# A Distributed Methodology for Approximate Uniform Global Minimum Sharing


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

[a] *Imperial College London, UK*
[b] *KIOS Research and Innovation Center of Excellence, University of Cyprus, Cyprus*
[c] *University of Trieste, Italy*





**Abstract**

The paper deals with the *distributed minimum sharing* problem, in which a network of decision-makers – exchanging information through a communication network – computes the minimum of some local quantities of interest in a distributed and decentralized way. The problem is equivalently cast into a cost-coupled distributed optimization problem, and an adjustable approximate (or sub-optimal) solution is presented which enjoys several properties of crucial importance in applications. In particular, the proposed solution is *scalable* in that the dimension of the state space 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 condition) asymptotic stability result is provided, as well as an attractiveness property towards a steady state which can be made arbitrarily close to the sought minimum. Exact asymptotic convergence is also recovered at the price, however, of loosing uniformity of the convergence 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 $\mathrm{M}_i \in \mathbb{R}$ 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

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

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

$$\begin{aligned} &\max \sum_{i\in\mathcal{N}} \psi_i(x_i) \\ &x_i \le \mathrm{M}_i, \qquad \forall i \in \mathcal{N} \\ &x_i = x_j, \qquad \forall i,j \in \mathcal{N} \end{aligned} \tag{2}$$

obtained with $\psi_i$, $i \in \mathcal{N}$, 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 [1].

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., [2–6] and the references therein). Perhaps the most elementary existing algorithms solving the minimum sharing problem are the *FloodMax* [2] and the *Max-Consensus* [4–6]. In its simplest form, Max-Consensus[1] requires each agent $i \in \mathcal{N}$ to store an estimate $x_i \in \mathbb{R}$ of $\mathrm{M}^\star$ which is updated iteratively on the basis of the following update rule

$$x_i^{t+1} = \min_{j\in[i]} x_j^t, \qquad x_i^{t_0} = \mathrm{M}_i, \qquad \forall i \in \mathcal{N}, \tag{3}$$

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 (3) is decentralized and scalable, in that each agent needs only information coming from its neighbors. However, although (3) guarantees convergence of each $x_i$ to $\mathrm{M}^\star$ when the estimates $x_i$ are initialized as specified, *convergence is not guaranteed for an arbitrary initialization.* Namely, (3) is not *globally convergent.* In fact, under the action of the same update law, if

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

then $x_i^t < \mathrm{M}^\star$ holds true for all subsequent $t$, so that $x_i^t \to \mathrm{M}^\star$ cannot hold[2]. While there are application domains for which attaining global convergence is not strictly

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

[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 (3). This, in turn, is intimately linked to the unfeasibility result of [3, Theorem 3.1.1], and to the *detectability* issues appearing in many control problems, such as *Extremum Seeking* [7, 8].

necessary, there are many others in which it is a crucial requirement. This is the case, for instance, when the quantities $\mathrm{M}_i$ 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 $\mathrm{M}^\star$, i.e. $x_i^{t_1} = \mathrm{M}^\star$ for all $i \in \mathcal{N}$, and assume that there is a unique $k \in \mathcal{N}$ such that $\mathrm{M}^\star = \mathrm{M}_k$. Now, suppose that at some $t_2 > t_1$ the value of $\mathrm{M}_k$ increases. Then, in view of the discussion above, the update law (3) fails to track the new minimum, since (4) holds with $t_0 = t_1$.

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*, ensuring 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. [9, Section 7]), making the minimum sharing method suitable for applications in which time-varying target steady states have to be tracked. 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 inspired and 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 $\mathrm{M}_i$, hence the minimum $\mathrm{M}^\star$, may change over time. In this contexts, (i) global attractiveness allows to track the changing minimum $\mathrm{M}^\star$, (ii) uniformity of convergence guarantees that the convergence rate is always the same and does not grow unbounded 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 congestion, ghost jams, and delays. 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 maximum cruise speed. At each time, the maximum attainable cruise speed of each vehicle $i$ is constrained by a *personal maximum value*, denoted by $\mathrm{M}_i$, 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, $\mathrm{M}^\star$, compatible with all the personal velocity constraints. At each time, the problem of estimating $\mathrm{M}^\star$ 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 [3, Theorem 3.1.1]. A standard additional assumption making the problem well-posed is that each agent is characterized by a *unique identifier* $\mathrm{M}_i$. Hence, the problem of leader election can be cast as finding the minimum, $\mathrm{M}^\star$, of such identifiers. The agent whose identifier coincides with $\mathrm{M}^\star$ 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* [2], the *Max-Consensus* [4–6] (see (3)), the *MegaMerger* [10], and the *Yo-Yo* algorithm. See [2, 3] 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 employ a number of variables which does not grow with the network size or topology, and do not depend on *centralized quantities* such as parameters that need to be known in advance by all the agents. 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 [1] for a recent overview). If the functions $\psi_i$ in (2) are convex, indeed, different approaches can be used, such as consensus-based (sub)gradient methods [11–16], second-order methods [17, 18], projected [19] and primal-dual [20, 21] methods with inequality constraints, methods based on the distributed Alternate Direction Method of Multipliers (ADMM) [1, 22–29], and methods based on gradient tracking [30–35]. Gradient methods typically achieve global attractiveness. However, among the cited references only [12] deals with constrained problems with different local constraints such as (2). Yet, [12] requires a vanishing stepsize, which makes convergence not uniform. Gradient methods employing a fixed stepsize thus guaranteeing uniformity are given in [11, 13–18]. However, they do not cover constrained problems of the kind (2). Moreover, the first-order methods in [11, 13, 16] 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 trade off is more generally established between uniformity, approximation error and convergence rate. Approaches [19–21] 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 [30–35] (which, anyway, are developed for unconstrained problems), and also for the "node-based" formulations of ADMM [23–26, 28]. Instead, the "edge-based" formulations of ADMM (e.g. [1, Section 3.3], [29]) 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 proved for any of the aforementioned approaches and typically the update laws employ coefficients (e.g. stepsizes) which must be common[3] to each 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 all 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 $\mathrm{M}^\star$ stored by agent $i$. We show that all the estimates $x_i$ converge, globally and uniformly, to a stable neighborhood of $\mathrm{M}^\star$ whose size can be reduced arbitrarily around $\mathrm{M}^\star$ by suitably tuning some control parameters. More precisely, the proposed approach enjoys the following properties:

(a) The algorithm is distributed, decentralized and scalable, since only one variable is stored for each agent.

(b) It provides a global, uniform, approximate convergence result, in which the agents estimates $x_i$ converge to a stable steady state that can be made arbitrarily close to $\mathrm{M}^\star$.

(c) An exact convergence result (i.e., the local estimates all converge to $\mathrm{M}^\star$) can be achieved, at the price, however, of losing uniformity.

The paper is organized as follows. After providing preliminary concepts, 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 where details on the design of the algorithms are provided as well. Numerical results showing the effectiveness of the proposed minimum sharing technique are reported in Section 5. Finally, Section 6 delivers some concluding remarks, followed by Section 7 and the Appendix where the proofs of the main theoretical results are provided.

## 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 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 the context. With $A \subset \mathbb{R}^n$ and $x \in \mathbb{R}^n$, $\operatorname{dist}(x, A) := \inf_{a\in A} |x - a|$ denotes the distance from $x$ to $A$. Sequences indexed by a set $S$ are denoted as $(x_s)_{s\in S}$. 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 $\operatorname{dom} x$ of $\mathbb{N}$. For ease of notation, we will use $x^t$ in place of $x(t)$ to denote the values of a signal $x : \mathbb{N} \to \mathbb{R}^n$. With $t_0 \in \mathbb{N}$, we say that $x$ *starts at* $t_0$ if $\min \operatorname{dom} x = t_0$.

[3]Exceptions are given in the gradient-tracking designs of [30, 32], 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., [9, Chapter 7]).

### 2.2. Communication Networks

Throughout the paper, $\mathcal{N}$ denotes the (finite) set of agents in the network, and we let $N := \operatorname{card}\mathcal{N}$. The network communication constraints are formally captured by the concept of "communication structure" defined below[4].

**Definition 1.** *A* communication structure *on $\mathcal{N}$ is a family $\mathcal{C} = \{[i]\}_{i\in\mathcal{N}}$ of subsets $[i]$ of $\mathcal{N}$ satisfying $i \in [i]$.* ◁

For each $i \in \mathcal{N}$ the set $[i] \in \mathcal{C}$ 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

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

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 $\mathrm{M}_i$ 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 $\mathrm{M}_i$ (hence, their minimum $\mathrm{M}^\star$) 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 $\mathrm{M}^\star$ provided that its dynamics is sufficiently slow. Moreover, classical results in the context of control under different time-scales (see, e.g., [8, 36–38]) also guarantee that approximate asymptotic stability is preserved under arbitrary 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 $\mathrm{M}_i$ are subject to few impulsive changes separated by relatively large intervals of time.

[4] 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* if for each $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that every solution of (7) satisfying $\operatorname{dist}(x^{t_0}, \mathcal{A}) \leq \delta(\epsilon)$ also satisfies $\operatorname{dist}(x^t, \mathcal{A}) \leq \epsilon$, for all $t \geq t_0$.

We say that the set $\mathcal{A}$ 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 $\operatorname{dist}(x^t, \mathcal{A}) \leq \epsilon$ 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.
- *$\varepsilon$-approximately attractive* (with $\varepsilon > 0$) if the set $\{x \in \mathbb{R}^n \mid \operatorname{dist}(x, \mathcal{A}) \leq \varepsilon\}$ 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 $\Gamma$, 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 $\varepsilon > 0$, there exists $\gamma^\star(\varepsilon) \in \Gamma$ such that the set $\mathcal{A}$ is $\varepsilon$-approximately attractive for the system (8) obtained with $\gamma = \gamma^\star(\varepsilon)$.

Finally, we remark that attractiveness implies finite-time $\varepsilon$-approximate attractiveness for each $\varepsilon > 0$, and practical attractiveness implies finite-time practical attractiveness.

## 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 $\mathrm{M}_i$, not known a priori by the others, and it stores and updates a local estimate $x_i \in \mathbb{R}$ of the quantity $\mathrm{M}^\star$ defined in (1). Thus, the problem at hand consists in *designing an update law* for each agent $i \in \mathcal{N}$ of the form

$$x_i^{t+1} = f_i(t, x^t), \tag{9}$$

such that the resulting estimates $x_i^t$ converge to $\mathrm{M}^\star$, 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 (9) as

$$x^{t+1} = f(t, x^t). \tag{10}$$

As each agent is allowed to exchange information only with the agents belonging to its neighborhood $[i] \in \mathcal{C}$, 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*

$$\mathcal{A} := \{\mathrm{M}^\star\}^N \tag{11}$$

*is globally attractive for* (10). ◁

**Remark 1.** *We stress that, if $f$ is $\mathcal{C}$-decentralized in the sense of Definition 4, then each function $f_i$ in* (9) *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 additional main assumptions specified hereafter. In the following we let

$$I^\star \;:=\; \operatorname*{argmin}_{i\in\mathcal{N}} \mathrm{M}_i. \tag{12}$$

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 $\mathrm{M}^\star$.

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

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 $\mathrm{M}^\star$. 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 $\mathrm{M}^\star$ is a mild requirement. For instance, in both the traffic control and leader election problems we can assume that the quantities $\mathrm{M}_i$ are integers, so that "$\mu_i \in (0,1)$ for all $i \in \mathcal{N}$" is a feasible choice requiring no further knowledge on $\mathrm{M}^\star$. Furthermore, this assumption is not in principle needed if an approximate or practical attractiveness result is sought. In fact, if for some $I \subset \mathcal{N}$, $\varepsilon := \max_{i\in I} \mu_i > \mathrm{M}^\star$, then $\mathrm{M}^\star \in (0, \varepsilon)$, and, as clarified later on by the asymptotic analysis, we are able to claim that the set $[0, \varepsilon]^N$ (which includes $\mathrm{M}^\star$) is practically attractive for $x$, with $\varepsilon$, however, that can be made arbitrarily small by choosing $\mu_i$ accordingly.

In the following we let

$$\underline{\mu} := \min_{i\in\mathcal{N}} \mu_i. \tag{13}$$

### 3.3. The Update Laws

The proposed update law is obtained by choosing $f$ so that, for each $i \in \mathcal{N}$, Equation (9) reads as follows[5]

$$x_i^+ = \Pi_{[\mu_i,\, \mathrm{M}_i]}\left[ \mathrm{e}^{h_i^t} x_i + k_i \sum_{j\in[i]} (x_j - x_i) \right], \tag{14}$$

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

$$0 < k_i \le \frac{1}{\operatorname{card}([i] \setminus i)} \tag{15}$$

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

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

Notice that, as in [12], the update laws (14) have the form of a projected (onto the interval $[\mu_i, \mathrm{M}_i]$) consensus-like protocol. Unlike [12], however, the resulting consensus matrix needs *not* be column- or row-stochastic, and the coefficients $k_i$ are only constrained by (15) 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*[6] is not marginally stable. Rather, it is deliberately made unstable by the terms $\mathrm{e}^{h_i^t}$.

### *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*

$$m \leq \frac{1}{\underline{h}(t_0)} \log\left(\frac{\mathrm{M}^\star}{\underline{\mu}}\right) \tag{16}$$

*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 *persistence of excitation* property.

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

Persistence 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.

**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 persistently exciting.* ◁

Lemma 1 is proved in Appendix A.

[6] 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* (14)*. Since $\mathrm{M}^\star \in \mathcal{M}$, this implies that $\mathrm{M}^\star$ is a well-defined steady state for* (14)*. 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 intimately linked with the impossibility result [3, 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 [7, 8].* ◁

## 4. Convergence Results

### *4.1. Main result*

For ease of notation, we write the update laws (14) in the compact form (10). The following theorem – which is the main fundamental result of the paper – relates the excitation properties the signals $h_i$ to the asymptotic convergence to $\mathrm{M}^\star$ of the estimates $x_i$ produced by the update laws (14). In particular, it shows that sufficiency of excitation implies convergence (possibly exact) and persistence 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 1 and 2, consider the update laws* (14)*, in which $k_i$ satisfies* (15)*. 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* (10) *starting at $t_0$ satisfies*

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

   *with $I^\star$ given by* (12)*.*
2. *For each $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that, if*

$$\limsup_{t\to\infty} h_i^t \leq \delta(\epsilon), \quad \forall i \in \mathcal{N}, \tag{17}$$

   *then each solution $x$ starting at $t_0$ satisfies*

$$\lim_{t\to\infty} |x_i^t - \mathrm{M}^\star| \leq \epsilon, \quad \forall i \in \mathcal{N}. \tag{18}$$

   *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* (10)*.*

3. *If the family* $(h_i)_{i\in\mathcal{N}}$ *is persistently exciting in the sense of Definition 6, then* $\mathcal{A}_\varepsilon$ *is globally uniformly attractive.*
4. *If all the signals* $h_i$ *are non-zero and periodic (with possibly different period), then there exists a compact set* $\mathcal{A}_\varepsilon^u \subset \mathcal{A}_\varepsilon$ *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* (11)*, is globally attractive for* (10)*, 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 $(h_i)_{i\in\mathcal{N}}$ carries enough excitation (in the sense of Definition 5), then, in 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 (14) 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 (11)). More precisely, let $\Gamma$ be the set of all the families $\gamma := (h_i)_{i\in\mathcal{N}}$ of functions $h_i : \mathbb{N} \to \mathbb{R}_{\ge 0}$, 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

$$f_\gamma^i(t,x) := \Pi_{[\mu_i,\,\mathrm{M}_i]}\left[\mathrm{e}^{h_i^t} x_i + k_i \sum_{j\in[i]} \left(x_j - x_i\right)\right]. \qquad (19)$$

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* (19)*.* ◁

Claim 3 of the theorem further strengthen (1) to a uniform global practical stability property of $\mathcal{A}$ in presence of persistence of excitation. Claim 4, moreover, also states that, in the relevant case in which the signals $h_i$ are periodic, with possibly different periods, there exists a compact set included in the approximation $\mathcal{A}_\epsilon$ 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 persistently 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 persistence of excitation: sufficiency of excitation allows exact convergence, but prevents uniformity in the initial time. Persistence of excitation, instead guarantees uniformity and stability. It, however, frustrates exact convergence, guaranteeing only a weaker practical result. This, in turn, reveals a somehow *necessary compromise between complexity, uniformity and convergence performance* which is an interesting insight.

### *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 (14). 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 persistence of excitation.

#### *4.3.1. Sufficiently Exciting Designs*

Sufficiency of excitation of the family $(h_i)_{i\in\mathcal{N}}$ (in the sense of Definition 5) is guaranteed if each $h_i$ takes "enough" positive values. According to Definition 5, and in particular to (16), 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 (16) is undesirable as inevitably centralized and not robust. A simple decentralized way to design a sufficiently exciting family $(h_i)_{i\in\mathcal{N}}$ is rather to choose the signals $h_i$ so as each $h_i$ is bounded and

$$\sum_{t\in\mathbb{N}} h_i^t = \infty, \qquad \forall i \in \mathcal{N}. \qquad (20)$$

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* (20)*. 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 (20), hence it is omitted.

In view of Claim 5 of Theorem 1, exact convergence of the estimates $x_i$ to $\mathrm{M}^\star$ 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

$$\sum_{t\in\mathbb{N}} \left(h_i^t\right)^2 < \infty. \qquad (21)$$

It is interesting to notice that Properties (20)-(21) are standard assumptions asked to the *step size* in classical *stochastic approximation algorithms* [39, 40], as well as in modern distributed optimization algorithms using vanishing step sizes [1, 12, 41]. 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. Persistently Exciting Designs*

In view of Lemma 6, if every signal $h_i$ is periodic, then $(h_i)_{i\in\mathcal{N}}$ is persistently exciting in the sense of Definition 6. While periodicity is not necessary for persistence 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 persistently 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

$$h_i^t = A_i \text{step}\left(\text{mod}(t, T_i) - (1 - \rho_i)(T_i)\right) \tag{22}$$

   in which $\text{mod}(s) := s - \max\{n \in \mathbb{N} \mid n(T_i + 1) \le s\}$, and $\text{step}(\cdot)$ denotes the *step function* satisfying $\text{step}(s) = 0$ for $s < 0$ and $\text{step}(s) = 1$ for $s \ge 0$. The signal (22) has period $T_i$ and $h_i^t = A_i$ holds for $\rho_i T_i$ seconds each period.

## 5. Numerical Simulations

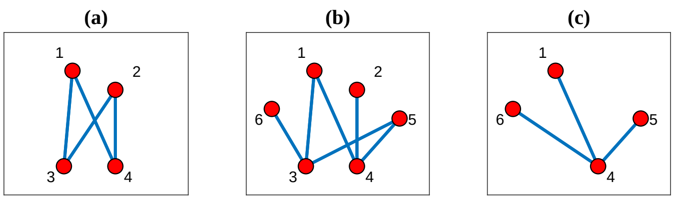


Figure 1: Communication structure of Simulation 1 **(a)** for $0 \le t < 500$, **(b)** for $500 \le t < 1500$, and **(c)** for $t \ge 1500$.

In this section, we present two illustrative numerical simulation scenarios of the proposed methodology. In Scenario 1, a network with a time-changing topology (see Figure 1) is considered while in Scenario 2, for a fixed network topology, the use of different signals $h_i$ is evaluated.

### *5.1. Scenario 1: Uniform Convergence*

The first simulation, shown in Figure 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 Figure 1-(a) and given by $[1] = \{1, 3, 4\}$, $[2] = \{2, 3, 4\}$, $[3] = \{1, 2, 3\}$, $[4] = \{1, 2, 4\}$, and with numbers $(\mathrm{M}_1, \mathrm{M}_2, \mathrm{M}_3, \mathrm{M}_4) = (10, 12, 13, 13)$, implying $\mathrm{M}^\star = \mathrm{M}_1 = 10$. The update laws (14) 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 Figure 1-(b), and given by $[1] = \{1, 3, 4\}$, $[2] = \{2, 4\}$, $[3] = \{1, 3, 5, 6\}$, $[4] = \{1, 2, 4, 5\}$, $[5] = \{3, 4, 5\}$ and $[6] = \{3, 6\}$. The new agents have numbers $(\mathrm{M}_5, \mathrm{M}_6) = (7, 11)$, 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 $(\mathrm{M}_1, \mathrm{M}_3) = (11, 13)$. The new optimum is thus $\mathrm{M}^\star = \mathrm{M}_5 = 7$.

- At time $t = 1500$, agents 2 and 3 leave the network, and the communication structure is changed to that depicted in Figure 1-(c), i.e. $[1] = \{1, 4\}$, $[4] = \{1, 4, 5, 6\}$, $[5] = \{4, 5\}$ and $[6] = \{4, 6\}$. Moreover, the numbers of the agents are changed to $(\mathrm{M}_1, \mathrm{M}_4, \mathrm{M}_5, \mathrm{M}_6) = (12, 16, 11, 16)$, leading to $\mathrm{M}^\star = \mathrm{M}_5 = 11$.

- Finally, at time $t = 5000$, the numbers of Agent 4 is changed to $\mathrm{M}_4 = 8$, so as $\mathrm{M}^\star = \mathrm{M}_4 = 8$.

As Figure 2 shows, the convergence to the (time-varying) optimum $\mathrm{M}^\star$ is approximate, with the trajectories of the agents which show residual oscillations. Figure 2 also underlines that the convergence to $\mathrm{M}^\star$ "from below" (i.e. when the initial values of the agent are smaller than $\mathrm{M}^\star$) is slower than convergence "from above" (i.e. when the initial values of the agent are larger than $\mathrm{M}^\star$). 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 in the contest of Claim 1 of Theorem 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 in the contest of Claims 2 and 3 of Theorem 1, is determined by the values of the coefficients $k_i$.

Finally, for the sake of comparison, Figure 3 shows a simulation in which the Max-Consensus (3) is employed in the same setting described above (see Figure 2). As shown in Figure 3, although showing a faster convergence within the first two changes of $\mathrm{M}^\star$, the Max-Consensus fails in

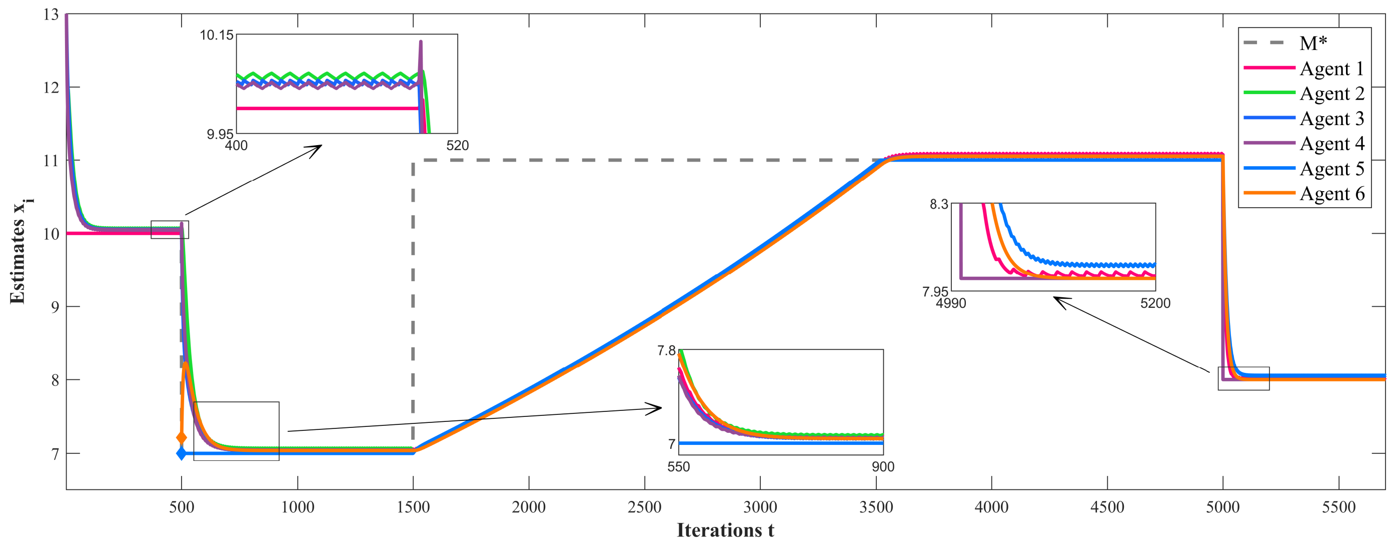


Figure 2: Evolution of the estimates $x_i$ in Scenario 1. The trajectory of the optimal value $\mathrm{M}^\star$ is shown in dashed, gray line. Colored lines depicts instead the trajectory of the estimates $x_i$, $i = 1, \dots, 6$. In abscissa: iteration variable $t$.

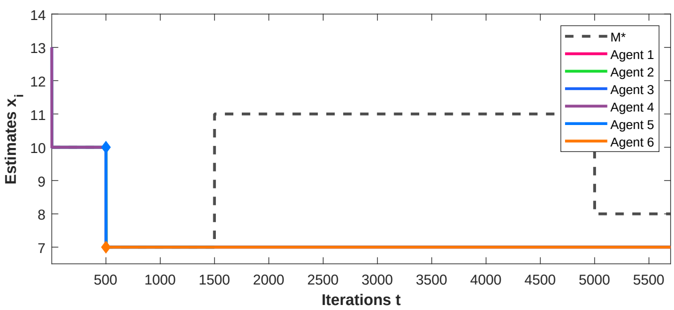


Figure 3: Evolution of the Max-Consensus estimates (update law (3)) in the setting of Scenario 1 (cf. Figure 2). In abscissa: iteration variable $t$.

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 Figure 1-(a), i.e. $[1] = \{1,3,4\}$, $[2] = \{2,3,4\}$, $[3] = \{1,2,3\}$, $[4] = \{1,2,4\}$. Initially, the agents are given numbers $(\mathrm{M}_1,\mathrm{M}_2,\mathrm{M}_3,\mathrm{M}_4) = (3,6,9,15)$, so as $\mathrm{M}^\star = \mathrm{M}_1 = 3$. At time $t = 500$, $\mathrm{M}_1$ is changed to 15, so as $\mathrm{M}^\star = \mathrm{M}_2 = 6$. At time $t = 20000$, $\mathrm{M}_2$ is changed to 15, so as $\mathrm{M}^\star = \mathrm{M}_3 = 9$. At time $t = 35000$, $\mathrm{M}_3$ is changed to 12, do as $\mathrm{M}^\star = M_3 = 12$. Finally, at time $t = 150000$, $\mathrm{M}_3$ is changed to 15, so as $\mathrm{M}^\star = \mathrm{M}_1 = \mathrm{M}_2 = \mathrm{M}_3 = \mathrm{M}_4 = 15$. 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 persistently exciting (in the sense of Definition 6) 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 $h^t_{i'} = (1+t)^{-1}$ for all $i' \in \mathcal{N}'$. The signals $h_{i'}$ satisfy (20)-(21) and, thus, $(h_{i'})_{i'\in\mathcal{N}'}$ is sufficiently exciting in the sense of Definition 5. However, it fails to be persistently exciting in the sense of Definition 6. The simulation shown in Figure 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 $\mathrm{M}^\star$ is followed by the estimates $x_i$ with the same convergence rate. On the contrary, $\mathrm{M}^{\star\prime} = \mathrm{M}^\star$ 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 persistently 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 $\mathrm{M}^{\star\prime}$ with higher precision (by Claim 5 of Theorem 1, indeed, since $h^t_{i'} \to 0$, the convergence of the estimates $x_{i'}$ is asymptotic if $\mathrm{M}^{\star\prime}$ remains constant), whereas the estimates $x_i$ keep a non-diminishing convergence 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, Figure 5 shows a simulation of the Max-Consensus (3) in this second setting (cf. Figure 4). Again, the Max-Consensus fails in tracking the time-varying $\mathrm{M}^\star$.

## 6. Concluding Remarks

In this paper, we proposed a novel methodology for the problem of distributed minimum sharing or, equivalently, for the class of distributed optimization problems (2). The

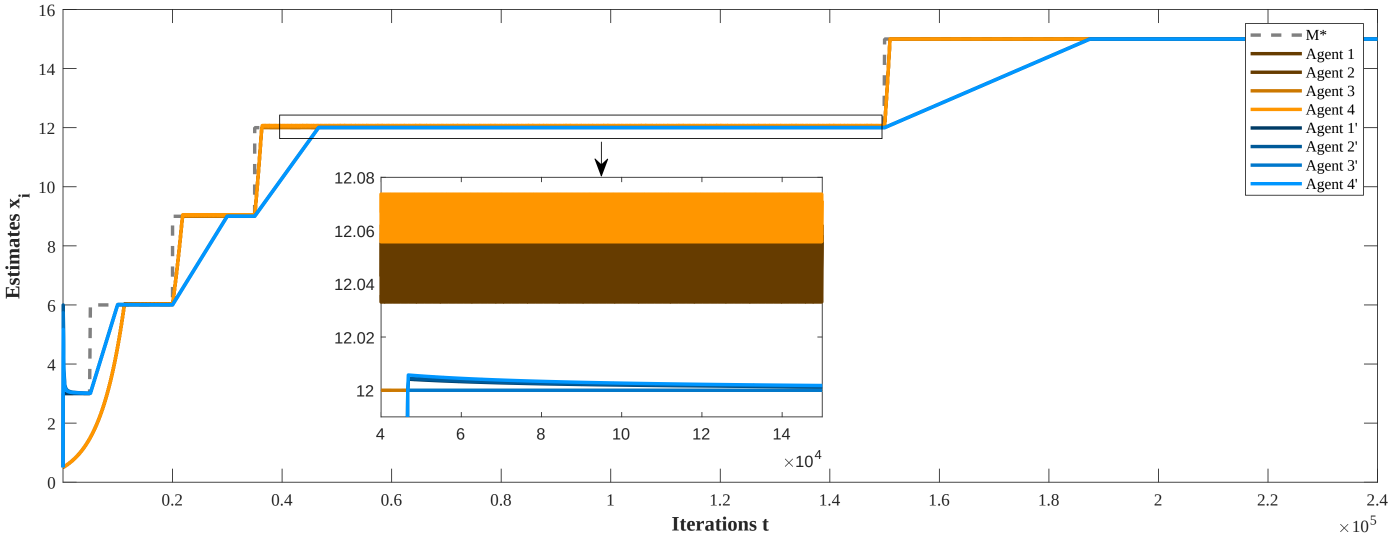


Figure 4: Evolution of the estimates $x_i$ in Scenario 2. The trajectory of the optimal value $\mathrm{M}^\star$ is shown in dashed, gray line. Dark to light orange lines depicts the trajectory of the estimates $x_i$, $i = 1, \ldots, 4$ of the first network. Dark to light blue lines depicts the trajectory of the estimates $x_i$, $i = 1', \ldots, 4'$ of the second network. In abscissa: iteration variable $t$.

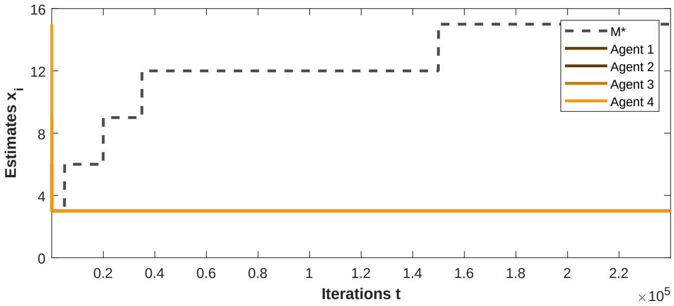


Figure 5: Evolution of the Max-Consensus estimates (update law (3)) in the setting of Scenario 2 (cf. Figure 4). In abscissa: iteration variable $t$.

proposed solution has many desirable properties: it is distributed, decentralized, scalable, global in the initial conditions, stable, and possibly asymptotically convergent or uniform in time. In particular, by Theorem 1 and Section 4.3.2, constructive conditions for global uniform practical, or global non-uniform exact, asymptotic stability are given, and numerical simulations are presented to illustrate the result. The entire analysis has been carried out under the assumption that the communication structure and the parameters remain constant during the execution time.

Although the aforementioned stability, globality, and uniformity properties guarantee a good behavior for "slowly varying" structures (shown also by the numerical simulations), additional work is needed to extend the proposed approach to structurally 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 future research directions are towards this aim.

## 7. Proof of Theorem 1

For the sake of clarity, we organize the proof in five subsections, one for each of the claims of the Theorem.

### 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 $t_0 \in \mathbb{N}$ (in the sense of Definition 5), and if for each $i \in \mathcal{N}$, (15) holds, then there exists $t^\star = t^\star(t_0) > t_0$ such that, for each $i \in \mathcal{N}$, $x_i^t \geq \mathrm{M}^\star$ holds for all $t \geq t^\star$ and, for each $i \in I^\star$, $x_i^t = \mathrm{M}^\star$ holds for all $t \geq t^\star$.

Define the function

$$\underline{i} : \mathbb{X} \to \mathcal{N}, \qquad x \mapsto \underline{i}(x) := \operatorname*{argmin}_{i\in\mathcal{N}} x_i.$$

Then, $x_j \geq x_{\underline{i}(x)}$ holds for all $j \in \mathcal{N}$. Moreover, $h_i^t \geq 0$ and (15) imply

$$\mathrm{e}^{h_i^t} - \mathrm{card}([i] \setminus i)k_i \geq 0, \qquad \forall i \in \mathcal{N}.$$

Since $\Pi_{[\mu_i,\mathrm{M}_i]}$ is increasing[7] we have

$$\begin{aligned} x_i^{t+1} &= \Pi_{[\mu_i,\mathrm{M}_i]}\left[\left(\mathrm{e}^{h_i^t} - \mathrm{card}([i]\setminus i)k_i\right)x_i^t + k_i \sum_{j\in[i]\setminus i} x_j^t\right] \\ &\geq \Pi_{[\mu_i,\mathrm{M}_i]}\left[\left(\mathrm{e}^{h_i^t} - \mathrm{card}([i]\setminus i)k_i\right)x_{\underline{i}(x^t)}^t \right.\\ &\qquad \left. + \mathrm{card}([i]\setminus i)k_i x_{\underline{i}(x^t)}^t\right] \\ &= \Pi_{[\mu_i,\mathrm{M}_i]}\left[\mathrm{e}^{h_i^t} x_{\underline{i}(x^t)}^t\right] \\ &= \max\left\{\mu_i, \min\left\{\mathrm{e}^{h_i^t} x_{\underline{i}(x^t)}^t, \mathrm{M}_i\right\}\right\} \\ &\geq \min\left\{\mathrm{e}^{h_i^t} x_{\underline{i}(x^t)}^t, \mathrm{M}_i\right\} \geq \min\left\{\mathrm{e}^{h_i^t} x_{\underline{i}(x^t)}^t, \mathrm{M}^\star\right\} \end{aligned} \tag{23}$$

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

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

First, notice that, if for some $\bar{t} \in \mathbb{N}$, $x^{\bar{t}}_{\underline{i}(x^{\bar{t}})} \geq \mathrm{M}^\star$, then (23) implies $x^{\bar{t}+1}_{\underline{i}(x^{\bar{t}+1})} \geq \mathrm{M}^\star$, so that by induction it is possible to conclude that $x_i^t \geq \mathrm{M}^\star$ 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

$$x^t_{\underline{i}(x^t)} < \mathrm{M}^\star, \qquad \forall t \geq t_0. \tag{24}$$

Then, we show that, if the signals $h_i$ are sufficiently exciting from $t_0$ (in the sense of Definition 5), then (24) leads to a contradiction, in this way proving the claim.

Thus, assume that (24) holds. Then, since $h_i^t \geq 0$ for all $i \in \mathcal{N}$, (23) yields

$$x_i^{t+s} \geq \mathrm{e}^{h_i^t} x^t_{\underline{i}(x^t)}, \qquad \forall t \geq t_0,\ s \geq 1. \tag{25}$$

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 (25), this yields

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

and thus, in particular,

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

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 (25), one has

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

By repeating the same arguments, it is thus possible to conclude that, for each $m \in \mathbb{N}$ satisfying (16), one has

$$x^{t_0+1+m\Delta(t_0)}_{\underline{i}\left(x^{t_0+1+m\Delta(t_0)}\right)} \geq \mathrm{e}^{m\underline{h}(t_0)}\, x^{t_0+1}_{\underline{i}(x^{t_0+1})} \geq \mathrm{e}^{m\underline{h}(t_0)}\underline{\mu}, \tag{26}$$

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

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

Then, with $\bar{t} := t_0 + 1 + m^\star(t_0)\Delta(t_0)$, from (26) we obtain

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

which contradicts (24) and, thus, proves that $x_i^t \geq \mathrm{M}^\star$ holds for all $i \in \mathcal{N}$ and all $t \geq t^\star := \bar{t}$.

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

### *7.2. Proof of Claim 2*

In this subsection, we prove the second claim of the theorem. We first show that all the estimates $x_i$ are asymptotically bounded by a function of $\mathrm{M}^\star$ and $\limsup_{t\to\infty} h_i^t$. Then, we show that this bound can be reduced arbitrarily by reducing $\limsup_{t\to\infty} h_i^t$, for all $i \in \mathcal{N}$, accordingly.

Since by Claim 1 each $x_i$ satisfies $x_i^t \geq \mathrm{M}^\star$ 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

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

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

$$x_i^t \leq \mathrm{M}_i \tag{27}$$

and

$$x_i^{t+1} \leq \mathrm{e}^{h_i^t} x_i^t + k_i \sum_{j\in[i]} \left(x_j^t - x_i^t\right) \tag{28}$$

for all $i \in \mathcal{N}$ and all $t \geq t^\star$. From (27) we also obtain

$$\limsup_{t\to\infty} |x_i^t| \leq \mathrm{M}_i < \infty, \qquad \forall i \in \mathcal{N}. \tag{29}$$

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*

$$x^{t+1} \leq \lambda^t x^t + y^t \tag{30}$$

*for all $t \geq t_0$. Then*

$$\limsup_{t\to\infty} |x^t| \leq \frac{1}{1 - \limsup_{t\to\infty} \lambda^t} \limsup_{t\to\infty} |y^t|. \tag{31}$$

◁

With $I^\star$ defined in (12), 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 focus on the case in which $n^\star > 0$.

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

$$\max_{i\in[I^\star]^m_{m-1}} \limsup_{t\to\infty} |x_i^t| \leq \alpha_m \max_{j\in[I^\star]^{m+1}_m} \limsup_{t\to\infty} |x_j^t| + \beta_m \mathrm{M}^\star. \tag{32}$$

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

First notice that, for each $i \in [I^\star]^{m+1}_m$, every $j \in [i]$ belongs to exactly one among the sets $[I^\star]^{m+2}_{m+1}$, $[I^\star]^{m+1}_m$, and $[I^\star]^m_{m-1}$. Hence, in view of (28), we can write

$$\begin{aligned} x_i^{t+1} \leq{} & \left(\mathrm{e}^{h_i^t} - k_i \operatorname{card}([i] \setminus i)\right) x_i^t + k_i \sum_{j\in[i]\cap[I^\star]^m} x_j^t \\ & + k_i \sum_{j\in([i]\setminus i)\cap[I^\star]^{m+1}_m} x_j^t + k_i \sum_{j\in[i]\cap[I^\star]^{m+2}_{m+1}} x_j^t \end{aligned} \tag{33}$$

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

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 (15) 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

$$h_i^t \leq \bar{h}_{i,1} + \nu_1 = \log(1 + k_i \operatorname{card}([i] \setminus i)) - \nu_1 \tag{34}$$

for all $t \geq T^\star$ and all $i \in \mathcal{N}$. Thus, (15) and (34) 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 (29), (33) and Lemma 3 imply

$$\begin{aligned}\limsup_{t\to\infty} |x_i^t| \leq \gamma_i &\sum_{j\in[i]\cap[I^\star]^m} \limsup_{t\to\infty} |x_j^t| \\ &+ \gamma_i \sum_{j\in([i]\setminus i)\cap[I^\star]_m^{m+1}} \limsup_{t\to\infty} |x_j^t| \\ &+ \gamma_i \sum_{j\in[i]\cap[I^\star]_{m+1}^{m+2}} \limsup_{t\to\infty} |x_j^t|\end{aligned} \tag{35}$$

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

$$\gamma_i := \frac{k_i}{1 - \limsup_{t\to\infty}\left(e^{h_i^t} - k_i \operatorname{card}([i] \setminus i)\right)} \tag{36}$$

which exists finite in view of Lemma 3. In view of (32), equation (35) implies

$$\begin{aligned}\limsup_{t\to\infty} |x_i^t| \leq \left(c_{i,1}\alpha_m + c_{i,2}\right) &\max_{j\in[I^\star]_m^{m+1}} \limsup_{t\to\infty} |x_j^t| \\ &+ c_{i,3} \max_{j\in[I^\star]_{m+1}^{m+2}} \limsup_{t\to\infty} |x_j^t| + c_{i,1}\beta_m \mathrm{M}^\star.\end{aligned} \tag{37}$$

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

$$\begin{aligned}c_{i,1} &:= \gamma_i \operatorname{card}\left([i] \cap [I^\star]^m\right) \\ c_{i,2} &:= \gamma_i \operatorname{card}\left(([i] \setminus i) \cap [I^\star]_m^{m+1}\right) \\ c_{i,3} &:= \gamma_i \operatorname{card}\left([i] \cap [I^\star]_{m+1}^{m+2}\right).\end{aligned} \tag{38}$$

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

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

If

$$\limsup_{t\to\infty} h_i^t \leq \bar{h}_i \tag{39}$$

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

$$\begin{aligned}1 - e^{\limsup_{t\to\infty} h_i^t} &\geq 1 - e^{\bar{h}_i} \geq -k_i(1 - \alpha_m) + \nu_2 \\ &\geq -k_i(1 - \alpha_m)\operatorname{card}([i] \cap [I^\star]^m) + \nu_2\end{aligned} \tag{40}$$

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

$$\begin{aligned}&\operatorname{card}\left(([i] \setminus i) \cap [I^\star]_m^{m+1}\right) \\ &= \operatorname{card}\left([i] \setminus i\right) - \operatorname{card}\left([i] \cap [I^\star]^m\right) - \operatorname{card}\left([i] \cap [I^\star]_{m+1}^{m+2}\right) \\ &\leq \operatorname{card}\left([i] \setminus i\right) - \operatorname{card}\left([i] \cap [I^\star]^m\right),\end{aligned}$$

then, we conclude that

$$\begin{aligned}&c_{i,1}\alpha_m + c_{i,2} \\ &\quad\leq \frac{k_i(\alpha_m - 1)\operatorname{card}\left([i] \cap [I^\star]^m\right) + k_i \operatorname{card}\left([i] \setminus i\right)}{1 - e^{\bar{h}_i} + k_i \operatorname{card}([i] \setminus i)} \\ &\quad\leq \frac{(\alpha_m - 1)k_i \operatorname{card}\left([i] \cap [I^\star]^m\right) + k_i \operatorname{card}([i] \setminus i)}{(\alpha_m - 1)k_i \operatorname{card}\left([i] \cap [I^\star]^m\right) + k_i \operatorname{card}([i] \setminus i) + \nu_2} \\ &\quad< 1.\end{aligned} \tag{41}$$

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

Now, since (37) holds for each $i \in [I^\star]_m^{m+1}$, it in particular holds for $\bar{i}$ satisfying

$$\bar{i} \in \operatorname*{argmax}_{i\in[I^\star]_m^{m+1}} \limsup_{t\to\infty} |x_i^t|, \tag{42}$$

so that (37) implies

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

which, in view of (41), yields

$$\begin{aligned}\max_{i\in[I^\star]_m^{m+1}} \limsup_{t\to\infty} |x_i^t| \leq \alpha_{m+1} &\max_{j\in[I^\star]_{m+1}^{m+2}} \limsup_{t\to\infty} |x_j^t| \\ &+ \beta_{m+1}\mathrm{M}^\star\end{aligned} \tag{43}$$

with

$$\begin{aligned}\alpha_{m+1} &= \frac{c_{\bar{i},3}}{1 - \left(c_{\bar{i},1}\alpha_m + c_{\bar{i},2}\right)}, \\ \beta_{m+1} &= \frac{c_{\bar{i},1}}{1 - \left(c_{\bar{i},1}\alpha_m + c_{\bar{i},2}\right)}\beta_m.\end{aligned} \tag{44}$$

Furthermore, since $\limsup_{t\to\infty} h_i^t \leq \bar{h}_i$, in view of (40), $\alpha_{m+1}$ satisfies

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

Therefore, we claim that if (32) holds for some $m \in \{0, \ldots, n^\star - 1\}$ with $\alpha_m < 1$ and $\beta_m \geq 0$, then (43) holds as well for $m + 1$ with $\alpha_{m+1} < 1$ and $\beta_{m+1}$ given above.

Since by Claim 1, Equation (32) trivially holds for $m = 0$ with $\beta_0 = 1$ and $\alpha_0 = 0$, then we claim by induction that, if

$$\limsup_{t\to\infty} h_i^t \leq \bar{h} := \min_{i\in\mathcal{N}} \bar{h}_i, \qquad \forall i \in \mathcal{N}, \tag{45}$$

then Equation (32) holds for each $m \in \{0, \ldots, n^\star\}$.

Now, for $m = n^\star$, we have $[I^\star]^{m+1} \setminus [I^\star]^m = \emptyset$, so that (32) yields

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

Thus, iterating (32) backwards and using (27) yield

$$\limsup_{t\to\infty} x_i^t \le \min\Big\{\mathrm{M}_i,\ (1+\varepsilon_i)\mathrm{M}^\star\Big\} \tag{46}$$

in which

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

and

$$\varepsilon_i = \sum_{\ell=0}^{n^\star-m} \left(\prod_{k=\ell+1}^{n^\star-m} \alpha_{n^\star-k}\right) \beta_{n^\star-\ell} - 1, \tag{47}$$

for all $i \in [I^\star]_{m-1}^m$ and all $m = 1, \dots, n^\star$. Moreover, by using (46), we deduce from (32) that the quantities $\varepsilon_i$ also satisfy the recursion

$$\max_{i\in[I^\star]_{m-1}^m} \varepsilon_i \le \alpha_m \left(1 + \max_{i\in[I^\star]_m^{m+1}} \varepsilon_i\right) + \beta_m - 1. \tag{48}$$

We now prove that $\varepsilon_i$ in (46)-(47) can be reduced arbitrarily by reducing $\limsup_{t\to\infty} h_i^t$ accordingly for each $i \in \mathcal{N}$. For convenience, let

$$\upsilon_i := \limsup_{t\to\infty} h_i^t \in [0, \bar{h}_i]. \tag{49}$$

Then, the quantities $\gamma_i$, defined in (36), depend on $\upsilon_i$ as follows

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

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

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

In view of the definitions (38), also the quantities $\alpha_m$ and $\beta_m$, as defined in (44), depend on $\upsilon_{\bar{i}}$ through $\gamma_{\bar{i}}$, in which $\bar{i}$ satisfies (42). We now prove by induction that, by letting $\upsilon := (\upsilon_1, \dots, \upsilon_N)$, the following holds

$$\lim_{\upsilon\to 0} \alpha_m(\upsilon) + \beta_m(\upsilon) = 1, \qquad \forall m = 0, \dots, n^\star. \tag{50}$$

First notice that (50) trivially holds for $m = 0$, as indeed $\alpha_m = 0$ and $\beta_m = 1$ despite the value of $\upsilon$. It thus suffices to show that if (50) holds for a given $m \in \{0, \dots, n^\star - 1\}$, then the same relation holds as well for $m+1$. For, assume that (50) 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 $\rho_1 := \operatorname{card}([\bar{i}] \cap [I^\star]^m)$, $\rho_2 := \operatorname{card}(([\bar{i}] \setminus \bar{i}) \cap ([I^\star]_m^{m+1}))$, $\rho_3 := \operatorname{card}([\bar{i}] \cap ([I^\star]_{m+1}^{m+2}))$, and noting that $\operatorname{card}([\bar{i}] \setminus \bar{i}) - \rho_2 = \rho_1 + \rho_3$, we obtain

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

Thus, by induction, we claim (50) 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 $[I^\star]_{n^\star}^{n^\star+1} = \emptyset$), then $\alpha_{n^\star} = 0$. Thus,

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

In view of (47), this implies

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

In view of (48), $\lim_{\upsilon\to 0} \max_{i\in[I^\star]_m^{m-1}} \varepsilon_i(\upsilon) = 0$ implies

$$\lim_{\upsilon\to 0} \max_{i\in[I^\star]_{m-2}^{m-1}} \varepsilon_i(\upsilon) = \lim_{\upsilon\to 0} (\alpha_{m-1}(\upsilon) + \beta_m(\upsilon)) - 1 = 0,$$

so that, by induction, we conclude that

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

i.e.

$$\lim_{\upsilon\to 0} \varepsilon_i(\upsilon) = 0, \quad \forall i \in \mathcal{N}. \tag{51}$$

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

$$\limsup_{t\to\infty} h_i^t \le \delta(\epsilon) := \min\left\{\bar{h},\ \frac{\delta'(\epsilon)}{N}\right\}, \qquad \forall i \in \mathcal{N} \tag{52}$$

then $|\upsilon| \le \delta'(\epsilon)$, which implies $\mathrm{M}^\star \varepsilon_i \le \epsilon$, which, in view of (46), implies

$$\limsup_{t\to\infty} x_i^t \le \min\Big\{\mathrm{M}_i,\ \mathrm{M}^\star + \epsilon\Big\}, \tag{53}$$

Claim 2 thus follows by (53) and by noticing that Claim 1 implies $\limsup_{t\to\infty} x_i \ge \mathrm{M}^\star$.

### 7.3. Proof of Claim 3

The third claim of the theorem, i.e., that persistence of excitation (in the sense of Definition 6) of $(h_i)_{i\in\mathcal{N}}$ implies uniform attractiveness of $\mathcal{A}_\epsilon := \prod_{i\in\mathcal{N}} \left[\mathrm{M}^\star, \min\{\mathrm{M}^\star + \epsilon, \mathrm{M}_i\}\right]$, directly follows by the fact that, if the family $(h_i)_{i\in\mathcal{N}}$ is persistently exciting, then in the above analysis $t^\star$ does not depend on $t_0$ and, therefore, the convergence (53) is uniform in the initial time.

### 7.4. Proof of Claim 4

In this subsection we prove the fourth claim of the theorem. In particular, we first show that when the signals $h_i$ are periodic, then thy can be generated by an autonomous difference equation. Then, we prove that the cascade of this new system on the update laws (14) has a well-defined limit set which has the desired stability properties.

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

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

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

Then, each $\tau_i$-periodic signal $h_i$ satisfies the following equations

$$\begin{array}{rcl} \xi_i^{t+1} & = & F_i \xi_i^t \\ h_i^t & = & C_i \xi_i^t \end{array} \tag{54}$$

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 persistently exciting for some $\underline{h} > 0$ in the sense of Definition 6. For a fixed $\epsilon > 0$, let $\delta(\epsilon)$ be defined as above in (52), and let

$$\Xi_i := \Big\{ \xi \in \mathbb{R}^{\tau_i} \mid \forall j \in \{1, \dots, r\},\ \xi_{i,j} \in [0, \delta(\epsilon)], \text{ and } \exists j \in \{1, \dots, r\},\ \xi_{i,j} \geq \underline{h} \Big\},$$

where $\xi_{i,j}$ denotes the $j$-th component of $x_i$. Then, $\Xi_i$ is compact and invariant for (54). We now consider the interconnection between (54) and the update laws (14) 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

$$z^{t+1} = \phi(z^t), \qquad z^t \in Z \tag{55}$$

with $\phi$ 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 (14) 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 (55) 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 (55) starting at 0 from $K$ and, for each $t \in \mathbb{N}$, define the reachable set from $K$ as $\mathcal{R}^t(K) := \big\{ (\xi^s, x^s) \in \Xi \times \mathbb{R}^N \mid (\xi, x) \in \mathcal{S}(K),\ s \geq t \big\}$. In view of the above analysis, and since $\Xi$ is invariant for (55), 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 $\phi$ is continuous by construction, then $\Omega(K)$ is also forward invariant, uniformly globally attractive for (55) from $K$ (see e.g. [9, Proposition 6.26]), and it is the smallest set having the above properties. Furthermore, we notice that, by definition of the update laws (14), $x_i^t \in [\mu_i, \mathrm{M}_i]$ 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 persistently exciting in the sense of Definition 6, by Claim 3 the convergence (53) holds uniformly in the initial time. By the properties of $\Omega$, this implies that $\Omega \subset \Xi \times \mathcal{A}_\epsilon$, and the projection

$$\mathcal{A}_\epsilon^u := \Big\{ x \in \mathbb{R}^N \mid (\xi, x) \in \Omega \Big\}$$

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 (55) satisfying $\operatorname{dist}\big(x^0,\ \mathcal{A}_\epsilon^u\big) \leq b(\ell)$ also satisfies $\operatorname{dist}\big(x^t,\ \mathcal{A}_\epsilon^u\big) \leq \ell$ for all $t \in \mathbb{N}$. This, in turn, can be proved by similar arguments of [9, 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 $\operatorname{dist}\big(x_m^0,\ \mathcal{A}_\epsilon^u\big) \leq 2^{-m}$ and $\operatorname{dist}\big(x_m^{\tau_m},\ \mathcal{A}_\epsilon^u\big) > \ell$. This, in turn implies

$$\operatorname{dist}\big(z_m^{\tau_m},\ \Omega\big) > \ell. \tag{56}$$

Since $X_0 := \{x \in \mathbb{R}^N \mid \operatorname{dist}(x,\ \mathcal{A}_\epsilon^u) \leq 1\}$ 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 $\Omega$, 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 $\{0, \dots, \bar{\tau}\}$, which satisfies $\lim_{m\to\infty} \operatorname{dist}\big(z_m^0,\ \Omega\big) = 0$. As $\phi$ is continuous, $Z$ is closed, and since $\Omega$ is forward invariant, then in view of [9, 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 $\lim_{m\to\infty} \operatorname{dist}\big(z_m^t,\ \Omega\big) = 0$ for all $t \in \{0, \dots, \bar{\tau}\}$. This, however, contradicts (56) 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 (49)-(51). ■

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

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

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

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

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

$$\begin{aligned} \Delta = (1-m)\Delta + m\Delta &< (1-m)\Delta + r_i + (n+1)T_i - t_0 \\ &< T_i + 1, \end{aligned}$$

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 $m \leq \log(\mathrm{M}^\star/\mu)/\underline{h}$) and $i \in \mathcal{N}$, there exists $s_i \in \{t_0 + 1 + (m-1)\Delta,\ \dots,\ t_0 + m\Delta\}$ 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

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

for all $t \geq t_1^\star$.

As $\lambda^t \leq \nu < 1$ for all $t \geq t_0$, by iterating (30), for $t > t_1^\star$, we obtain

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

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

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

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

$$|x^t| \leq \frac{\bar{y}}{1-\bar{\lambda}} + 2\epsilon + p(\epsilon)$$

for all $t \geq t_2^\star$, and the claim follows by arbitrariness of $\epsilon$. ■

## References


[1] G. Notarstefano, I. Notarnicola, A. Camisa, Distributed optimization for smart cyber-physical networks, Found. Trends Sys. Contr. 7 (2019) 253–383.

[2] F. Bullo, J. Cortes, S. Martinez, Distributed control of robotic networks: a mathematical approach to motion coordination algorithms, Princeton University Press, 2009.

[3] N. Santoro, Design and Analysis of Distributed Algorithms, John Wiley & Sons, 2006.

[4] B. M. Nejad, S. A. Attia, J. Raisch, Max-consensus in a maxplus algebraic setting: The case of fixed communication topologies, in: 2009 XXII Int. Symposium on Information, Communication and Automation Technologies, 2009, pp. 1–7.

[5] F. Iutzeler, P. Ciblat, J. Jakubowicz, Analysis of Max-Consensus Algorithms in Wireless Channels, IEEE Trans. Sig. Proc. 60 (11) (2012) 6103–6107.

[6] A. Golfar, J. Ghaisari, Convergence analysis of max-consensus algorithm in probabilistic communication networks with Bernoulli dropouts, Int. J. Sys. Sci. 50 (7) (2019) 1313–1326.

[7] K. B. Ariyur, M. Krstić, Real-Time Optimization by Extremum-Seeking Control, Wiley-Blackwell, 2003.

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

[9] R. Goebel, R. G. Sanfelice, A. R. Teel, Hybrid Dynamical Systems. Modeling, Stability, and Robustness, Princeton University Press, 2012.

[10] R. G. Gallager, P. A. Humblet, P. M. Spira, A distributed algorithm for minimum-weight spanning trees, ACM Trans. Program. Lang. syst. 5 (1983) 66–77.

[11] A. Nedic, A. Ozdaglar, Distributed Subgradient Methods for Multi-Agent Optimization, IEEE Trans. Autom. Control 54 (1) (2009) 48–61.

[12] A. Nedic, A. Ozdaglar, P. A. Parrilo, Constrained Consensus and Optimization in Multi-Agent Networks, IEEE Transactions on Automatic Control 55 (4) (2010) 922–938.

[13] I. Lobel, A. Ozdaglar, Distributed Subgradient Methods for Convex Optimization Over Random Networks, IEEE Transactions on Automatic Control 56 (6) (2011) 1291–1306.

[14] W. Shi, Q. Ling, G. Wu, W. Yin, EXTRA: An Exact First-Order Algorithm for Decentralized Consensus Optimization, SIAM J. Optimization 25 (2) (2015) 944–966.

[15] W. Shi, Q. Ling, G. Wu, W. Yin, A Proximal Gradient Algorithm for Decentralized Composite Optimization, IEEE Transactions on Signal Processing 63 (22) (2015) 6013–6023.

[16] K. Yuan, Q. Ling, W. Yin, On the Convergence of Decentralized Gradient Descent, SIAM J. Optimization 26 (3) (2016) 1835–1854.

[17] D. Varagnolo, F. Zanella, A. Cenedese, G. Pillonetto, L. Schenato, Newton-Raphson Consensus for Distributed Convex Optimization, IEEE Trans. Autom. Control 61 (4) (2016) 994–1009.

[18] A. Mokhtari, Q. Ling, A. Ribeiro, Network Newton Distributed Optimization Methods, IEEE Trans. Sig. Proc. 65 (1) (2017) 146–161.

[19] P. Xie, K. You, R. Tempo, S. Song, C. Wu, Distributed Convex Optimization with Inequality Constraints over Time-Varying Unbalanced Digraphs, IEEE Trans. Autom. Control 63 (12) (2018) 4331–4337.

[20] M. Zhu, S. Martinez, On Distributed Convex Optimization Under Inequality and Equality Constraints, IEEE Trans. Autom. Control 57 (1) (2012) 151–164.

[21] T.-H. Chang, A. Nedić, A. Scaglione, Distributed Constrained Optimization by Consensus-Based Primal-Dual Perturbation Method, IEEE Trans. Autom. Control 59 (6) (2014) 1524–1538.

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

[23] J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar, M. Püschel, D-ADMM: A Communication-Efficient Distributed Algorithm for Separable Optimization, IEEE Trans. Sig. Proc. 61 (10) (2013) 2718–2723.

[24] W. Shi, Q. Ling, K. Yuan, G. Wu, W. Yin, On the Linear Convergence of the ADMM in Decentralized Consensus Optimization, IEEE Trans. Sig. Proc. 62 (7) (2014) 1750–1761.

[25] D. Jakovetić, J. M. F. Moura, J. Xavier, Linear Convergence Rate of a Class of Distributed Augmented Lagrangian Algorithms, IEEE Trans. Autom. Control 60 (4) (2015) 922–936.

[26] Q. Ling, W. Shi, G. Wu, A. Ribeiro, DLM: Decentralized Linearized Alternating Direction Method of Multipliers, IEEE Trans. Sig. Proc. 63 (15) (2015) 4051–4064.

[27] T.-H. Chang, A Proximal Dual Consensus ADMM Method for Multi-Agent Constrained Optimization, IEEE Trans. Sig. Proc. 64 (14) (2016) 3719–3734.

[28] A. Makhdoumi, A. Ozdaglar, Convergence Rate of Distributed ADMM Over Networks, IEEE Trans. Autom. Control 62 (10) (2017) 5082–5095.

[29] N. Bastianello, R. Carli, L. Schenato, M. Todescato, Asynchronous Distributed Optimization over Lossy Networks via Relaxed ADMM: Stability and Linear Convergence, IEEE Trans. Autom. ControlEarly Access (2020). doi:10.1109/TAC.2020.3011358.

[30] J. Xu, S. Zhu, Y. C. Soh, L. Xie, Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes, in: 54th IEEE Conf. Decision and Control (CDC), 2015, pp. 2055–2060.

[31] A. Nedić, A. Olshevsky, W. Shi, Achieving Geometric Convergence for Distributed Optimization Over Time-Varying Graphs, SIAM J. Optimization 27 (4) (2017) 2597–2633.

[32] A. Nedić, A. Olshevsky, W. Shi, C. A. Uribe, Geometrically convergent distributed optimization with uncoordinated stepsizes, in: 2017 American Control Conference (ACC), 2017, pp. 3950–3955.

[33] G. Qu, N. Li, Harnessing Smoothness to Accelerate Distributed Optimization, IEEE Trans. Contr. Net. Sys. 5 (3) (2018) 1245–1260.

[34] C. Xi, R. Xin, U. A. Khan, ADD-OPT: Accelerated Distributed Directed Optimization, IEEE Trans. Autom. Control 63 (5) (2018) 1329–1339.

[35] M. Bin, I. Notarnicola, L. Marconi, G. Notarstefano, A system theoretical perspective to gradient-tracking algorithms for distributed quadratic optimization, in: IEEE 58th Conf. Decision and Control (CDC), 2019, pp. 2994–2999.

[36] P. Kokotović, H. K. Khalil, J. O'Reilly, Singular Perturbation Methods in Control. Analysis and Design, SIAM, 1999.

[37] A. R. Teel, L. Moreau, D. Nešić, A unified framework for input-to-state stability in systems with two time scales, IEEE Trans. Autom. Control 48 (2003) 1526–1544.

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

[39] H. Robbins, S. Monro, A stochastic approximation method, The Annals of Mathematical Statistics 22 (3) (1951) 400–407.

[40] H. J. Kushner, G. G. Yin, Stochastic Approximation Algorithms and Applications, Springer Verlag, 1997.

[41] A. Simonetto, H. Jamali-Rad, Primal recovery from consensus-based dual decomposition for distributed convex optimization, J. Optim. Theory Appl. 168 (2016) 172–197.