the references therein). Perhaps the most elementary existing algorithms solving the minimum sharing problem are the *FloodMax* (Bullo et al., 2009) and the *Max-Consensus* (Golfar & Ghaisari, 2019; Iutzeler et al., 2012; Nejad et al., 2009). In its simplest form, Max-Consensus[1] requires each

 * Corresponding author.
 E-mail addresses: m.bin@imperial.ac.uk (M. Bin), t.parisini@imperial.ac.uk (T. Parisini).

[1] For brevity, we only focus on Max-Consensus. However, the same conclusions applies also to the FloodMax.

agent $i \in \mathcal{N}$ to store an estimate $x_i \in \mathbb{R}$ of $M^\star$ which is updated iteratively on the basis of the following update rule

$$x_i^{t+1} = \min_{j \in [i]} x_j^t, \qquad \forall i \in \mathcal{N}, \tag{3a}$$

with the initialization

$$x_i^{t_0} = M_i, \qquad \forall i \in \mathcal{N}, \tag{3b}$$

where t is the iteration variable, t_0 its initial value, and $[i] \subset \mathcal{N}$ denotes the neighborhood of agent i (we assume $i \in [i]$). The update law (3a) is decentralized and scalable, in that each agent needs only information coming from its neighbors and each agent stores only one variable. However, although (3a) guarantees convergence of each x_i to $M^\star$ when the estimates x_i are initialized as specified in (3b), *convergence is not guaranteed for an arbitrary initialization*. In fact, if

$$\exists i \in \mathcal{N} \text{ s.t. } x_i^{t_0} < M^\star, \tag{4}$$

then the corresponding estimate x_i^t produced by (3a) satisfies $x_i^t < M^\star$ all subsequent t, so that $x_i^t \to M^\star$ cannot hold.[2] Therefore, since convergence to $M^\star$ holds only for some specific initial values $x_i^{t_0}$, the Max-Consensus algorithm (3) is not *globally convergent*. While there are application domains for which attaining global convergence is not strictly necessary, there are many others in which it is a crucial requirement. This is the case, for instance, when the quantities M_i can change at run time (see the two use-cases illustrated in Section 1.2). To see how this may be a problem for the update law (3), assume by way of example that the estimates x_i^t have reached at a given t_1 the value $M^\star$, i.e. $x_i^{t_1} = M^\star$ for all $i \in \mathcal{N}$, and assume that there is a unique $k \in \mathcal{N}$ such that $M^\star = M_k$. Now, suppose that at some $t_2 > t_1$ the value of M_k increases, thus determining an increment also of $M^\star$. Then, the condition (4) holds for $t_0 = t_2$ so as, in view of the discussion above, the update law (3a) fails to track the new minimum.

Global attractiveness is not the only desirable property one may be interested in when the minimum sharing problem is considered over large networks with possibly changing conditions. In fact, a crucial role is also played by

1. *Uniformity of the convergence*: the convergence rate does not depend on the initial value t_0 of the iteration variable and is constant over compact subsets of initial conditions.
2. *Stability of the steady state*: ensures that small variations in the parameters and initial conditions map into small deviations from the unperturbed trajectories.
3. *Scalability*: the number of variables stored by each agent does not grow with the network size or the number of interconnections.
4. *Decentralization of the updates*: the update law of each agent uses only local information and depends on parameters that are independent from those of the other agents.

Indeed, uniform global attractiveness and stability of the steady state confer robustness against uncertain and time-varying conditions and parameters (see e.g. Goebel, Sanfelice, and Teel (2012, Chapter 7)), making the minimum sharing method suitable for applications in which the quantities M_i vary in time. Moreover, scalability and decentralization enable the application to large-scale networks. In this direction, in this paper we look for a

novel solution to the minimum sharing problem having scalability and decentralization properties similar to those of Max-Consensus (3), but, in addition, possessing the aforementioned globality, uniformity and stability properties.

1.2. Motivating applications

Our methodology is motivated by two application contexts described below. In both cases, a key element consists in solving an instance of the minimum-sharing problem (2) in which the parameters M_i, hence the minimum $M^\star$, may change over time. In this contexts, (i) global attractiveness allows to track the changing minimum $M^\star$, (ii) uniformity of convergence guarantees that the convergence rate is always the same and does not decrease with time, and (iii) stability guarantees that relatively small variations of the parameters lead to small transitory deviations from the optimal steady state.

1.2.1. Cooperative control of traffic networks

Consider a traffic network consisting of a set of vehicles driving on a highway in an intense traffic situation. Some of the vehicles have self-driving capabilities, and we can assign their driving policies. The other vehicles are instead human-driven and, thus, they are not controlled. The whole traffic network is seen as a *plant* that, when not properly controlled, may exhibit undesired behaviors, such as ghost jams. The control goal consists in finding a control policy, distributed among the self-driving vehicles, which guarantees that the "closed-loop" traffic network behaves properly, leading to a smooth traffic flow where all the vehicles hold a common maximal cruise speed. At each time, the maximum attainable cruise speed of each vehicle i is constrained by a *personal maximum value*, denoted by M_i, which may depend on mechanical constraints, on the traffic conditions, on standing speed limitations, or other exogenous factors. A key part of the control task consists in the distributed computation of the maximum common cruise speed, $M^\star$, compatible with all the personal velocity constraints. At each time, the problem of estimating $M^\star$ is an instance of (2), whose solution is precisely (1).

1.2.2. Dynamic leader election

Another important motivating application is the distributed *leader election* problem in dynamic networks, which shares many similarities with the previous application. Single-leader election has been proved to be an unsolvable problem in general, even under bi-directionality, connectivity, and total reliability assumptions on the communication networks (Santoro, 2006, Theorem 3.1.1). A standard additional assumption making the problem well-posed is that each agent is characterized by a *unique identifier* M_i. Hence, the problem of leader election can be cast as finding the minimum, $M^\star$, of such identifiers. The agent whose identifier coincides with $M^\star$ declares itself the leader, the others the followers.

1.3. Related works and state of the art

Classical algorithmic approaches to the minimum sharing problem in arbitrary networks have been developed in the context of distributed algorithms and robotic applications. They include the *FloodMax* (Bullo et al., 2009), the *Max-Consensus* (Golfar & Ghaisari, 2019; Iutzeler et al., 2012; Nejad et al., 2009) (see (3)), the *MegaMerger* (Gallager, Humblet, & Spira, 1983), and the *Yo-Yo* algorithm. See Bullo et al. (2009) and Santoro (2006) for a more detailed overview. Some of these approaches, such as the basic Max-Consensus (3), have nice scalability and decentralization properties: the update laws do not depend on *centralized quantities*, such as parameters that need to be known in advance by all

---

[2] In this specific case, we also observe that any *consensual* configuration (i.e., $x_i = x_j$ for all $i, j \in \mathcal{N}$) is an equilibrium of (3a). This, in turn, is intimately linked to the unfeasibility result of Santoro (2006, Theorem 3.1.1), and to the *detectability* issues appearing in many control problems, such as *Extremum Seeking* (Ariyur & Krstić, 2003; Tan, Nešić, & Mareels, 2006).

the agents, and employ a number of local variables which does not grow with the network size or topology. However, all such approaches require a correct initialization or a pre-processing synchronization phase, which are undesired limitations in applications of interest such as, for example, the ones discussed in Section 1.2.

If the minimum sharing problem is cast in terms of the optimization problem (2), then one can rely on a well-developed literature on discrete-time distributed optimization (see Notarstefano et al. (2019) for a recent overview). If the functions ψ_i in (2) are convex, indeed, different approaches can be used, such as consensus-based (sub)gradient methods (Lobel & Ozdaglar, 2011; Nedic & Ozdaglar, 2009; Nedic, Ozdaglar, & Parrilo, 2010; Shi, Ling, Wu, & Yin, 2015a, 2015b; Yuan, Ling, & Yin, 2016), second-order methods (Mokhtari, Ling, & Ribeiro, 2017; Varagnolo, Zanella, Cenedese, Pillonetto, & Schenato, 2016), projected (Xie, You, Tempo, Song, & Wu, 2018) and primal–dual (Chang, Nedić, & Scaglione, 2014; Zhu & Martinez, 2012) methods with inequality constraints, methods based on the distributed Alternate Direction Method of Multipliers (ADMM) (Bastianello, Carli, Schenato, & Todescato, 2020; Boyd, Parikh, Chu, Peleato, & Eckstein, 2011; Chang, 2016; Jakovetić, Moura, & Xavier, 2015; Ling, Shi, Wu, & Ribeiro, 2015; Makhdoumi & Ozdaglar, 2017; Mota, Xavier, Aguiar, & Püschel, 2013; Notarstefano et al., 2019; Shi, Ling, Yuan, Wu, & Yin, 2014), and methods based on gradient tracking (Bin, Notarnicola, Marconi, & Notarstefano, 2019; Nedić, Olshevsky and Shi, 2017; Nedić, Olshevsky, Shi and Uribe, 2017; Qu & Li, 2018; Xi, Xin, & Khan, 2018; Xu, Zhu, Soh, & Xie, 2015). Gradient methods typically achieve global attractiveness. However, among the cited references only (Nedic et al., 2010) deals with constrained problems with different *local constraints*[3] such as (2). Yet, Nedic et al. (2010) require a vanishing stepsize, which makes convergence not uniform. Gradient methods employing a fixed stepsize thus guaranteeing uniformity are given in Lobel and Ozdaglar (2011), Mokhtari et al. (2017), Nedic and Ozdaglar (2009), Shi et al. (2015a, 2015b), Varagnolo et al. (2016) and Yuan et al. (2016). However, they do not cover constrained problems of the kind (2). Moreover, the first-order methods in Lobel and Ozdaglar (2011), Nedic and Ozdaglar (2009) and Yuan et al. (2016) lead to an approximate convergence result in which the convergence speed and the approximation error need to be traded off. This, in turn, is consistent with our results in which a compromise is more generally established between uniformity, approximation error and convergence rate. The approaches (Chang et al., 2014; Xie et al., 2018; Zhu & Martinez, 2012) deal with inequality constraints including Problem (2). Nevertheless, they require a correct initialization and, hence, they do not provide global attractiveness. The same issue applies to gradient-tracking methods (Bin et al., 2019; Nedić, Olshevsky and Shi, 2017; Nedić, Olshevsky, Shi and Uribe, 2017; Qu & Li, 2018; Xi et al., 2018; Xu et al., 2015) (which, anyway, are developed for unconstrained problems), and also for the "node-based" formulations of ADMM (Jakovetić et al., 2015; Ling et al., 2015; Makhdoumi & Ozdaglar, 2017; Mota et al., 2013; Shi et al., 2014). Instead, the "edge-based" formulations of ADMM (e.g. Bastianello et al. (2020) and Notarstefano et al. (2019, Section 3.3)) do not suffer from this initialization issue, and they provide a solution which is global and uniform. Nevertheless, the number of variables that each agent has to store grows with the dimension of its neighborhood, thus incurring in scalability issues. Moreover, stability is not usually considered in the analysis of the aforementioned designs, and typically the update laws employ coefficients

(e.g. stepsizes) which must be common[4] to all agent (i.e., they are *centralized* quantities).

1.4. Contributions & organization of the paper

We propose a new approach to the minimum sharing problem that provides an adjustable *approximate* (or *sub-optimal* in terms of (2)) solution enjoying the globality, uniformity, scalability and decentralization properties stated in Section 1.1, which do not seem to be possessed altogether by any existing algorithm. The proposed update laws have the form

$$x_i^{t+1} = f_i(t, x^t), \tag{5}$$

for some suitable functions f_i, where $x_i \in \mathbb{R}$ represents the estimate of $M^\star$ stored by agent i, and $x := (x_i)_{i \in \mathcal{N}}$ is the aggregate estimate. As formally specified later on in Section 2.2, the actual structure of the functions f_i encodes the decentralization constraints, allowing an agent update to depend only on the estimates of a subset of other agents (see Remark 1). We show that all the estimates x_i converge, globally and uniformly, to a stable neighborhood of $M^\star$ whose size can be reduced arbitrarily around $M^\star$ by suitably tuning some control parameters. More precisely, the proposed approach enjoys the following properties:

(a) The algorithm is distributed and scalable, since the only one variable is stored for each agent.
(b) The update law of each agent employs a gain which can be tuned independently from the others.
(c) The estimates x_i converge globally and uniformly to a stable steady state which can be made arbitrarily close to $M^\star$.
(d) Exact convergence (i.e., all the estimates converge to $M^\star$) can be achieved, at the price, however, of losing uniformity.

In view of Item (b), the proposed method has good decentralization properties compared to most of the approaches mentioned in Section 1.3. Nevertheless, we underline that the proposed method is not fully decentralized, as the agents are supposed to know a lower-bound on $M^\star$ (Assumption 2) which explicitly enters in the update laws.

The paper is organized as follows. After providing preliminary definitions and remarks in Section 2, in Section 3 we formulate the minimum-sharing problem and we describe the proposed solution methodology. The main convergence results are given in Section 4 and proved in Section 7. Finally, numerical results and concluding remarks are reported in Sections 5 and 6, respectively.

2. Preliminaries

2.1. Notation

We denote by $\mathbb{R}$ and $\mathbb{N}$ the set of real and natural numbers respectively. If $a \in \mathbb{R}$, $\mathbb{R}_{\geq a}$ denotes the set of all real numbers larger than or equal to a, and similar definitions apply to other ordered sets and ordering relations. We denote by $\operatorname{card} A$ the cardinality of a set A. If $A, B \subset \mathbb{R}$, $A \setminus B := \{a \in A \mid a \notin B\}$ denotes the set difference between A and B. We identify singletons with their unique element and, for a $b \in \mathbb{R}$, we thus write $A \setminus b$ in place of $A \setminus \{b\}$. We denote norms by $|\cdot|$ whenever they are clear from

---

[3] By the term "local constraints" we refer to private constraints an agent may have on its own variables that do not depend on the other agents' variables, e.g. the constraints $x_i \leq M_i$ in (2).

[4] Exceptions are given in the gradient-tracking designs of Nedić, Olshevsky, Shi and Uribe (2017) and Xu et al. (2015), where agents employ uncoordinated stepsizes. In both the designs, the discrepancy between the stepsizes must be small enough. Hence, these results may be seen as a "robustness" property relative to variations of the stepsizes with respect to their average. In turn, this property comes *for free* if the algorithm is proved to be *asymptotically stable* with a common stepsize (see, e.g., Goebel et al. (2012, Chapter 7)).

the context. With $A \subset \mathbb{R}^n$ and $x \in \mathbb{R}^n$, dist $(x, A) := \inf_{a \in A} |x - a|$ denotes the distance from x to A. Sequences indexed by a set S are denoted by $(x_s)_{s \in S}$. For a non-empty interval $[a, b] \subset \mathbb{R}$, we define the projection map $\Pi_{[a,b]} : \mathbb{R} \to [a, b]$ as $\Pi_{[a,b]}(s) := \min\{\max\{s, a\}, b\}$. A function $f : \mathbb{R}^n \to \mathbb{R}^m$, $n, m \in \mathbb{N}$, is *locally bounded* if $f(K)$ is bounded for each compact set $K \subset \mathbb{R}^n$. In this paper, we consider discrete-time systems whose solutions are signals defined on a non-empty subset dom x of $\mathbb{N}$. For ease of notation, we will use x^t in place of $x(t)$ to denote the values of a signal x. With $t_0 \in \mathbb{N}$, we say that x *starts at* t_0 if $\min \mathrm{dom}\, x = t_0$.

2.2. Communication networks

Throughout the paper, $\mathcal{N}$ denotes the (finite) set of agents in the network, and we let $N := \mathrm{card}\, \mathcal{N}$. The network communication constraints are formally captured by the concept of "communication structure" defined below.[5]

Definition 1. A *communication structure* on $\mathcal{N}$ is a sequence $\mathcal{C} = ([i])_{i \in \mathcal{N}}$ of subsets $[i]$ of $\mathcal{N}$ which satisfy $i \in [i]$.

For each $i \in \mathcal{N}$, the set $[i]$ is called the *neighborhood* of i. A *communication network* is a pair $(\mathcal{N}, \mathcal{C})$, in which $\mathcal{N}$ is a set and $\mathcal{C}$ is a communication structure on $\mathcal{N}$.

For a given $I \subset \mathcal{N}$, we define the sequence of sets

$$
\begin{aligned}
[I]^0 &:= I \\
[I]^n &:= \bigcup_{j \in [I]^{n-1}} [j], \quad n \in \mathbb{N}_{\geq 1}
\end{aligned}
\tag{6}
$$

so as, in particular, $[\{i\}]^1 = [i]$. If $I = \{i\}$ is a singleton, we use the short notation $[\{i\}]^n = [i]^n$. Moreover, for $n, m \in \mathbb{N}$ we let

$$[I]^n_m := [I]^n \setminus [I]^m.$$

We consider networks that are *connected* according to the following definition.

Definition 2. With $I \subset \mathcal{N}$, a communication network $(\mathcal{N}, \mathcal{C})$ is said to be *I-connected* if there exists $n_I \leq N$ such that $[I]^{n_I} = \mathcal{N}$.

The notion of *I-connectedness* is in general weaker than usual *strong connectedness*, which requires the existence of a path between any two agents. Later on, we shall assume that $\mathcal{N}$ is given a communication structure $\mathcal{C}$ which is $I^\star$-connected for a specific subset $I^\star \subset \mathcal{N}$. For the purpose of analysis, this communication structure is assumed static. Likewise also the quantities M_i are supposed constant. In fact, this corresponds to a well-defined "nominal setting" for the proposed method in which we can prove the desired uniform global attractiveness and stability properties. Proving such properties in the nominal case, in turn, guarantees that the proposed method can be applied also to relevant classes of problems where the communication structure and the parameters M_i (hence, their minimum $\mathsf{M}^\star$) may change over time. Indeed, as already mentioned in Section 1.1, uniform global attractiveness and stability ensure a proper approximate tracking of a time-varying minimum $\mathsf{M}^\star$ provided that its dynamics is sufficiently slow. Moreover, classical results in the context of control under different time-scales (see, e.g., Kokotović, Khalil, and O'Reilly (1999), Tan et al. (2006), Teel, Moreau, and Nešić (2003) and Wang, Teel, and Nešić (2012)) also guarantee good tracking performances under changes of the communication structure $\mathcal{C}$ that are, on average, sufficiently slow with respect to the dynamics of the update laws. In this respect, Section 5 provides numerical results in a scenario in which the communication structure and the numbers M_i are subject to impulsive changes separated by relatively large intervals of time.

---

[5] A common way to define a communication structure on $\mathcal{N}$ is to consider an undirected graph $(\mathcal{N}, \mathcal{E})$ with vertices set equal to $\mathcal{N}$ and edges set $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$ such that if $(i, j) \in \mathcal{E}$ then agents i and j can communicate. In this case, $[i] := \{i\} \cup \{j \in \mathcal{N} \mid (j, i) \in \mathcal{E}\}$.

2.3. Stability and convergence notions

We consider discrete-time systems of the form

$$x^{t+1} = f(t, x^t), \tag{7}$$

with state $x^t \in \mathbb{R}^n$, $n \in \mathbb{N}$. Given a closed set $\mathcal{A} \subset \mathbb{R}^n$, we say that $\mathcal{A}$ is *stable* for (7) if for each $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that every solution of (7) satisfying dist $(x^{t_0}, \mathcal{A}) \leq \delta(\epsilon)$ also satisfies dist $(x^t, \mathcal{A}) \leq \epsilon$, for all $t \geq t_0$. We say that $\mathcal{A}$ is *attractive* for (7) if there exists an open superset $\mathcal{O}$ of $\mathcal{A}$ and, for every $t_0 \in \mathbb{N}$, every solution x to (7) with $x^{t_0} \in \mathcal{O}$, and every $\epsilon > 0$, there exists $t^\star(t_0, x^{t_0}, \epsilon) \in \mathbb{N}$, such that dist $(x^t, \mathcal{A}) \leq \epsilon$ holds for all $t \geq t_0 + t^\star(t_0, x^{t_0}, \epsilon)$. Different qualifiers can enrich this attractiveness property. In particular, the set $\mathcal{A}$ is said to be:

- *Globally attractive* if $\mathcal{O} = \mathbb{R}^n$.
- *Finite-time attractive* if the condition "$\epsilon > 0$" can be replaced by "$\epsilon \geq 0$".
- *Uniformly attractive in the initial time* t_0 if the map $t^\star(\cdot)$ does not depend on t_0.
- *Uniformly attractive in the initial conditions* x^{t_0} if for each $(t_0, \epsilon) \in \mathbb{N} \times \mathbb{R}_{\geq 0}$, the map $t^\star(t_0, \cdot, \epsilon)$ is locally bounded.
- *Uniformly attractive* if it is both uniformly attractive in the initial time and in the initial conditions.
- *ϵ-approximately attractive* (with $\epsilon > 0$) if the set $\{x \in \mathbb{R}^n \mid \mathrm{dist}\,(x, \mathcal{A}) \leq \epsilon\}$ is attractive.

If $\mathcal{A}$ is both stable and attractive, it is said to be *asymptotically stable*. Moreover, with $(f_\gamma)_{\gamma \in \Gamma}$ representing a family of functions $f_\gamma : \mathbb{N} \times \mathbb{R}^n \to \mathbb{R}^n$ indexed by a set Γ, consider the family of systems

$$x^{t+1} = f_\gamma(t, x^t), \qquad \gamma \in \Gamma. \tag{8}$$

Then, we say that the set $\mathcal{A}$ is *practically attractive* for the family (8), if for each $\epsilon > 0$, there exists $\gamma^\star(\epsilon) \in \Gamma$ such that the set $\mathcal{A}$ is ϵ-approximately attractive for the system (8) obtained with $\gamma = \gamma^\star(\epsilon)$.

3. Distributed minimum sharing

3.1. Problem formulation

We are given a communication network $(\mathcal{N}, \mathcal{C})$. Each agent $i \in \mathcal{N}$ is provided with a number M_i, not known a priori by the others, and it stores and updates a local estimate $x_i \in \mathbb{R}$ of the quantity $\mathsf{M}^\star$ defined in (1). Thus, the problem at hand consists in designing an update law for each agent $i \in \mathcal{N}$ of the form (5) such that the resulting estimates x_i^t converge to $\mathsf{M}^\star$, in some of the senses defined in Section 2.3. The resulting family $f := (f_i)_{i \in \mathcal{N}}$ is called the *distributed methodology*. In the following, we let $x := (x_i)_{i \in \mathcal{N}}$ and we compactly rewrite (5) as

$$x^{t+1} = f(t, x^t). \tag{9}$$

As each agent is allowed to exchange information only with the agents belonging to its neighborhood $[i]$, the functions f_i must respect this constraint. This is formally expressed by the following definitions.

Definition 3. With $V \subset \mathcal{N}$, a function g on $\mathbb{N} \times \mathbb{R}^N$ is said to be *adapted to V* if it satisfies $g(t, x) = g(t, z)$ for every $t \in \mathbb{N}$, and every $x, z \in \mathbb{R}^N$ satisfying $x_i = z_i$ for all $i \in V$.

Definition 4. The function $f = (f_i)_{i \in \mathcal{N}}$ is said to be *$\mathcal{C}$-decentralized* if, for each $i \in \mathcal{N}$, the map f_i is adapted to $[i]$.

Then, the *distributed minimum sharing* problem is defined as follows.

Problem 1. Design a $\mathcal{C}$-decentralized function f, such that the set

$$\mathcal{A} := \{M^\star\}^N \tag{10}$$

is globally attractive for (9).

Remark 1. We stress that, if f is $\mathcal{C}$-decentralized, then each function f_i in (5) depends only on $(x_j)_{j \in [i]}$ and not on the whole state x.

Remark 2. Depending on the additional qualifiers that may characterize the attractiveness property of $\mathcal{A}$ in Problem 1, we may have solutions to Problem 1 in "different senses". In the forthcoming section, we propose a methodology obtaining both global attractiveness and global uniform practical attractiveness of $\mathcal{A}$, depending on the value of some user-decided control parameters. We will show that a compromise between how close we can get to $\mathcal{A}$ and uniformity in the initial time is necessary. In particular, we show that attractiveness is possible only at the price of losing uniformity in the initial time, and that, if such property is needed, then global practical uniform attractiveness is the best we can achieve.

3.2. Standing assumptions

We consider Problem 1 under two main assumptions specified hereafter. We define the set

$$I^\star := \underset{i \in \mathcal{N}}{\arg\min}\, M_i. \tag{11}$$

With the following assumption, we require the communication network to be connected with respect to $I^\star$.

Assumption 1 (*Connectedness*). The communication network $(\mathcal{N}, \mathcal{C})$ is $I^\star$-connected in the sense of Definition 2.

The second assumption, instead, requires each agent to know a lower-bound on $M^\star$.

Assumption 2 (*Consistency*). Each agent $i \in \mathcal{N}$ knows a number $\mu_i \in \mathbb{R}_{>0}$ such that $\mu_i \leq M^\star$.

It is worth noting that Assumption 2 is a "centralized" assumption, in that it asks each agent to know a lower bound on the common, unknown quantity $M^\star$. Nevertheless, it introduces almost no loss of generality in different applications of interest, including those mentioned in Section 1.2, where knowing a lower-bound on $M^\star$ is a mild requirement. For instance, in both the traffic control and leader election problems we can assume that the quantities M_i are integers, so that "$\mu_i \in (0, 1)$ for all $i \in \mathcal{N}$" is a feasible choice requiring no further knowledge on $M^\star$. Furthermore, this assumption is not in principle needed if an approximate or practical attractiveness result is sought. In fact, if for some $I \subset \mathcal{N}$, $\epsilon := \max_{i \in I} \mu_i > M^\star$, then $M^\star \in [0, \epsilon)$, and, as clarified later on by the asymptotic analysis, we are able to claim that the set $[0, \epsilon]^N$ (which includes $M^\star$) is practically attractive for x, with ϵ, however, that can be made arbitrarily small by choosing μ_i accordingly.

In the following we let

$$\underline{\mu} := \min_{i \in \mathcal{N}} \mu_i. \tag{12}$$

3.3. The update laws

The proposed update law is obtained by choosing f so that, for each $i \in \mathcal{N}$, Eq. (5) reads as follows[6]

$$x_i^+ = \Pi_{[\mu_i,\, M_i]}\left[e^{h_i^t} x_i + k_i \sum_{j \in [i]} (x_j - x_i) \right], \tag{13}$$

in which $\mu_i > 0$ is the same quantity of Assumption 2, $k_i > 0$ is a free control gain chosen to satisfy

$$0 < k_i \leq \frac{1}{\mathrm{card}([i] \setminus i)} \tag{14}$$

and $h_i : \mathbb{N} \to \mathbb{R}_{\geq 0}$ is a time signal to be designed later on.

Notice that, as in Nedic et al. (2010), the update laws (13) have the form of a projected (onto the interval $[\mu_i, M_i]$) consensus-like protocol. Unlike Nedic et al. (2010), however, the matrix defining the estimates dynamics needs *not* be column or row-stochastic, and the coefficients k_i are only constrained by (14) and, hence, they can be chosen in a completely decentralized way. Moreover, unlike all the aforementioned distributed optimization approaches, the restriction of the dynamics onto the *consensus manifold*[7] is not marginally stable. Rather, it is deliberately made unstable by the terms $e^{h_i^t}$.

3.4. Excitation properties

The signals h_i will be chosen to guarantee one of the following *excitation properties*.

Definition 5 (*Sufficiency of Excitation*). With $t_0 \in \mathbb{N}$, the family $(h_i)_{i \in \mathcal{N}}$, is said to be *sufficiently exciting from t_0* if there exist $\underline{h}(t_0) > 0$ and $\Delta(t_0) \in \mathbb{N}_{\geq 1}$ such that, for each $m \in \mathbb{N}_{\geq 1}$ satisfying

$$m \leq \frac{1}{\underline{h}(t_0)} \log\left(\frac{M^\star}{\underline{\mu}} \right) \tag{15}$$

and each $i \in \mathcal{N}$, there exists at least one $s_i \in \{t_0 + 1 + (m - 1)\Delta(t_0), \ldots, t_0 + m\Delta(t_0)\}$ such that $h_i^{s_i} \geq \underline{h}(t_0)$.

In qualitative terms, given an initial time t_0, sufficiency of excitation implies that the signals h_i are positive "frequently enough" for a "large enough" amount of time succeeding t_0. When $(h_i)_{i \in \mathcal{N}}$ is sufficiently exciting from *every* t_0, and independently on it, then we say that $(h_i)_{i \in \mathcal{N}}$ enjoys the *uniformity of excitation* property.

Definition 6 (*Uniformity of Excitation*). The family $(h_i)_{i \in \mathcal{N}}$ is said to be *uniformly exciting* if it is sufficiently exciting from every t_0, with $\underline{h}$ and Δ not dependent on t_0.

Uniformity of excitation can be seen as a "uniform in t_0" version of sufficiency of excitation and, in particular, it implies that all the signals h_i take positive values infinitely often. Defined in this way, both these properties are "centralized", in that they employ quantities common to all the agents. However, both can be easily obtained by means of decentralized design policies in which the signals h_i are chosen independently on each other. This is the case, for instance, when the signals h_i are *periodic* (with possibly different periods) and not identically zero, as formalized in the following lemma (proved in Appendix A).

Lemma 1. *Suppose that, for each $i \in \mathcal{N}$, h_i is periodic and there exists $t \in \mathbb{N}$ for which $h_i^t > 0$. Then, the family $(h_i)_{i \in \mathcal{N}}$ is uniformly exciting.*

---

[6] Recall that $\Pi_{[a,b]}(s) := \min\{\max\{s, a\}, b\}$.

[7] That is, the set $\{x \in \mathbb{R}^N \mid x_i = x_j, \ \forall i, j \in \mathcal{N}\}$.

Remark 3. If $h_i^t = 0$ for all $i \in \mathcal{N}$ and $t \in \mathbb{N}$, each of the infinite points of the consensus manifold $\mathcal{M}$ is an equilibrium for (13). Since $\mathrm{M}^\star \in \mathcal{M}$, this implies that $\mathrm{M}^\star$ is a well-defined steady state for (13). However, in this case $\mathrm{M}^\star$ cannot be reached by any of the initial conditions in $\mathcal{M}$, as they are indeed equilibria. This, in turn, is related to the impossibility result (Santoro, 2006, Theorem 3.1.1) in the leader election problem in absence of unique identifiers, and is at the basis of the non-globality of the FloodMax and Max-Consensus algorithms (see Section 1.1). In order to prevent the consensual states in $\mathcal{M}$ to be equilibria, the signals h_i^t must carry enough excitation, in the sense of Definitions 5 or 6. As formally stated later on in Theorem 1, indeed, this permits to recover globality, although it ruins "exactness" of convergence of each estimate x_i to $\mathrm{M}^\star$, being it a consensual state. In these terms, the signals h_i play the same role of the *dithering* signals in Extremum Seeking approaches (Ariyur & Krstić, 2003; Tan et al., 2006).

4. Convergence results

4.1. Main result

For ease of notation, we write the update laws (13) in the compact form (9). The following theorem – which is the main result of the paper – relates the excitation properties of the signals h_i to the asymptotic convergence of the estimates x_i produced by the update laws (13) to $\mathrm{M}^\star$. In particular, it shows that sufficiency of excitation implies convergence (possibly exact) and uniformity of excitation implies uniform convergence, but ruins exactness. Further remarks and insights on the results given in the theorem follow thereafter in Section 4.2.

Theorem 1. *Under Assumptions Assumptions 1 and 2, consider the update laws (13), in which k_i satisfies (14). Suppose that, for a given $t_0 \in \mathbb{N}$, the family $(h_i)_{i \in \mathcal{N}}$ is sufficiently exciting from t_0 in the sense of Definition 5. Then, the following claims hold:*

1. *There exists $t^\star = t^\star(t_0)$ such that every solution x to (9) starting at t_0 satisfies*

$$\begin{aligned} x_i^t &\geq \mathrm{M}^\star, && \forall t \geq t^\star(t_0), \ \forall i \in \mathcal{N} \setminus I^\star \\ x_i^t &= \mathrm{M}^\star, && \forall t \geq t^\star(t_0), \ \forall i \in I^\star, \end{aligned}$$

   with $I^\star$ given by (11).
2. *For each $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that, if*

$$\limsup_{t \to \infty} h_i^t \leq \delta(\epsilon), \quad \forall i \in \mathcal{N}, \tag{16}$$

   then each solution x starting at t_0 satisfies

$$\lim_{t \to \infty} |x_i^t - \mathrm{M}^\star| \leq \epsilon, \quad \forall i \in \mathcal{N}. \tag{17}$$

   In particular, the set

$$\mathcal{A}_\epsilon := \prod_{i \in \mathcal{N}} [\mathrm{M}^\star, \ \min\{\mathrm{M}^\star + \epsilon, \ \mathrm{M}_i\}]$$

   is globally attractive for (9).
3. *If the family $(h_i)_{i \in \mathcal{N}}$ is uniformly exciting in the sense of Definition 6, then $\mathcal{A}_\epsilon$ is globally uniformly attractive.*
4. *If all the signals h_i are non-zero and periodic (with possibly different periods), then there exists a compact set $\mathcal{A}_\epsilon^u \subset \mathcal{A}_\epsilon$ which is globally uniformly attractive and stable, hence, globally uniformly asymptotically stable.*
5. *If*

$$\lim_{t \to \infty} h_i^t = 0, \quad \forall i \in \mathcal{N}$$

   then, the set $\mathcal{A}$, given by (10), is globally attractive for (9), i.e.

$$\lim_{t \to \infty} x_i^t = \mathrm{M}^\star, \quad \forall i \in \mathcal{N}.$$

For the reader's convenience, the proof of Theorem 1 is postponed to Section 7.

4.2. Remarks on the result

Claim 1 of Theorem 1 states that, if the family $(h_i)_{i \in \mathcal{N}}$ is sufficiently exciting, then, in a finite time $t^\star$ the estimates x_i of the agents $i \in I^\star$ satisfying $\mathrm{M}_i = \mathrm{M}^\star$ reach the target value $\mathrm{M}^\star$, while all the other estimates x_i of the remaining agents $i \in \mathcal{N} \setminus I^\star$ become larger than $\mathrm{M}^\star$. The time $t^\star$ is, however, a centralized quantity which depends on the excitation properties of all the signals h_i.

Claim 2 characterizes the asymptotic behavior of the remaining agents, by stating that the update laws (13) are able to drive the estimates x_i arbitrarily close to $\mathrm{M}^\star$, provided that the amplitude of the signals h_i^t is eventually reduced accordingly. As the approximation $\mathcal{A}_\epsilon$ can be made arbitrarily tight, by acting on the asymptotic bounds of h_i accordingly, it turns out that this is a *global practical attractiveness* result of the target set $\mathcal{A}$ (defined in (10)). More precisely, let Γ be the set of all the families $\gamma := (h_i)_{i \in \mathcal{N}}$ of functions $h_i : \mathbb{N} \to \mathbb{R}_{\geq 0}$, and consider a family of systems of the form (8), with $x^t \in \mathbb{R}^N$ and $f_\gamma := (f_\gamma^i)_{i \in \mathcal{N}}$ satisfying

$$f_\gamma^i(t, x) := \Pi_{[\mu_i, \mathrm{M}_i]} \left[e^{h_i^t} x_i + k_i \sum_{j \in [i]} (x_j - x_i) \right]. \tag{18}$$

Then, the second claim of the theorem can be restated as follows.

Corollary 1. *Under the assumptions of Theorem 1, the set $\mathcal{A}$ is globally practically attractive for the family (18).*

Claim 3 of the theorem further strengthen Corollary 1 to a *uniform* global practical asymptotic stability property of $\mathcal{A}$ in presence of uniformity of excitation. Moreover, in the relevant case in which the signals h_i are periodic, Claim 4 guarantees the existence of a compact set included in $\mathcal{A}_\epsilon$ which is globally uniformly asymptotically stable.

Finally, Claim 5 states that, if all the signals h_i^t converge to zero, then a *global attractiveness* result of the target set $\mathcal{A}$ holds (i.e. $x_i^t \to \mathrm{M}^\star$ for all $i \in \mathcal{N}$). However, we observe that, if $h_i^t \to 0$ for some $i \in \mathcal{N}$, then the family $(h_i)_{i \in \mathcal{N}}$ fails to be uniformly exciting, and thus the convergence of the estimates x_i to $\mathrm{M}^\star$ is *not* in general uniform in the initial time t_0. This underlines an important difference between sufficiency and uniformity of excitation: sufficiency of excitation allows exact convergence, but prevents uniformity in the initial time. Uniformity of excitation, instead, guarantees uniform convergence and stability but frustrates exact convergence, guaranteeing only a weaker practical result. This, in turn, reveals a somehow necessary compromise between complexity, uniformity and convergence.

4.3. On the design of the signals h_i

The signals h_i are the only degrees of freedom left to be chosen in the update laws (13). In this respect, Theorem 1 links their amplitude and excitation properties to the corresponding asymptotic behavior of the estimates x_i, thus providing guidelines for their design. Based on the claims of Theorem 1, in this section we discuss some possible designs guaranteeing sufficiency or uniformity of excitation.

4.3.1. Sufficiently exciting designs

Sufficiency of excitation of the family $(h_i)_{i \in \mathcal{N}}$ is guaranteed if each h_i takes "enough" positive values. According to Definition 5,

and in particular to (15), how much is "enough" depends on centralized quantities. In turn, a design of the signals h_i based on the knowledge of t_0 and of the quantities appearing in (15) is undesirable as inevitably centralized and not robust. A simple decentralized way to design a sufficiently exciting family $(h_i)_{i \in \mathcal{N}}$ amounts to choose bounded signals h_i satisfying

$$\sum_{t \in \mathbb{N}} h_i^t = \infty, \qquad \forall i \in \mathcal{N}. \tag{19}$$

This, for instance, can be achieved by simply letting $h_i^t = a_i/(1+t)$ for some arbitrary $a_i > 0$.

Lemma 2. *Suppose that, for each $i \in \mathcal{N}$, the signal h_i is bounded and satisfies (19). Then, the family $(h_i)_{i \in \mathcal{N}}$ is sufficiently exciting in the sense of Definition 5.*

The proof of Lemma 2 follows directly from (19), hence it is omitted.

In view of Claim 5 of Theorem 1, exact convergence of the estimates x_i to $M^\star$ is obtained if $\lim_{t \to \infty} h_i^t = 0$ for all $i \in \mathcal{N}$. Moreover, convergence of h_i to zero is implied by (although not equivalent to) the following property

$$\sum_{t \in \mathbb{N}} (h_i^t)^2 < \infty. \tag{20}$$

It is interesting to notice that Properties (19)–(20) are standard assumptions asked to the *stepsize* in classical *stochastic approximation algorithms* (Kushner & Yin, 1997; Robbins & Monro, 1951), as well as in modern distributed optimization algorithms using vanishing step sizes (Nedic et al., 2010; Notarstefano et al., 2019; Simonetto & Jamali-Rad, 2016). In the context of this paper, these two conditions are simply sufficient conditions for sufficiency of excitation, which can be easily satisfied by decentralized designs of the signals h_i.

4.3.2. Uniformly exciting designs

In view of Definition 6, if every signal h_i is periodic, then $(h_i)_{i \in \mathcal{N}}$ is uniformly exciting. While periodicity is not necessary for uniformity of excitation, it certainly is a relevant design choice due its simplicity and effectiveness. Possible decentralized design choices for periodic signals h_i leading to a uniformly exciting family $(h_i)_{i \in \mathcal{N}}$ are listed below, where the quantities $A_i, T_i, \rho_i > 0$ are arbitrary. From the theoretical viewpoint, all the following options are equally fine. Depending on the application domain, however, some choices may be more convenient than others.

1. *Constant signals*: is the simplest design choice and consists in choosing $h_i^t = A_i$ for all $i \in \mathcal{N}$.
2. *Rectified sinusoids*: different versions can be defined, for instance $h_i^t = A_i|\sin(\pi t/T_i)|$ and $h_i = A_i \max\{0, \sin(2\pi t/T_i)\}$ both have period T_i.
3. *Square waves*: with $\rho_i \in (0, 1]$ playing the role of a duty cycle, square waves have the form

$$h_i^t = A_i \text{step} \left(\text{mod}(t, T_i) - (1 - \rho_i)(T_i) \right) \tag{21}$$

in which $\text{mod}(s) := s - \max\{n \in \mathbb{N} \mid n(T_i + 1) \leq s\}$, and step$(\cdot)$ denotes the *step function* satisfying step$(s) = 0$ for $s < 0$ and step$(s) = 1$ for $s \geq 0$. The signal (21) has period T_i and $h_i^t = A_i$ holds for $\rho_i T_i$ seconds each period.

5. Numerical simulations

In this section, we present two illustrative numerical simulation scenarios. In Scenario 1, a network with a time-changing topology (see Fig. 1) is considered while in Scenario 2, for a fixed network topology, the use of different signals h_i is evaluated.
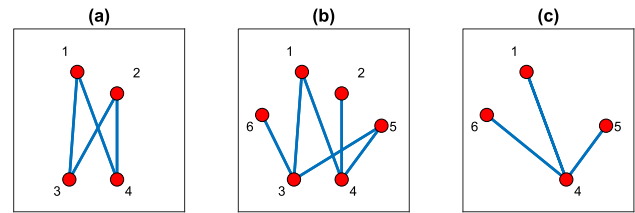
Fig. 1. Communication structure of Simulation 1: (a) $[1] = \{1, 3, 4\}$, $[2] = \{2, 3, 4\}$, $[3] = \{1, 2, 3\}$, $[4] = \{1, 2, 4\}$; (b) $[1] = \{1, 3, 4\}$, $[2] = \{2, 4\}$, $[3] = \{1, 3, 5, 6\}$, $[4] = \{1, 2, 4, 5\}$, $[5] = \{3, 4, 5\}$ and $[6] = \{3, 6\}$; (c) $[1] = \{1, 4\}$, $[4] = \{1, 4, 5, 6\}$, $[5] = \{4, 5\}$ and $[6] = \{4, 6\}$.

5.1. Scenario 1: Uniform convergence

The first simulation, shown in Fig. 2, is obtained as follows. The simulation starts with a network of 4 agents (Agents 1, 2, 3, and 4), provided with a communication structure shown in Fig. 1-(a) and with numbers $(M_1, M_2, M_3, M_4) = (10, 12, 13, 13)$, implying $M^\star = M_1 = 10$. The update laws (13) are implemented with $\mu_i = 1/2$ for all $i \in \{1, \ldots, 4\}$, with $(k_1, k_2, k_3, k_4) = (0.1, 0.08, 0.05, 0.09)$, and with the signals h_i chosen as the square waves discussed in Section 4.3.2 with parameters $(T_1, A_1, \rho_1) = (15, 10^{-3}, 0.2)$, $(T_2, A_2, \rho_2) = (10, 5 \cdot 10^{-4}, 0.5)$, $(T_3, A_3, \rho_3) = (5, 10^{-3}, 0.3)$, $(T_4, A_4, \rho_4) = (10, 5 \cdot 10^{-4}, 0.5)$.

At time $t = 500$, two new agents (Agents 5 and 6) are added to the network, and the communication structure is changed to the one shown in Fig. 1-(b). The new agents have numbers $(M_5, M_6) = (7, 11)$, lower bounds $\mu_5 = \mu_6 = 1/2$, coefficients $(k_5, k_6) = (0.07, 0.1)$, and signals h_i given by the square waves presented in Section 4.3.2 with $(T_5, A_5, \rho_5) = (5, 10^{-3}, 0.4)$ and $(T_6, A_6, \rho_6) = (7, 25 \cdot 10^{-4}, 0.1)$. Furthermore, the numbers of agents 1 and 3 are changed to $(M_1, M_3) = (11, 13)$. The new optimum is thus $M^\star = M_5 = 7$.

At time $t = 1500$, Agents 2 and 3 leave the network, and the communication structure is changed to that depicted in Fig. 1-(c). Moreover, the numbers of the agents are changed to $(M_1, M_4, M_5, M_6) = (12, 16, 11, 16)$, leading to $M^\star = M_5 = 11$.

Finally, at time $t = 5000$, the number of Agent 4 is changed to $M_4 = 8$, so as $M^\star = M_4 = 8$.

As Fig. 2 shows, convergence to the (time-varying) optimum $M^\star$ is approximate, and the trajectories of the agents show residual oscillations. Fig. 2 also underlines that convergence to $M^\star$ "from below" (i.e. when the initial values of the agent are smaller than $M^\star$) is slower than convergence "from above" (i.e. when the initial values of the agent are larger than $M^\star$). As shown in the analysis of Section 7, this is due to the fact that (i) the convergence rate "from below", proved in Section 7.1, is determined by the values of the signals h_i^t, while (ii) the convergence rate "from above", proved in Sections 7.2–7.3, is determined by the values of the coefficients k_i.

For the sake of comparison, Fig. 3 shows a simulation in which the Max-Consensus (3) is employed in the same setting. As shown in Fig. 3, although showing a faster convergence for the first two changes of $M^\star$, the Max-Consensus fails in tracking the other changes. As illustrated in Section 1.1, this is due to the fact that it is not globally attractive.

5.2. Scenario 2: Non-uniform convergence

In the second scenario, we compare two simple networks having the same data and communication structures, but different signals h_i. The first network, $\mathcal{N}$, includes Agents 1, 2, 3 and 4, and it is given the communication structure depicted in Fig. 1(a). Initially, the agents are given numbers (M_1, M_2, M_3, M_4)
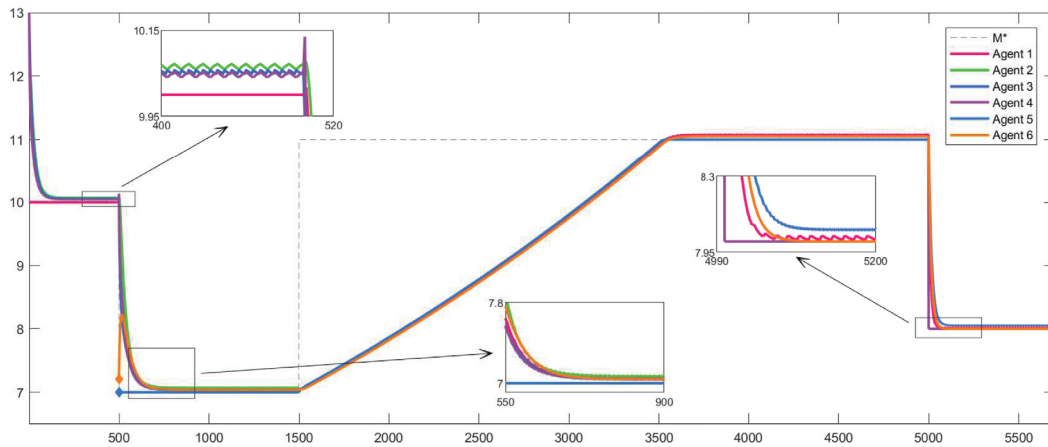
Fig. 2. Evolution of the estimates x_i in Scenario 1. The trajectory of the optimal value M* is shown in dashed gray line. Colored lines depict instead the trajectory of the estimates x_i, $i = 1, \ldots, 6$. In abscissa: iteration variable t. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
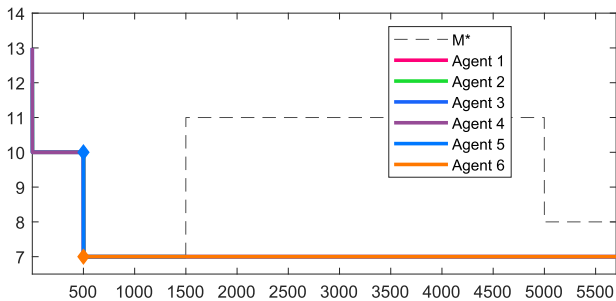
Fig. 3. Evolution of the Max-Consensus estimates (update law (3)) in the setting of Scenario 1 (cf. Fig. 2). In abscissa: iteration variable t.

$= (3, 6, 9, 15)$, so as M$^\star = M_1 = 3$. At time $t = 500$, M$_1$ is changed to 15, so as M$^\star = M_2 = 6$. At time $t = 20\,000$, M$_2$ is changed to 15, so as M$^\star = M_3 = 9$. At time $t = 35\,000$, M$_3$ is changed to 12, so as M$^\star = M_3 = 12$. Finally, at time $t = 150\,000$, M$_3$ is changed to 15, so as M$^\star = M_1 = M_2 = M_3 = M_4 = 15$. The update laws are implemented with $(k_1, k_2, k_3, k_4) = (0.1, 0.08, 0.05, 0.09)$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 1/2$, and with a family $(h_i)_{i \in \mathcal{N}_1}$ of uniformly exciting signals defined as square waves with parameters $(T_1, A_1, \rho_1) = (15, 10^{-3}, 0.2)$, $(T_2, A_2, \rho_2) = (10, 5 \cdot 10^{-4}, 0.5)$, $(T_3, A_3, \rho_3) = (5, 10^{-3}, 0.3)$, $(T_4, A_4, \rho_4) = (10, 5 \cdot 10^{-4}, 0.5)$.

The second network, $\mathcal{N}'$, includes Agents 1′, 2′, 3′ and 4′ and has the same communication structure and data of $\mathcal{N}$. The update laws have the same parameters $k_{i'} = k_i$ and $\mu_{i'} = \mu_i$, $i \in \mathcal{N}$, except for the family $(h_{i'})_{i' \in \mathcal{N}'}$ which is given by $h_{i'}^t = (1 + t)^{-1}$ for all $i' \in \mathcal{N}'$. The signals $h_{i'}$ satisfy (19)–(20) and, thus, $(h_{i'})_{i' \in \mathcal{N}'}$ is sufficiently exciting. However, it fails to be uniformly exciting. The simulation shown in Fig. 4 compares the time behavior of the update laws x_i, $i \in \mathcal{N}$ and $x_{i'}$, $i' \in \mathcal{N}'$. As shown in the figure, each "step" of M* is followed by the estimates x_i with the same convergence rate. On the contrary, M$^{\star\prime} = $M* is followed by the estimates $x_{i'}$ with a convergence rate which degrades in time. This is due to the fact that the family $(h_i)_{i \in \mathcal{N}}$ is uniformly exciting, while the family $(h_{i'})_{i' \in \mathcal{N}'}$ is only sufficiently exciting. Thus, uniformity of convergence is not guaranteed for the estimates $x_{i'}$. Nevertheless, the zoomed part of the plot clearly shows that the estimates $x_{i'}$ reach M$^{\star\prime}$ with higher precision (by Claim 5 of Theorem 1, indeed, since $h_{i'}^t \to 0$ the convergence of the estimates $x_{i'}$ is asymptotic if M$^{\star\prime}$ remains constant), whereas the

estimates x_i exhibit a non-zero residual error. The above simulations underline the necessary compromise, already mentioned in different parts of the paper, and formally characterized by Claims 3 and 5 of Theorem 1, between *exact convergence* and *uniformity in time*, which characterizes the proposed methodology.

Finally, Fig. 5 shows a simulation of the Max-Consensus (3) in the same setting (cf. Fig. 4). Again, the Max-Consensus fails in tracking the time-varying M*. To see why this is the case, consider for instance the change of value of M$_1$ at $t = 500$. This determines an increment of M*, bringing the Max-Consensus algorithm in a situation in which (4) holds at $t_0 = 500$. Hence, as explained in Section 1.1, x^{500} falls outside the domain of attraction of the new M*, and thus convergence fails.

6. Concluding remarks

As detailed in the proof of the main result (Section 7) and shown in the numerical simulations, the proposed solution is characterized by a necessary compromise between convergence rate and asymptotic error, as both are determined in the worst case by the signals h_i. In particular, if $(h_i)_{i \in \mathcal{N}}$ is uniformly exciting, uniform convergence is guaranteed, but the estimates will have a non-zero steady-state error. We stress that this residual error can be reduced arbitrarily by reducing the maximum value of the signals h_i accordingly. But we also remark that, in general, this results in a reduction of the convergence rate. Larger values of the signals h_i are associated instead with faster convergence but lead to larger steady-state errors. Moreover, in the limit case in which $h_i^t \to 0$ for all $i \in \mathcal{N}$, asymptotic convergence is obtained whenever $(h_i)_{i \in \mathcal{N}}$ is sufficiently exciting. The convergence rate, however, is superlinear and not lower-bounded, and thus uniformity is lost.

Clearly, "smart" choices of the signals h_i are possible adapting their value at run time to increase them when fast convergence is needed and decrease them when, instead, we desire a low residual error. "Adaptive" design choices of this kind will be the subject of future research.

We prove all the proposed solution properties under the assumption that the communication structure and the parameters remain constant during the execution. Although uniform global asymptotic stability already guarantees a good behavior for "slowly varying" structures (also shown by the numerical simulations), additional work is needed to extend the analysis to handle time-varying networks with communication delays and noise. This extension, in turn, calls for a stochastic framework in which the aleatory nature of those phenomena is fully captured and is the subject of future research.
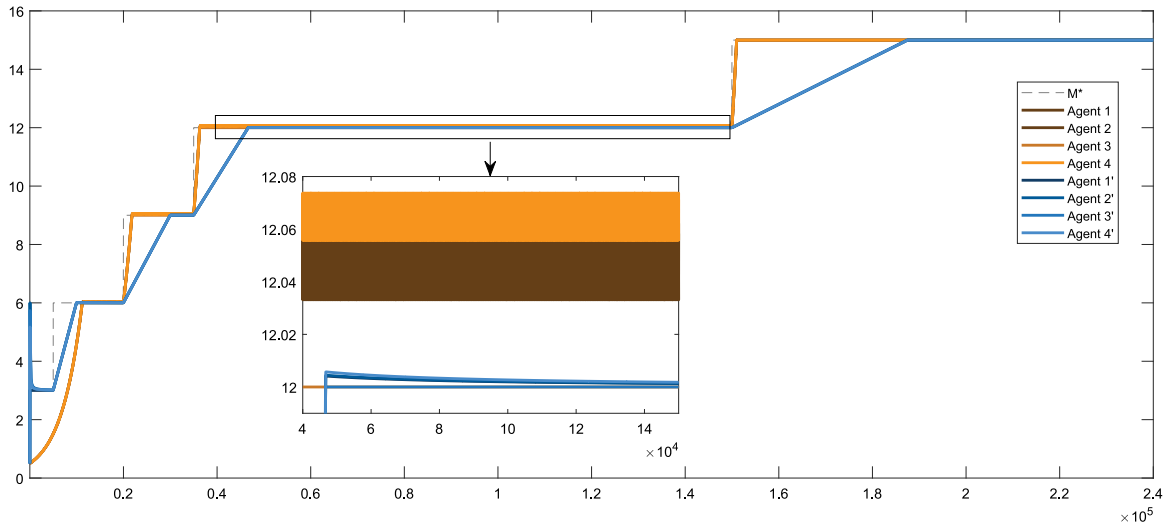
Fig. 4. Evolution of the estimates x_i in Scenario 2. The trajectory of the optimal value $M^\star$ is shown in dashed gray line. Dark to light orange lines depict the trajectory of the estimates x_i, $i = 1, \ldots, 4$ of the first network. Dark to light blue lines depict the trajectory of the estimates x_i, $i = 1', \ldots, 4'$ of the second network. In abscissa: iteration variable t. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

7. Proof of Theorem 1

7.1. Proof of Claim 1

In this subsection we prove Claim 1. In particular, we show that if the family $(h_i)_{i \in \mathcal{N}}$ is sufficiently exciting from some $t_0 \in \mathbb{N}$, then there exists $t^\star = t^\star(t_0) > t_0$ such that, for each $i \in \mathcal{N}$, $x_i^t \geq M^\star$ holds for all $t \geq t^\star$ and, for each $i \in I^\star$, $x_i^t = M^\star$ holds for all $t \geq t^\star$.

Define the function $\underline{i} : \mathbb{R}^N \to \mathcal{N}$, $x \mapsto \underline{i}(x) := \operatorname{argmin}_{i \in \mathcal{N}} x_i$. Then, $x_j \geq x_{\underline{i}(x)}$ holds for all $j \in \mathcal{N}$. Moreover, $h_i^t \geq 0$ and (14) imply $e^{h_i^t} - \operatorname{card}([i] \setminus i) k_i \geq 0$ for all $i \in \mathcal{N}$. Since $\Pi_{[\mu_i, M_i]}$ is increasing, we have

$$
\begin{aligned}
x_i^{t+1} &= \Pi_{[\mu_i, M_i]} \left[\left(e^{h_i^t} - \operatorname{card}([i] \setminus i) k_i \right) x_i^t + k_i \sum_{j \in [i] \setminus i} x_j^t \right] \\
&\geq \Pi_{[\mu_i, M_i]} \left[\left(e^{h_i^t} - \operatorname{card}([i] \setminus i) k_i \right) x_{\underline{i}(x^t)}^t \right. \\
&\qquad\qquad \left. + \operatorname{card}([i] \setminus i) k_i x_{\underline{i}(x^t)}^t \right] \\
&= \Pi_{[\mu_i, M_i]} \left[e^{h_i^t} x_{\underline{i}(x^t)}^t \right] \\
&= \max \left\{ \mu_i, \min \left\{ e^{h_i^t} x_{\underline{i}(x^t)}^t, M_i \right\} \right\} \\
&\geq \min \left\{ e^{h_i^t} x_{\underline{i}(x^t)}^t, M_i \right\} \geq \min \left\{ e^{h_i^t} x_{\underline{i}(x^t)}^t, M^\star \right\}
\end{aligned}
\tag{22}
$$

for all $t \geq t_0$ and all $i \in \mathcal{N}$.

First, notice that, if for some $\bar{t} \in \mathbb{N}$, $x_{\underline{i}(x^{\bar{t}})}^{\bar{t}} \geq M^\star$, then (22) implies $x_{\underline{i}(x^{\bar{t}+1})}^{\bar{t}+1} \geq M^\star$, so that by induction it is possible to conclude that $x_i^t \geq M^\star$ holds for all $t \geq \bar{t}$. Namely, the claim holds with $t^\star = \bar{t}$. It thus suffices to show that such $\bar{t}$ exists. In doing so, we proceed by contradiction. We first assume that

$$
x_{\underline{i}(x^t)}^t < M^\star, \qquad \forall t \geq t_0. \tag{23}
$$

Then, we show that, if the signals h_i are sufficiently exciting from t_0 (in the sense of Definition 5), then (23) leads to a contradiction, in this way proving the claim.
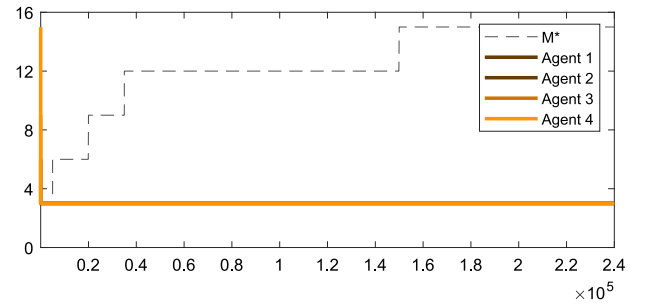
Fig. 5. Evolution of the Max-Consensus estimates (update law (3)) in the setting of Scenario 2 (cf. Fig. 4). In abscissa: iteration variable t.

Thus, assume that (23) holds. Then, since $h_i^t \geq 0$ for all $i \in \mathcal{N}$, (22) yields

$$
x_i^{t+s} \geq e^{h_i^t} x_{\underline{i}(x^t)}^t, \qquad \forall t \geq t_0, \ s \geq 1. \tag{24}
$$

Suppose that the signals h_i are sufficiently exciting from t_0, for some parameters $\underline{h}(t_0)$ and $\Delta(t_0)$. Then, for each $i \in \mathcal{N}$, there exists $s_i \in \{t_0 + 1, \ldots, t_0 + \Delta(t_0)\}$, such that $h_i^{s_i} \geq \underline{h}(t_0)$. In view of (24), this yields

$$
x_i^{t_0+1+\Delta(t_0)} \geq e^{\underline{h}(t_0)} x_{\underline{i}(x^{t_0+1})}^{t_0+1}, \qquad \forall i \in \mathcal{N},
$$

and thus, in particular,

$$
x_{\underline{i}\left(x^{t_0+1+\Delta(t_0)}\right)}^{t_0+1+\Delta(t_0)} \geq e^{\underline{h}(t_0)} x_{\underline{i}(x^{t_0+1})}^{t_0+1}.
$$

In the same way, in view of sufficiency of excitation of the signals h_i, for each $i \in \mathcal{N}$, there exists $s_i \in \{t_0 + 1 + \Delta(t_0), \ldots, t_0 + 2\Delta(t_0)\}$, such that $h_i^{s_i} \geq \underline{h}(t_0)$. Then, in view of (24), one has

$$
x_{\underline{i}\left(x^{t_0+1+2\Delta(t_0)}\right)}^{t_0+1+2\Delta(t_0)} \geq e^{\underline{h}(t_0)} x_{\underline{i}\left(x^{t_0+1+\Delta(t_0)}\right)}^{t_0+1+\Delta(t_0)} \geq e^{2\underline{h}(t_0)} x_{\underline{i}(x^{t_0+1})}^{t_0+1}.
$$

By repeating the same arguments, it is thus possible to conclude that, for each $m \in \mathbb{N}$ satisfying (15), one has

$$
x_{\underline{i}\left(x^{t_0+1+m\Delta(t_0)}\right)}^{t_0+1+m\Delta(t_0)} \geq e^{m\underline{h}(t_0)} x_{\underline{i}(x^{t_0+1})}^{t_0+1} \geq e^{m\underline{h}(t_0)} \underline{\mu}, \tag{25}
$$

in which we used the fact that, by definition of $\Pi_{[\underline{\mu}_i, M_i]}$, $x_i^t \geq \mu_i \geq \underline{\mu}$ for all $i \in \mathcal{N}$ and all $t \geq t_0 + 1$. Since the latter relation holds

in particular for

$$m^\star(t_0) = \frac{1}{\underline{h}(t_0)} \log\left(\frac{M^\star}{\underline{\mu}}\right).$$

Then, with $\bar{t} := t_0 + 1 + m^\star(t_0)\Delta(t_0)$, from (25) we obtain

$$x_i^{\bar{t}} \geq x_{\underline{i}(x^{\bar{t}})}^{\bar{t}} \geq e^{m^\star(t_0)\underline{h}(t_0)}\underline{\mu} = M^\star, \qquad \forall i \in \mathcal{N}$$

which contradicts (23) and, thus, proves that $x_i^t \geq M^\star$ holds for all $i \in \mathcal{N}$ and all $t \geq t^\star := \bar{t}$.

Finally, for all $i \in I^\star$, we have $x_i^t \in [\underline{\mu}, M_i] \leq M^\star$ for all $t \geq t_0 + 1$ and this, together with the bound $x_i^t \geq M^\star$ above, implies $x_i^t = M^\star$ for all $i \in I^\star$ and $t \geq t^\star$.

7.2. Proof of Claim 2

Since by Claim 1 each x_i satisfies $x_i^t \geq M^\star$ for all $t \geq t^\star$, then, in view of Assumption 2, each x_i also satisfies $x_i^t \geq \mu_i$ for all $t \geq t^\star$. This, in turn, allows us to write

$$x_i^{t+1} = \min\left\{M_i, \ e^{h_i^t}x_i^t + k_i \sum_{j\in[i]}(x_j^t - x_i^t)\right\}$$

for all $i \in \mathcal{N}$ and all $t \geq t^\star$, which implies both

$$x_i^t \leq M_i \tag{26}$$

and

$$x_i^{t+1} \leq e^{h_i^t}x_i^t + k_i \sum_{j\in[i]}(x_j^t - x_i^t) \tag{27}$$

for all $i \in \mathcal{N}$ and all $t \geq t^\star$. From (26) we also obtain

$$\limsup_{t\to\infty} |x_i^t| \leq M_i < \infty, \qquad \forall i \in \mathcal{N}. \tag{28}$$

In the following we rely on the forthcoming lemma, whose proof is postponed to Appendix B.

Lemma 3. *With $n \in \mathbb{N}$, let $x, y : \mathbb{N} \to \mathbb{R}^n$. Suppose that y is bounded and that, for some $t_0 \in \mathbb{N}$ and some $\lambda : \mathbb{N} \to \mathbb{R}_{\geq 0}$ fulfilling $\lambda^t \leq \nu \in [0, 1)$ for all $t \geq t_0$, x and y satisfy*

$$x^{t+1} \leq \lambda^t x^t + y^t \tag{29}$$

for all $t \geq t_0$. Then

$$\limsup_{t\to\infty} |x^t| \leq \frac{1}{1 - \limsup_{t\to\infty} \lambda^t} \limsup_{t\to\infty} |y^t|. \tag{30}$$

With $I^\star$ defined in (11), let $n^\star$ be the least integer such that $[I^\star]^{n^\star} = \mathcal{N}$ (which exists finite in view of Assumption 1). The case in which $n^\star = 0$ (i.e. $I^\star = \mathcal{N}$) directly follows from Claim 1. Hence, we consider $n^\star > 0$.

Assume that, for some $m \in \{0, \dots, n^\star - 1\}$, there exist $\alpha_m \in [0, 1)$ and $\beta_m > 0$ such that[8]

$$\max_{i\in[I^\star]_{m-1}^m} \limsup_{t\to\infty} |x_i^t| \leq \alpha_m \max_{j\in[I^\star]_m^{m+1}} \limsup_{t\to\infty} |x_j^t| + \beta_m M^\star. \tag{31}$$

We will now prove that, if this is the case, then a similar property holds also for $m + 1$.

First notice that, for each $i \in [I^\star]_m^{m+1}$, every $j \in [i]$ belongs to exactly one among the sets $[I^\star]_{m+1}^{m+2}$, $[I^\star]_m^{m+1}$, and $[I^\star]_{m-1}^m$. Hence,

---

[8] Here we let $[I^\star]^{-1} := \emptyset$.

in view of (27), we can write

$$x_i^{t+1} \leq \left(e^{h_i^t} - k_i \operatorname{card}([i] \setminus i)\right)x_i^t + k_i \sum_{j\in[i]\cap[I^\star]^m} x_j^t$$
$$+ k_i \sum_{j\in([i]\setminus i)\cap[I^\star]_m^{m+1}} x_j^t + k_i \sum_{j\in[i]\cap[I^\star]_{m+1}^{m+2}} x_j^t \tag{32}$$

for all $i \in [I^\star]_m^{m+1}$ and all $t \geq t^\star$, in which we used the fact that $[i] \cap [I^\star]_{m-1}^m = [i] \cap [I^\star]^m$, for all $i \in [I^\star]_m^{m+1}$. If (14) holds, then $1 + k_i \operatorname{card}([i] \setminus i) > 1$. With $\nu_1 > 0$ sufficiently small so that $\log(1 + k_i \operatorname{card}([i] \setminus i)) - 2\nu_1 > 0$, let

$$\bar{h}_{i,1} := \log(1 + k_i \operatorname{card}([i] \setminus i)) - 2\nu_1.$$

If $\limsup_{t\to\infty} h_i^t \leq \bar{h}_{i,1}$ for all $i \in \mathcal{N}$, then there exists $T^\star > t^\star$ such that

$$h_i^t \leq \bar{h}_{i,1} + \nu_1 = \log(1 + k_i \operatorname{card}([i] \setminus i)) - \nu_1 \tag{33}$$

for all $t \geq T^\star$ and all $i \in \mathcal{N}$. Thus, (14) and (33) imply

$$0 \leq e^{h_i^t} - k_i \operatorname{card}([i] \setminus i) \leq e^{\bar{h}_{i,1}+\nu_1} - k_i \operatorname{card}([i] \setminus i) < 1,$$

for all $t \geq T^\star$ and all $i \in \mathcal{N}$, so that (28), (32) and Lemma 3 imply

$$\limsup_{t\to\infty} |x_i^t| \leq \gamma_i \sum_{j\in[i]\cap[I^\star]^m} \limsup_{t\to\infty} |x_j^t|$$
$$+ \gamma_i \sum_{j\in([i]\setminus i)\cap[I^\star]_m^{m+1}} \limsup_{t\to\infty} |x_j^t| \tag{34}$$
$$+ \gamma_i \sum_{j\in[i]\cap[I^\star]_{m+1}^{m+2}} \limsup_{t\to\infty} |x_j^t|$$

for all $i \in [I^\star]_m^{m+1}$, in which we let

$$\gamma_i := \frac{k_i}{1 - \limsup_{t\to\infty}\left(e^{h_i^t} - k_i \operatorname{card}([i] \setminus i)\right)} \tag{35}$$

which exists finite in view of Lemma 3. In view of (31), Eq. (34) implies

$$\limsup_{t\to\infty} |x_i^t| \leq \left(c_{i,1}\alpha_m + c_{i,2}\right) \max_{j\in[I^\star]_m^{m+1}} \limsup_{t\to\infty} |x_j^t|$$
$$+ c_{i,3} \max_{j\in[I^\star]_{m+1}^{m+2}} \limsup_{t\to\infty} |x_j^t| + c_{i,1}\beta_m M^\star. \tag{36}$$

for all $i \in [I^\star]_m^{m+1}$, in which we let for convenience

$$c_{i,1} := \gamma_i \operatorname{card}\left([i] \cap [I^\star]^m\right)$$
$$c_{i,2} := \gamma_i \operatorname{card}\left(([i] \setminus i) \cap [I^\star]_m^{m+1}\right) \tag{37}$$
$$c_{i,3} := \gamma_i \operatorname{card}\left([i] \cap [I^\star]_{m+1}^{m+2}\right).$$

With $\nu_2 > 0$ sufficiently small so that $k_i(1 - \alpha_m) - \nu_2 > 0$ for all $i \in \mathcal{N}$ (recall that $\alpha_m < 1$ by assumption), define

$$\bar{h}_i := \min\left\{\bar{h}_{i,1}, \ \log(1 + k_i(1 - \alpha_m) - \nu_2)\right\}.$$

If

$$\limsup_{t\to\infty} h_i^t \leq \bar{h}_i \tag{38}$$

for all $i \in [I^\star]_m^{m+1}$, then, since $\operatorname{card}([i] \cap [I^\star]^m) \geq 1$, it holds that

$$1 - e^{\limsup_{t\to\infty} h_i^t} \geq 1 - e^{\bar{h}_i} \geq -k_i(1 - \alpha_m) + \nu_2$$
$$\geq -k_i(1 - \alpha_m) \operatorname{card}([i] \cap [I^\star]^m) + \nu_2 \tag{39}$$

for all $i \in [I^\star]_m^{m+1}$. Since for all $i \in [I^\star]_m^{m+1}$,

$$\text{card}\big(([i] \setminus i) \cap [I^\star]_m^{m+1}\big)$$

$$= \text{card}\,([i] \setminus i) - \text{card}\left([i] \cap [I^\star]^m\right) - \text{card}\left([i] \cap [I^\star]_{m+1}^{m+2}\right)$$

$$\leq \text{card}\,([i] \setminus i) - \text{card}\left([i] \cap [I^\star]^m\right),$$

then, we conclude that

$$
\begin{aligned}
c_{i,1}&\alpha_m + c_{i,2}\\
&\leq \frac{k_i(\alpha_m - 1)\,\text{card}\left([i] \cap [I^\star]^m\right) + k_i\,\text{card}\,([i] \setminus i)}{1 - e^{\bar{h}_i} + k_i\,\text{card}([i] \setminus i)}\\
&\leq \frac{(\alpha_m - 1)k_i\,\text{card}\left([i] \cap [I^\star]^m\right) + k_i\,\text{card}([i] \setminus i)}{(\alpha_m - 1)k_i\,\text{card}\left([i] \cap [I^\star]^m\right) + k_i\,\text{card}([i] \setminus i) + \nu_2}\\
&< 1.
\end{aligned}
\tag{40}
$$

for all $i \in [I^\star]_m^{m+1}$.

Now, since (36) holds for each $i \in [I^\star]_m^{m+1}$, it in particular holds for $\bar{i}$ satisfying

$$\bar{i} \in \operatorname*{argmax}_{i \in [I^\star]_m^{m+1}} \limsup_{t \to \infty} |x_i^t|, \tag{41}$$

so that (36) implies

$$
\begin{aligned}
\max_{i \in [I^\star]_m^{m+1}} \limsup_{t \to \infty} |x_i^t| &\leq (c_{\bar{i},1}\alpha_m + c_{\bar{i},2}) \max_{i \in [I^\star]_m^{m+1}} \limsup_{t \to \infty} |x_i^t|\\
&\quad + c_{\bar{i},3} \max_{j \in [I^\star]_{m+1}^{m+2}} \limsup_{t \to \infty} |x_j^t| + c_{\bar{i},1}\beta_m M^\star
\end{aligned}
$$

which, in view of (40), yields

$$
\begin{aligned}
\max_{i \in [I^\star]_m^{m+1}} \limsup_{t \to \infty} |x_i^t| &\leq \alpha_{m+1} \max_{j \in [I^\star]_{m+1}^{m+2}} \limsup_{t \to \infty} |x_j^t|\\
&\quad + \beta_{m+1}M^\star
\end{aligned}
\tag{42}
$$

with

$$
\begin{aligned}
\alpha_{m+1} &= \frac{c_{\bar{i},3}}{1 - (c_{\bar{i},1}\alpha_m + c_{\bar{i},2})},\\
\beta_{m+1} &= \frac{c_{\bar{i},1}}{1 - (c_{\bar{i},1}\alpha_m + c_{\bar{i},2})}\beta_m.
\end{aligned}
\tag{43}
$$

Furthermore, since $\limsup_{t \to \infty} h_i^t \leq \bar{h}_i$, in view of (39), α_{m+1} satisfies

$$
\begin{aligned}
\alpha_{m+1} &\leq \frac{k_{\bar{i}}\,\text{card}\left([\bar{i}] \cap [I^\star]_{m+1}^{m+2}\right)}{k_{\bar{i}}\,\text{card}([\bar{i}] \setminus \bar{i}) - k_{\bar{i}}\,\text{card}(([\bar{i}] \setminus \bar{i}) \cap [I^\star]^{m+1}) + \nu_2}\\
&\leq \frac{k_{\bar{i}}\,\text{card}\left([\bar{i}] \cap [I^\star]_{m+1}^{m+2}\right)}{k_{\bar{i}}\,\text{card}\left([\bar{i}] \cap [I^\star]_{m+1}^{m+2}\right) + \nu_2} < 1.
\end{aligned}
$$

Therefore, we claim that if (31) holds for some $m \in \{0, \dots, n^\star - 1\}$ with $\alpha_m < 1$ and $\beta_m \geq 0$, then (42) holds as well for $m + 1$ with $\alpha_{m+1} < 1$ and $\beta_{m+1} \geq 0$ given above. Since by Claim 1, Eq. (31) trivially holds for $m = 0$ with $\beta_0 = 1$ and $\alpha_0 = 0$, then we claim by induction that, if

$$\limsup_{t \to \infty} h_i^t \leq \bar{h} := \min_{i \in \mathcal{N}} \bar{h}_i, \qquad \forall i \in \mathcal{N}, \tag{44}$$

then Eq. (31) holds for each $m \in \{0, \dots, n^\star\}$.

Now, for $m = n^\star$, we have $[I^\star]^{m+1} \setminus [I^\star]^m = \emptyset$, so that (31) yields

$$\limsup_{t \to \infty} x_i^t \leq \beta_{n^\star} M^\star, \qquad \forall i \in [I^\star]_{n^\star - 1}^{n^\star}.$$

Thus, iterating (31) backwards and using (26) yield

$$\limsup_{t \to \infty} x_i^t \leq \min\left\{ M_i, \ (1 + \varepsilon_i)M^\star \right\} \tag{45}$$

in which

$$\varepsilon_i = 0, \qquad \forall i \in I^\star$$

and

$$\varepsilon_i = \sum_{\ell=0}^{n^\star - m} \left(\prod_{k=\ell+1}^{n^\star - m} \alpha_{n^\star - k} \right) \beta_{n^\star - \ell} - 1, \tag{46}$$

for all $i \in [I^\star]_{m-1}^m$ and all $m = 1, \dots, n^\star$. Moreover, (46) directly implies that the quantities ε_i also satisfy

$$\max_{i \in [I^\star]_{m-1}^m} \varepsilon_i = \alpha_m \left(1 + \max_{i \in [I^\star]_m^{m+1}} \varepsilon_i \right) + \beta_m - 1 \tag{47}$$

for all $m = 1, \dots, n^\star$. We now prove that ε_i in (45)–(46) can be reduced arbitrarily by reducing $\limsup_{t \to \infty} h_i^t$ accordingly for each $i \in \mathcal{N}$. For convenience, let

$$\upsilon_i := \limsup_{t \to \infty} h_i^t \in [0, \bar{h}_i]. \tag{48}$$

Then, the quantities γ_i, defined in (35), satisfy

$$\gamma_i(\upsilon_i) = \frac{k_i}{1 - e^{\upsilon_i} + k_i\,\text{card}([i] \setminus i)}.$$

Thus, γ_i is continuous in $[0, \infty)$, and

$$\lim_{\upsilon_i \to 0} \gamma_i(\upsilon_i) = \frac{1}{\text{card}([i] \setminus i)}.$$

In view of the definitions (37), also the quantities α_m and β_m, as defined in (43), depend on υ_i through γ_i, in which $\bar{i}$ satisfies (41). We now prove by induction that, by letting $\upsilon := (\upsilon_1, \dots, \upsilon_N)$, the following holds

$$\lim_{\upsilon \to 0} \alpha_m(\upsilon) + \beta_m(\upsilon) = 1, \qquad \forall m = 0, \dots, n^\star. \tag{49}$$

First notice that (49) trivially holds for $m = 0$, as indeed $\alpha_m = 0$ and $\beta_m = 1$ despite the value of υ. It thus suffices to show that if (49) holds for a given $m \in \{0, \dots, n^\star - 1\}$, then the same relation holds as well for $m + 1$. For, assume that (49) holds for a given $m \in \{0, \dots, n^\star - 1\}$. Then, we can write $\lim_{\upsilon \to 0} \beta_m(\upsilon) = 1 - \lim_{\upsilon \to 0} \alpha_m(\upsilon)$. Thus, by letting for convenience $\rho_1 := \text{card}([\bar{i}] \cap [I^\star]^m)$, $\rho_2 := \text{card}(([\bar{i}] \setminus \bar{i}) \cap ([I^\star]_m^{m+1}))$, $\rho_3 := \text{card}([\bar{i}] \cap ([I^\star]_{m+1}^{m+2}))$, and noting that $\text{card}([\bar{i}] \setminus \bar{i}) - \rho_2 = \rho_1 + \rho_3$, we obtain

$$
\begin{aligned}
&\lim_{\upsilon \to 0} \alpha_{m+1}(\upsilon) + \beta_{m+1}(\upsilon)\\
&\quad = \frac{\rho_3 + (1 - \lim_{\upsilon \to 0} \alpha_m(\upsilon))\,\rho_1}{\text{card}([\bar{i}] \setminus \bar{i}) - \lim_{\upsilon \to 0} \alpha_m(\upsilon)\rho_1 - \rho_2}\\
&\quad = \frac{\rho_3 + (1 - \lim_{\upsilon \to 0} \alpha_m(\upsilon))\,\rho_1}{\rho_3 + (1 - \lim_{\upsilon \to 0} \alpha_m(\upsilon))\,\rho_1} = 1.
\end{aligned}
$$

Thus, by induction, we claim (49) for all $m \in \{0, \dots, n^\star\}$.

Since for every $i \in [I^\star]_{n^\star - 1}^{n^\star}$, $c_{i,3} = 0$ (in fact $[I^\star]^{n^\star + 1} = \emptyset$), then $\alpha_{n^\star} = 0$. Thus,

$$\lim_{\upsilon \to 0} \beta_{n^\star}(\upsilon) = 1.$$

In view of (46), this implies

$$\lim_{\upsilon \to 0} \max_{i \in [I^\star]_{n^\star - 1}^{n^\star}} \varepsilon_i(\upsilon) = 0.$$

In view of (47), $\lim_{\upsilon \to 0} \max_{i \in [I^\star]_m^{m+1}} \varepsilon_i(\upsilon) = 0$ implies

$$\lim_{\upsilon \to 0} \max_{i \in [I^\star]_{m-1}^m} \varepsilon_i(\upsilon) = \lim_{\upsilon \to 0} (\alpha_m(\upsilon) + \beta_m(\upsilon)) - 1 = 0,$$

so that, by induction, we conclude that

$$\lim_{\upsilon \to 0} \max_{i \in [I^\star]_{m-1}^m} \varepsilon_i(\upsilon) = 0, \quad \forall m \in \{0, \dots, n^\star\},$$

i.e.

$$\lim_{\upsilon \to 0} \varepsilon_i(\upsilon) = 0, \quad \forall i \in \mathcal{N}. \tag{50}$$

The latter equation thus implies that, given any $\epsilon \geq 0$, there exists $\delta'(\epsilon) \geq 0$ such that $|\upsilon| \leq \delta'(\epsilon)$ implies $\mathrm{M}^\star \varepsilon_i \leq \epsilon$ for all $i \in \mathcal{N}$. Therefore, if

$$\limsup_{t \to \infty} h_i^t \leq \delta(\epsilon) := \min\left\{\bar{h}, \frac{\delta'(\epsilon)}{N}\right\}, \qquad \forall i \in \mathcal{N} \tag{51}$$

then $|\upsilon| \leq \delta'(\epsilon)$, which implies $\mathrm{M}^\star \varepsilon_i \leq \epsilon$. In turn, in view of (45), this implies

$$\limsup_{t \to \infty} x_i^t \leq \min\left\{\mathrm{M}_i, \ \mathrm{M}^\star + \epsilon\right\}. \tag{52}$$

Claim 2 thus follows from (52) and by noticing that Claim 1 implies $\limsup_{t \to \infty} x_i \geq \mathrm{M}^\star$.

7.3. Proof of Claim 3

The third claim of the theorem, i.e., that uniformity of excitation (in the sense of Definition 6) of $(h_i)_{i \in \mathcal{N}}$ implies uniform attractiveness of $\mathcal{A}_\epsilon := \prod_{i \in \mathcal{N}} [\mathrm{M}^\star, \min\{\mathrm{M}^\star + \epsilon, \mathrm{M}_i\}]$, directly follows by the fact that, if the family $(h_i)_{i \in \mathcal{N}}$ is uniformly exciting, then in the above analysis $t^\star$ does not depend on t_0 and, therefore, the convergence (52) is uniform in the initial time.

7.4. Proof of Claim 4

In this subsection we prove the fourth claim of the theorem. With $(\tau_i)_{i \in \mathcal{N}} \in \mathbb{N}^N$ arbitrary, let $F_i \in \mathbb{R}^{\tau_i \times \tau_i}$ and $C_i \in \mathbb{R}^{1 \times \tau_i}$ denote the matrices

$$F_i := \begin{bmatrix} 0_{(\tau_i - 1) \times 1} & I_{(\tau_i - 1) \times (\tau_i - 1)} \\ 1 & 0_{1 \times (\tau_i - 1)} \end{bmatrix}, \qquad C_i := \begin{bmatrix} 1 & 0_{1 \times (\tau_i - 1)} \end{bmatrix}.$$

Then, each τ_i-periodic signal h_i satisfies

$$\xi_i^{t+1} = F_i \xi_i^t, \quad h_i^t = C_i \xi_i^t \tag{53}$$

for a suitable initial condition $\xi_i^{t_0} \in \mathbb{R}^{\tau_i}$. Moreover, if all the signals h_i are non-zero, then, by Lemma 1, $(h_i)_{i \in \mathcal{N}}$ is uniformly exciting in the sense of Definition 6 for some $\underline{h} > 0$. For a fixed $\epsilon > 0$, let $\delta(\epsilon)$ be defined as above in (51), and let

$$\Xi_i := \Big\{ \xi_i \in \mathbb{R}^{\tau_i} \mid \forall j \in \{1, \ldots, \tau_i\}, \ \xi_{i,j} \in [0, \delta(\epsilon)], \ \text{and}$$

$$\exists j \in \{1, \ldots, \tau_i\}, \ \xi_{i,j} \geq \underline{h} \Big\},$$

where $\xi_{i,j}$ denotes the jth component of ξ_i. Then, Ξ_i is compact and invariant for (53). We now consider the interconnection between (53) and the update laws (13) for all $i \in \mathcal{N}$, with the dynamics restricted to the invariant set $Z := \Xi \times \mathbb{R}^N$, being $\Xi := \prod_{i \in \mathcal{N}} \Xi_i$. We compactly rewrite this interconnections as follows

$$z^{t+1} = \phi(z^t), \qquad z^t \in Z \tag{54}$$

with ϕ suitably defined and $z^t := (\xi^t, x^t) \in \mathbb{R}^r \times \mathbb{R}^N$, being $\xi := (\xi_i)_{i \in \mathcal{N}}$ and $r := \sum_{i \in \mathcal{N}} \tau_i$. Clearly, for every solution x_a to (13) starting at a given $t_0 \in \mathbb{N}$ and subject to the signals $(h_i)_{i \in \mathcal{N}}$, there is a solution $z_b = (\xi_b, x_b)$ to (54) starting at 0 and such that $x_b(t) = x_a(t_0 + t)$ for all $t \in \mathbb{N}$. For each compact $K \subset \Xi \times \mathbb{R}^N$, let $\mathcal{S}(K)$ denote the set of solutions to (54) starting at 0 from K and, for each $t \in \mathbb{N}$, define the reachable set from K as $\mathcal{R}^t(K) := \left\{ (\xi^s, x^s) \in \Xi \times \mathbb{R}^N \mid (\xi, x) \in \mathcal{S}(K), \ s \geq t \right\}$. In view of the above analysis, and since Ξ is invariant for (54), it follows that $\mathcal{R}_t(K)$ is included in $\Xi \times \mathbb{R}^N$ and bounded uniformly in K and t for each $t \geq 1$. Thus, the limit set $\Omega(K) := \bigcap_{t \in \mathbb{N}} \overline{\mathcal{R}^t(K)}$ (where $\overline{\mathcal{R}^t(K)}$ denotes the closure of $\mathcal{R}^t(K)$) is compact, non-empty, and included in

$\Xi \times \mathbb{R}^N$. Moreover, since ϕ is continuous by construction, then $\Omega(K)$ is also forward invariant, uniformly globally attractive for (54) from K (see e.g. Goebel et al. (2012, Proposition 6.26)), and it is the smallest set having the above properties. Furthermore, we notice that, by definition of the update laws (13), $x_i^t \in [\mu_i, \mathrm{M}_i]$ for all $t \geq t_0$ despite the value of the initial conditions and of t_0, so that we conclude that $\Omega(K_1) = \Omega(K_2)$ for all K_1, K_2 supersets of $K^\star := \prod_{i \in \mathcal{N}} [\mu_i, \mathrm{M}_i]$. In the following we let $\Omega := \Omega(K^\star)$.

As $(h_i)_{i \in \mathcal{N}}$ is uniformly exciting, by Claim 3 the convergence (52) holds uniformly in the initial time. By the properties of Ω, this implies that $\Omega \subset \Xi \times \mathcal{A}_\epsilon$, and the projection $\mathcal{A}_\epsilon^u := \left\{ x \in \mathbb{R}^N \mid (\xi, x) \in \Omega \right\}$ satisfies $\mathcal{A}_\epsilon^u \subset \mathcal{A}_\epsilon$. Therefore, it remains to show that $\mathcal{A}_\epsilon^u$ is stable for x, i.e. that for each $\ell > 0$, there exists $b(\ell) > 0$, such that every solution to (54) satisfying $\mathrm{dist}\left(x^0, \ \mathcal{A}_\epsilon^u\right) \leq b(\ell)$ also satisfies $\mathrm{dist}\left(x^t, \ \mathcal{A}_\epsilon^u\right) \leq \ell$ for all $t \in \mathbb{N}$. This, in turn, can be proved by similar arguments of Goebel et al. (2012, Proposition 7.5)). In particular, suppose that the above stability property does not hold, and fix an $\ell > 0$ arbitrarily. If $\mathcal{A}_\epsilon^u$ is not stable, then for each $m \in \mathbb{N}$ there exist $\tau_m \in \mathbb{N}$ and a solution $z_m = (\xi_m, x_m) \in \mathcal{S}(Z)$ such that $\mathrm{dist}\left(x_m^0, \ \mathcal{A}_\epsilon^u\right) \leq 2^{-m}$ and $\mathrm{dist}\left(x_m^{\tau_m}, \ \mathcal{A}_\epsilon^u\right) > \ell$. This, in turn implies

$$\mathrm{dist}\left(z_m^{\tau_m}, \ \Omega\right) > \ell. \tag{55}$$

Since $X_0 := \{x \in \mathbb{R}^N \mid \mathrm{dist}\left(x, \ \mathcal{A}_\epsilon^u\right) \leq 1\}$ is compact, $Z_0 := \Xi \times X_0$ is compact. Thus, since $z_m^0 \in Z_0$ for all $m \in \mathbb{N}$, by uniform attractiveness of Ω, there exists $\bar{\tau} = \bar{\tau}(\ell) \in \mathbb{N}$ such that $\tau_m \leq \bar{\tau}$ for all $m \in \mathbb{N}$. We are thus given a sequence $(z_m|_{\leq \bar{\tau}})_{m \in \mathbb{N}}$ of uniformly bounded signals $z_m|_{\leq \bar{\tau}}$, obtained by restricting the solutions z_m to $\{0, \ldots, \bar{\tau}\}$, which satisfies $\lim_{m \to \infty} \mathrm{dist}\left(z_m^0, \ \Omega\right) = 0$. As ϕ is continuous, Z is closed, and since Ω is forward invariant, then in view of Goebel et al. (2012, Theorem 6.8) we can extract a subsequence of $(z_m|_{\leq \bar{\tau}})_{m \in \mathbb{N}}$ (which we do not re-index) that satisfies $\lim_{m \to \infty} \mathrm{dist}\left(z_m^t, \ \Omega\right) = 0$ for all $t \in \{0, \ldots, \bar{\tau}\}$. This, however, contradicts (55) and proves the claim.

7.5. Proof of Claim 5

The last claim of the theorem, i.e. that if $(h_i)_{i \in \mathcal{N}}$ is sufficiently exciting according to Definition 5 and $\lim_{t \to \infty} h_i^t = 0$, then $\lim_{t \to \infty} x_i^t = \mathrm{M}^\star$ for all $i \in \mathcal{N}$, follows directly from (48)–(50). ∎

Appendix A. Proof of Lemma 1

For each $i \in \mathcal{N}$, let $T_i \in \mathbb{N}_{\geq 1}$ be the period of h_i and, with $t_i^\star$ and $h_i^\star > 0$ such that $h_i^{t_i^\star} \geq h_i^\star$, let

$$r_i := t_i^\star - T_i \max\{n \in \mathbb{N} \mid T_i n \leq t_i^\star\}.$$

Then $r_i \in \{0, \ldots, T_i\}$ and, since h_i is T_i-periodic, for every $i \in \mathcal{N}$ we have

$$h_i^{r_i + n T_i} \geq h_i^\star \qquad \forall n \in \mathbb{N}. \tag{A.1}$$

Let $\Delta := \max_{i \in \mathcal{N}} T_i + 1$ and $\underline{h} := \min_{i \in \mathcal{N}} h_i^\star$. Fix arbitrarily $m \in \mathbb{N}_{\geq 1}$ and $t_0 \in \mathbb{N}$. Then we claim that, for each $i \in \mathcal{N}$, there exists $n_i \in \mathbb{N}$ such that

$$s_i := r_i + n_i T_i \in \Big\{ t_0 + 1 + (m-1)\Delta, \ \ldots, \ t_0 + m\Delta \Big\}.$$

In fact, if this is not true, there exist $m, t_0, n \in \mathbb{N}$ and $i \in \mathcal{N}$ such that $r_i + n T_i < t_0 + 1 + (m-1)\Delta$ and $r_i + (n+1)T_i > t_0 + m\Delta$ hold. This, however, implies

$$\Delta = (1 - m)\Delta + m\Delta < (1-m)\Delta + r_i + (n+1)T_i - t_0$$
$$< T_i + 1,$$

which contradicts the fact that, by definition, $\Delta \geq T_i + 1$ for all $i \in \mathcal{N}$. Since (A.1) implies that $h_i^{s_i} \geq \underline{h}$ for all $i \in \mathcal{N}$, then we claim that, for every $t_0 \in \mathbb{N}$, $m \in \mathbb{N}_{\geq 1}$ (and thus, in particular, for those satisfying $m \leq \log(M^\star/\mu)/\underline{h}$) and $i \in \mathcal{N}$, there exists $s_i \in \{t_0 + 1 + (m-1)\Delta, \ldots, t_0 + m\Delta\}$ such that $h_i^{s_i} \geq \underline{h}$, which proves the claim. ∎

Appendix B. Proof of Lemma 3

As $\nu \in [0, 1)$, then for each $\epsilon \in (0, 1 - \nu)$ there exists $t_1^\star \geq t_0$ such that

$$\nu^{t-t_0}x^{t_0} \leq \epsilon, \qquad |y^t| \leq \limsup_{t\to\infty}|y^t| + \epsilon, \qquad \lambda^t \leq \limsup_{t\to\infty}\lambda^t + \epsilon$$

for all $t \geq t_1^\star$. As $\lambda^t \leq \nu < 1$ for all $t \geq t_0$, by iterating (29), for $t > t_1^\star$, we obtain

$$
\begin{aligned}
|x^t| &\leq \left(\prod_{s=t_0}^{t-1}\lambda^s\right)|x^{t_0}| + \sum_{s=t_0}^{t-1}\left(\prod_{\ell=s+1}^{t-1}\lambda^\ell\right)|y^s| \\
&\leq \nu^{t-t_0}|x^{t_0}| + \sum_{s=t_0}^{t_1^\star-1}\left(\prod_{\ell=s+1}^{t-1}\lambda^\ell\right)|y^s| + \sum_{s=t_1^\star}^{t-1}\left(\prod_{\ell=s+1}^{t-1}\lambda^\ell\right)|y^s| \qquad \text{(B.1)} \\
&\leq \epsilon + \sum_{s=t_0}^{t_1^\star-1}\nu^{t-s-1}|y^s| + \sum_{s=t_1^\star}^{t-1}\left(\prod_{\ell=s+1}^{t-1}\lambda^\ell\right)|y^s|.
\end{aligned}
$$

As y is bounded, there exists c such that $|y^t| \leq c$ for all $t \in \mathbb{N}$. Hence, the second term of the sum satisfies

$$\sum_{s=t_0}^{t_1^\star-1}\nu^{t-s-1}y^s = \nu^{t-t_1^\star}\sum_{s=t_0}^{t_1^\star-1}\nu^{t_1^\star-s-1}y^s \leq \nu^{t-t_1^\star}\frac{c}{1-\nu}.$$

Therefore, there exists $t_2^\star \geq t_1^\star$ such that

$$\sum_{s=t_0}^{t_1^\star-1}\nu^{t-s-1}y^s \leq \epsilon, \qquad \forall t \geq t_2^\star.$$

Denote for convenience $\bar{y} := \limsup_{t\to\infty}|y^t|$ and $\bar{\lambda} := \limsup_{t\to\infty}\lambda^t$. As $\lambda^t \leq \nu$ for all $t \geq t_0$, then $\bar{\lambda} \leq \nu$. As $\epsilon < 1 - \nu$ by assumptions, then $\bar{\lambda} + \epsilon < 1$. Therefore, since $t \geq t_1^\star$, then the last term of (B.1) satisfies

$$
\begin{aligned}
\sum_{s=t_1^\star}^{t-1}\left(\prod_{\ell=s+1}^{t-1}\lambda^\ell\right)|y^s| &\leq \sum_{s=t_1^\star}^{t-1}(\bar{\lambda}+\epsilon)^{t-s-1}(\bar{y}+\epsilon) \\
&\leq \frac{\epsilon}{1-\nu} + \frac{\bar{y}}{1-(\bar{\lambda}+\epsilon)} \qquad \text{(B.2)} \\
&\leq \frac{\bar{y}}{1-\bar{\lambda}} + \frac{\epsilon}{1-\nu} + \frac{\bar{y}}{1-(\bar{\lambda}+\epsilon)} - \frac{\bar{y}}{1-\bar{\lambda}} \\
&\leq \frac{\bar{y}}{1-\bar{\lambda}} + \bar{y}p(\epsilon)
\end{aligned}
$$

in which $p : [0, 1-\nu) \to \mathbb{R}$, defined as

$$p(\epsilon) := \frac{\epsilon}{1-\nu} + \frac{\epsilon}{(1-\bar{\lambda})(1-\bar{\lambda}-\epsilon)},$$

is continuous and satisfies $\lim_{\epsilon\to 0} p(\epsilon) = 0$. From (B.1) we get $|x^t| \leq \bar{y}/(1-\bar{\lambda}) + 2\epsilon + \bar{y}p(\epsilon)$ for all $t \geq t_2^\star$, and the claim follows by arbitrariness of ϵ. ∎

References

Ariyur, K. B., & Krstić, M. (2003). *Real-time optimization by extremum-seeking control.* Wiley-Blackwell.

Bastianello, N., Carli, R., Schenato, L., & Todescato, M. (2020). Asynchronous distributed optimization over lossy networks via relaxed ADMM: Stability and linear convergence. *IEEE Transactions on Automatic Control*, http://dx.doi.org/10.1109/TAC.2020.3011358.

Bin, M., Notarnicola, I, Marconi, L., & Notarstefano, G. (2019). A system theoretical perspective to gradient-tracking algorithms for distributed quadratic optimization. In *Proc. IEEE 58th Conf. decision and control* (pp. 2994–2999).

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning, 3*, 1–122.

Bullo, F., Cortes, J., & Martinez, S. (2009). *Distributed control of robotic networks: A mathematical approach to motion coordination algorithms.* Princeton University Press.

Chang, T.-H. (2016). A proximal dual consensus ADMM method for multi-agent constrained optimization. *IEEE Transactions on Signal Processing, 64*(14), 3719–3734.

Chang, T.-H., Nedić, A., & Scaglione, A. (2014). Distributed constrained optimization by consensus-based primal-dual Perturbation method. *IEEE Transactions on Automatic Control, 59*(6), 1524–1538.

Gallager, R. G., Humblet, P. A., & Spira, P. M. (1983). A Distributed Algorithm for minimum-weight spanning trees. *ACM Transactions on Programming Languages and Systems, 5*, 66–77.

Goebel, R., Sanfelice, R. G., & Teel, A. R. (2012). *Hybrid dynamical systems. modeling, stability, and robustness.* Princeton University Press.

Golfar, A., & Ghaisari, J. (2019). Convergence analysis of max-consensus algorithm in probabilistic communication networks with Bernoulli dropouts. *International Journal of Systems Science, 50*(7), 1313–1326.

Iutzeler, F., Ciblat, P., & Jakubowicz, J. (2012). Analysis of Max-Consensus Algorithms in wireless channels. *IEEE Transactions on Signal Processing, 60*(11), 6103–6107.

Jakovetić, D., Moura, J. M. F., & Xavier, J. (2015). Linear convergence rate of a class of distributed augmented Lagrangian algorithms. *IEEE Transactions on Automatic Control, 60*(4), 922–936.

Kokotović, P., Khalil, H. K., & O'Reilly, J. (1999). *Singular Perturbation methods in control. Analysis and design.* SIAM.

Kushner, H. J., & Yin, G. G. (1997). *Stochastic approximation algorithms and applications.* Springer Verlag.

Ling, Q., Shi, W., Wu, G., & Ribeiro, A. (2015). DLM: Decentralized linearized alternating direction method of multipliers. *IEEE Transactions on Signal Processing, 63*(15), 4051–4064.

Lobel, I., & Ozdaglar, A. (2011). Distributed subgradient methods for convex optimization over random networks. *IEEE Transactions on Automatic Control, 56*(6), 1291–1306.

Makhdoumi, A., & Ozdaglar, A. (2017). Convergence rate of distributed ADMM over networks. *IEEE Transactions on Automatic Control, 62*(10), 5082–5095.

Mokhtari, A., Ling, Q., & Ribeiro, A. (2017). Network Newton distributed optimization methods. *IEEE Transactions on Signal Processing, 65*(1), 146–161.

Mota, J. F. C., Xavier, J. M. F., Aguiar, P. M. Q., & Püschel, M. (2013). D-ADMM: A communication-Efficient Distributed Algorithm for separable optimization. *IEEE Transactions on Signal Processing, 61*(10), 2718–2723.

Nedić, A., Olshevsky, A., & Shi, W. (2017). Achieving geometric convergence for distributed optimization over time-varying graphs. *SIAM Journal on Optimization, 27*(4), 2597–2633.

Nedić, A., Olshevsky, A., Shi, W., & Uribe, C. A. (2017). Geometrically convergent distributed optimization with uncoordinated step-sizes. In *Proc. 2017 American control conference* (pp. 3950–3955).

Nedic, A., & Ozdaglar, A. (2009). Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control, 54*(1), 48–61.

Nedic, A., Ozdaglar, A., & Parrilo, P. A. (2010). Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, [ISSN: 1558-2523] *55*(4), 922–938.

Nejad, B. M., Attia, S. A., & Raisch, J. (2009). Max-consensus in a max-plus algebraic setting: The case of fixed communication topologies. In *Proc. 2009 XXII Int. symposium on information, communication and automation technologies* (pp. 1–7).

Notarstefano, G., Notarnicola, I., & Camisa, A. (2019). Distributed optimization for smart cyber-physical networks. *Foundations and Trends® in Systems and Control, 7*, 253–383.

Qu, G., & Li, N. (2018). Harnessing smoothness to accelerate distributed optimization. *IEEE Transactions on Control of Network Systems, 5*(3), 1245–1260.

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics*, *22*(3), 400–407.

Santoro, N. (2006). *Design and analysis of distributed algorithms*. John Wiley & Sons.

Shi, W., Ling, Q., Wu, G., & Yin, W. (2015a). EXTRA: An exact first-order algorithm for decentralized consensus optimization. *SIAM Journal on Optimization*, *25*(2), 944–966.

Shi, W., Ling, Q., Wu, G., & Yin, W. (2015b). A Proximal Gradient Algorithm for decentralized composite optimization. *IEEE Transactions on Signal Processing*, *63*(22), 6013–6023.

Shi, W., Ling, Q., Yuan, K., Wu, G., & Yin, W. (2014). On the linear convergence of the ADMM in decentralized consensus optimization. *IEEE Transactions on Signal Processing*, *62*(7), 1750–1761.

Simonetto, A., & Jamali-Rad, H. (2016). Primal recovery from consensus-based dual decomposition for distributed convex optimization. *Journal of Optimization Theory and Applications*, *168*, 172–197.

Tan, Y., Nešić, D., & Mareels, I. (2006). On non-local stability properties of extremum seeking control. *Automatica*, *6*, 889–903.

Teel, A. R., Moreau, L., & Nešić, D. (2003). A unified framework for input-to-state stability in systems with two time scales. *IEEE Transactions on Automatic Control*, *48*, 1526–1544.

Varagnolo, D., Zanella, F., Cenedese, A., Pillonetto, G., & Schenato, L. (2016). Newton-raphson consensus for distributed convex optimization. *IEEE Transactions on Automatic Control*, *61*(4), 994–1009.

Wang, W., Teel, A. R., & Nešić, D. (2012). Analysis for a class of singularly perturbed hybrid systems via averaging. *Automatica*, *48*, 1057–1068.

Xi, C., Xin, R., & Khan, U. A. (2018). ADD-OPT: Accelerated distributed directed optimization. *IEEE Transactions on Automatic Control*, *63*(5), 1329–1339.

Xie, P., You, K., Tempo, R., Song, S., & Wu, C. (2018). Distributed convex optimization with inequality constraints over time-varying unbalanced digraphs. *IEEE Transactions on Automatic Control*, *63*(12), 4331–4337.

Xu, J., Zhu, S., Soh, Y. C., & Xie, L. (2015). Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes. In *Proc. 54th IEEE Conf. decision and control* (pp. 2055–2060).

Yuan, K., Ling, Q., & Yin, W. (2016). On the convergence of decentralized gradient descent. *SIAM Journal on Optimization*, *26*(3), 1835–1854.

Zhu, M., & Martinez, S. (2012). On distributed convex optimization under inequality and equality constraints. *IEEE Transactions on Automatic Control*, *57*(1), 151–164.

Michelangelo Bin received the Master's degree in Automation Engineering and the Ph.D. in Control Theory from the University of Bologna, Italy, in 2015 and 2019 respectively. From September 2019 he is with the Department of Electrical and Electronic Engineering, Imperial College London, UK, where he is currently a Research Associate. He is the recipient of the SIDRA 2019 Best Italian Ph.D. Thesis Award in Automatica. He serves as Associate Editor of Systems & Control Letters. His research interests include systems theory, nonlinear control and regulation, adaptive systems, and large-scale network science and control.

Thomas Parisini received the Ph.D. degree in Electronic Engineering and Computer Science in 1993 from the University of Genoa. He was with Politecnico di Milano and since 2010 he holds the Chair of Industrial Control and is Director of Research at Imperial College London. He is a Deputy Director of the KIOS Research and Innovation Centre of Excellence, University of Cyprus. Since 2001 he is also Danieli Endowed Chair of Automation Control with University of Trieste. In 2009–2012 he was Deputy Rector of University of Trieste. In 2018 he received an *Honorary Doctorate* from University of Aalborg, Denmark. In 2020 he has been appointed as Deputy Chair of the Employment & Education Task Force of the B20–Italy. He authored or co-authored more than 350 research papers in archival journals, book chapters, and international conference proceedings. He is a co-recipient of the IFAC Best Application Paper Prize of the Journal of Process Control, Elsevier, for the three-year period 2011–2013 and of the 2004 Outstanding Paper Award of the IEEE Trans. on Neural Networks. He is also a recipient of the 2007 IEEE Distinguished Member Award. In 2016, he was awarded as Principal Investigator at Imperial of the H2020 European Union flagship Teaming Project KIOS Research and Innovation Centre of Excellence led by University of Cyprus. Thomas Parisini serves as 2021–2022 President of the IEEE Control Systems Society and has served as Vice-President for Publications Activities. During 2009–2016 he was the Editor-in-Chief of the IEEE Trans. on Control Systems Technology. Since 2017, he is Editor for Control Applications of Automatica and since 2018 he is the Editor in Chief of the European Journal of Control. Among other activities, he was the Program Chair of the 2008 IEEE Conference on Decision and Control and General Co-Chair of the 2013 IEEE Conference on Decision and Control. Prof. Parisini is a Fellow of the IEEE and of the IFAC.