Università degli Studi di Padova

Facoltà di Ingegneria

Corso di Laurea in Ingegneria dell'Informazione

Conditions for Convergence in Consensus:

an Analysis of Limit Cases

Author

Stefano Patron

Matr. number

591285

Supervisor

Dr. Francesco Ticozzi

A.Y. **2010/2011**

Abstract

In this thesis, we first review and discuss the problem of convergence to consensus for a model that represents a network of individual agents interacting through time-dependent communication links. This model properly describes many situations in cooperative control, where real-world communication topologies are usually not fully connected and are dinamically changing over time.

The convergence to consensus problem, namely the search of conditions that ensure the agents will reach "agreement" on a common state, is presented emphasizing the two major aspects on which constraints must be imposed: the individual behavior and the global network connectivity.

We next analyze two approaches to the study of convergence to consensus: leaving out the most technical parts of the proofs, we focus on the fundamental results and differences about them. The first one is based upon set-valued Lyapunov theory, with a notion of (local) convexity playing a central role, while the second one relies on a theorem by Birkhoff on the contraction properties for positive maps on cones.

Finally, we proceed to a thorough comparison of these approaches, highlighting their similarities, differences and applicability to limit cases with the aid of numerous simple examples.

Contents

1

Introduction to Consensus

1.1 Practical Applications

The present work can be collocated in the field of cooperative control, which studies dynamical systems composed of many different individuals (*agents*) that can communicate with each other, respecting some constraints.
In this framework, global results about the multi-agent system are closely linked to the individual behavior and to particular communication features.

In particular, we focus on the problem of proving the convergence to consensus, i.e.

determining a set of conditions that allow the system agents to reach, at least asymptotically, a common state.

This can be interpreted as the single systems reaching a common point of interest, or agreeing on some quantitative statement. Intuitively, we can expect these conditions to regard the communication features on the one hand, and the individual dynamical behavior on the other.

Consensus problems arise in many type of systems: in particular it is of key interest in understanding how natural groups coordinate themselves, like in swarming, a cooperative behaviors observed for a variety of living beings such as birds, fish or bacteria.
Of course, it could be hard to determine exact mathematical models to describe those dynamics, however the research work about it can be useful to find methods which can be applied in control systems. Another interesting field regards human behavior: in a theoretical way, if people opinions could be synthesized in a certain "state", it would be possible to apply the consensus theory in order to study those dynamics which allow to achieve

unanimous opinions and common point of views.
Even if this could appear as science fiction, because of the impossibility of containing human behaviors in mathematical terms, it is still conceivable to find simplified quantitative models that could illustrate the basic mechanisms causing aggregated behaviors to emerge, e.g. in economic or social sciences.

We refer to [4] for a thorough discussion of the key aspects of the problem from a control-theoretic perspective. Cooperative control of multiple-vehicle systems has potential impact in numerous civilian, homeland security, and military applications where communication bandwidth and power constraints will preclude centralized command and control. The interest is typically in information consensus, where a team of vehicles must communicate with its neighbors in order to agree on key pieces of information that enable them to work together in a coordinated fashion. The problem is particularly challenging because real-world communication topologies are usually not fully connected: in many cases, they depend on the relative positions of the vehicles and on other environmental factors and are therefore dynamically changing in time. In addition, wireless communication channels are subject to multipath, fading and drop-out. Such principles can be used in a robotic context to help enable a large group of autonomously functioning vehicles in the air (UAV), on land or sea or underwater, to collectively accomplish, in a safe and coordinated manner, useful tasks such as distributed, adaptive scientific data gathering, search and rescue, and reconnaissance.

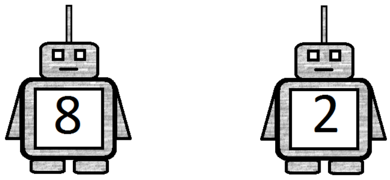
Consensus is reached when the system achieves agreement among agents. It is customary to associate to the system a communication network, based on the real communication links existing between the agents.
To achieve consensus, there must be a shared variable of interest, called the information state, as well as appropriate algorithmic methods for negotiating to reach consensus on the value of that variable, called the consensus algorithms; agents update the value of their information states based on the information states of their neighbors, i.e. the agents which can communicate with them.

The topology of the communication network can change continuously, and algorithms need to be robust against those changes, because the information states of all the agents in the network converge to a common value.

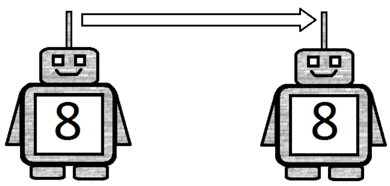
1.2 An Introductory Example

We start introducing the object of our analysis with an informal example, before proceeding with a formal mathematical description.
Let us suppose we have two programmable robots, and each of them can store a real number; let us take for example the initial values 8 and 2.

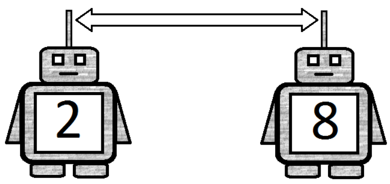
Let us suppose we can program the robots at the moment they are created, and then they will act autonomously, without any supervision. *We want the two robots to reach autonomously the same number, indipendently from the initial values.*



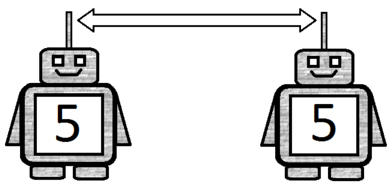
Of course, sharing information is a necessary step for this cooperation, and so the robots should be able to comunicate. For example, provided a robot can broadcast its value, it is sufficient program them saying: "*when you receive a number, store it*", and so the target will be reached.



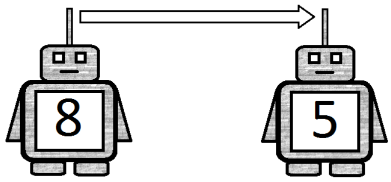
However, in general we may have no knowledge about which robot will transmit; if the second robot is able to broadcast its value too, the effect will be a *number exchanging* and the target will not be reached.



Let us change our strategy, and program the robots saying: "*when you receive a number, compute the average between that and the number you have in memory, and store the result*". With this algorithm the target will be reached.

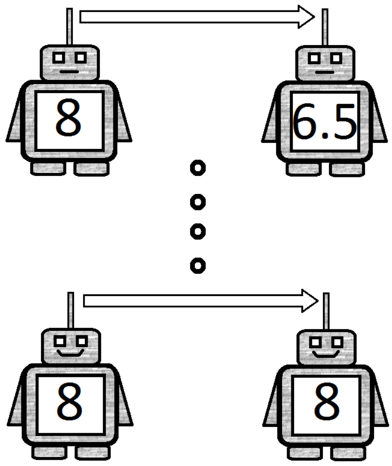


We could object that returning to the unidirectional transmition consensus will not be reached using the mean method: $(8+2)/2 = 5$



But let us observe that if the connection link is stable in time, and if the robot continues to compute the mean...

$(8+2)/2 = 5$
$(8+5)/2 = 6.5$
$(8+6.5)/2 = 7.25$
$\vdots$
$\vdots$
$\vdots$
... agreement will be reached asymptotically.



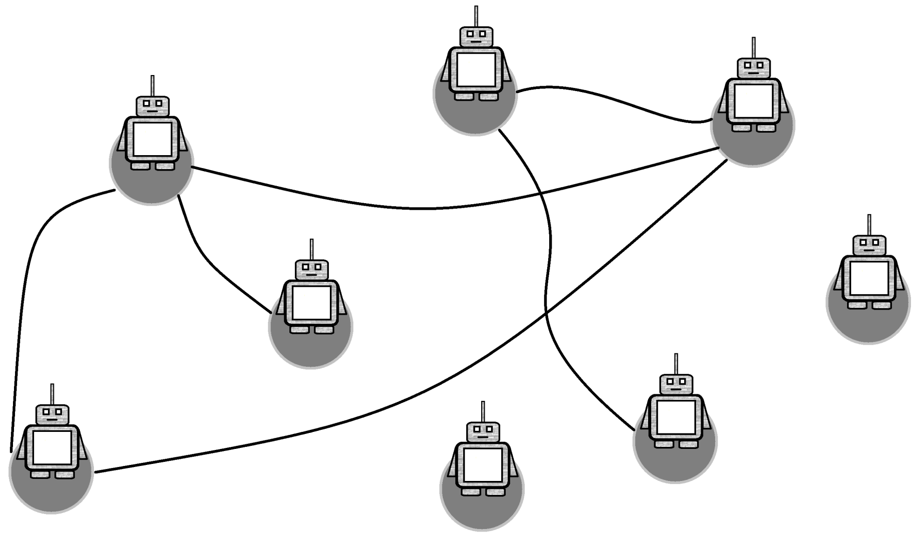
The second method ("compute the mean") appears more general than the first one ("copy the number").

In this simple example, the **consensus problem** can be traduced with: *make the robots reach the same number*, or *the same "state"*. We are not interested in how much time will be employed: we are looking for criteria that permit to verificate sufficient conditions to consensus.

From this first example, we observe the dipendence from:
1) communication settings
2) consensus algorithm
and so the criteria we are looking for will be based upon these aspects.

Let us try to generalize the problem: let us suppose we have n robots, and each of them can store a number. Let us suppose we have no information about the evolution of the connection links between these robots in time: the links can modificate time by time. In other words, if in instant t there is a certain link between robots i and j, it is not necessarily true that this link will exist in instant $t+1$.

Which hypotesis, about *communication settings* and *consensus algorithm*, consensus will be reached under?



In order to model this problem, we set the n robots on the nodes of a **graph**: *the i-th node corresponds to the i-th robot.* Then for each instant we *associate an arc of the graph to each communication link*; having supposed time-varying links, the arc set will be time-varying too, and so we will have a *sequence of arc sets.*

In this way, our model is a **sequence of graphs**:

$$\mathcal{G}(t) = (\mathcal{N}, \mathcal{A}(t))$$

$\mathcal{N} = \{1, 2, ..., n\}$ the set of nodes/robots
$\mathcal{A}(t) \subseteq \mathcal{N} \times \mathcal{N}$ the set of arcs/links, $t \in \mathbb{N}$

Henceforward, we will call the robots **agents**, and we will generalize the problem saying that they have a **state** that is not necessarily a number, but an element of an Euclidean space X.

Our goal is to study consensus criteria for **multiagent systems with time-dependent communication links**.

1.3 The Multiagent System

The starting point is a system of n interacting agents; this is a dynamic model in state form with state space X, that is Euclidean and finite-dimensional

$x_1(t+1) = f_1(t, x_1(t), x_2(t), ..., x_n(t))$
$x_2(t+1) = f_2(t, x_1(t), x_2(t), ..., x_n(t))$
...
$x_n(t+1) = f_n(t, x_1(t), x_2(t), ..., x_n(t))$

In a more compact form, we define the **discrete-time system** on X^n:

$$x(t+1) = f(t, x(t)) \tag{1.1}$$

where $x(t) = [x_1(t), x_2(t), ..., x_n(t)]^T \in X^n$ is the *state vector* of the n agents, and $f : \mathbb{N} \times X^n \to X^n$ is the continuous *update function*.

This formulation is very general, and does not consider the constraints imposed by the agents' communication links. The communication topology can be represented each instant t by a *communication graph*, where the nodes correspond to the agents, and where the arc (i, j) exists whenever there is a communication link from agent i to agent j.
We suppose time-dependent communications, therefore the communication graph changes over time.
Considering these aspects, we can introduce the **sequence of graphs**

$$\mathcal{G}(t) = (\mathcal{N}, \mathcal{A}(t))$$

where every node in $\mathcal{N} = \{1, 2, ..., n\}$ corresponds to an agent and each arc $(i, j) \in \mathcal{A}(t) \subseteq \mathcal{N} \times \mathcal{N}$ represents a communication channel, from agent i to agent j, existing in a given instant $t \in \mathbb{N}$.
The communication graph $(\mathcal{N}, \mathcal{A}(t))$ determines what information is available for which agent in instant t.

Definition 1 (Neighbors) *Consider a node* $i \in \mathcal{N}$*; we define* $Neighbors(i, t)$ *as the* set of those nodes $j \in \mathcal{N}$ for which $(j, i) \in \mathcal{A}(t)$, with $t \in \mathbb{N}$.
In the consensus context, $Neighbors(i, t)$ *is the set of agents which could influence* i *(because of the existence of a link starting from them).*

It is important to emphasize the difference between *directional* and *bidirectional* graphs:
if the graphs are considered **symmetric**, i.e. $(i, j) \in \mathcal{A}(t) \Leftrightarrow (j, i) \in \mathcal{A}(t)$, then when i is a neighbor of j, j is a neighbor of i;
if the graphs are considered **asymmetric**, then when i is a neighbor of j, j is not necessarily a neighbor of i.

Furthermore, using the definition about *neighbors*, the time-dependence of the links means that $Neighbors(i, t)$ can change over time.

Moreover, in order to give a connection between the system 1.1 and the graph sequence $\mathcal{G}(t)$, we informally detail the update function components above as:

$x_1(t+1) = f_1(t, \{x_j(t) : j \in Neighbors(1,t)\})$
$x_2(t+1) = f_2(t, \{x_j(t) : j \in Neighbors(2,t)\})$
...
$x_n(t+1) = f_n(t, \{x_j(t) : j \in Neighbors(n,t)\})$

It means that *agents update the value of their states based on the states of their neighbors.*

1.4 Asymptotic Consensus and Fundamental Assumptions

When multiple agents agree on the value of a variable of interest (the *state*), they are said to have reached **consensus**.
Considering the vector state $[x_1(t), x_2(t), ..., x_n(t)]^T \in X^n$, consensus will be reached whenever

$$[x_1(t), x_2(t), ..., x_n(t)]^T \rightarrow [\alpha, \alpha, ..., \alpha]^T \quad with \quad \alpha \in X$$

Definition 2 (Asymptotic Consensus) *The update function f guarantees* asymptotic consensus *in system 1.1 if for every initial vector state $x(0) \in X^n$ and for every associated sequence of graphs $\mathcal{G}(t)$, there exists some $\alpha \in X$ such that*

$$\lim_{t \to \infty} x_i(t) = \alpha \quad \forall i \in \mathcal{N}$$

To achieve consensus, there must be an appropriate algorithmic method, called *consensus algorithm.* Typically, in order to model situations of interest, consensus algorithms are designed to be distributed, assuming only neighbor-to-neighbor interaction between agents. The goal is to design an update law so that the states of all the agents in the network converge to a common value.

Another notion is necessary for the following definitions:

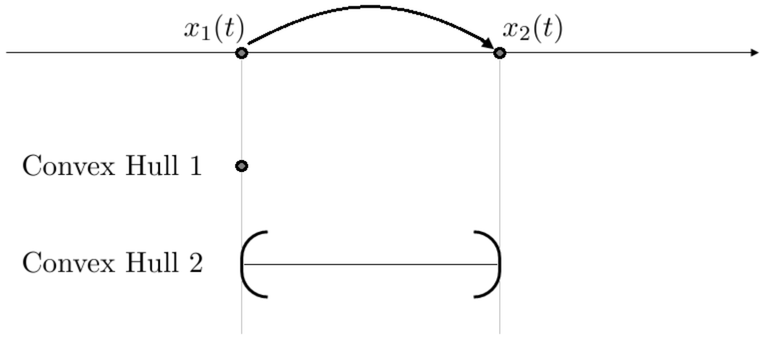
Definition 3 (Convex Hull) *In an Euclidean space X, a set is* convex *if for every pair of points within the set, every point on the straight line segment that joins them is also within the set.*
Given a finite set of points $P = \{\alpha_1, \alpha_2, ...\alpha_n\}$ of X^n, the **convex hull** *of P is the smallest convex set containing P, namely the intersection of all convex sets containing P.*

Let us anticipate that there are *three fundamental assumptions* necessary to reach consensus in system 1.1; many variations of them can be found in the literature: we choose the following formulation.

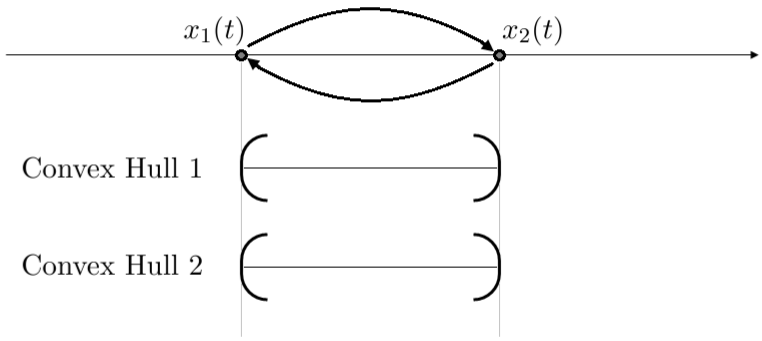
Local Convexity Assumption The next state of i, $x_i(t+1)$, is strictly contained in the *convex hull* of the states of agent i at time t and of agents in $Neighbors(i,t)$.
Let us emphasize with the following example the local nature of this property, showing that the convex hull can take different forms considering symmetric or asymmetric graphs.

▷ **Example** Let us consider a system composed of two agents; each of them has a certain state in the real line.

- Let us suppose $\mathcal{A}(t) = \{(1,2)\}$, namely the graph is *directional*. Considering agent 1, $Neighbors(1,t) = \{1\}$ and so the relative convex hull is $x_1(t)$; considering agent 2, $Neighbors(2,t) = \{1,2\}$, and so the relative convex hull is the set $(x_1(t), x_2(t))$.



- Let us suppose $\mathcal{A}(t) = \{(1,2);(2,1)\}$, namely the graph is *bidirectional.*
  In this case, $Neighbors(1,t) = Neighbors(2,t) = \{1,2\}$ and so the relative convex hull is the set $(x_1(t), x_2(t))$ for both agents.



Local convexity represents a "good behavior" of each individual agent, which updates its state approaching the states of its neighbors (whenever it has some neighbors).

Global Asymptotic Connectivity In order to avoid a lack of communication, it is necessary to suppose that the graph $\mathcal{G}_{(t,\infty)} = (\mathcal{N}, \bigcup_{\tau \geq t} \mathcal{A}(\tau))$ is *connected* $\forall t \in \mathbb{N}$, which means that exists a node in $\mathcal{N}$ that can reach every other nodes of $\mathcal{N}$ through paths of $\mathcal{G}_{(t,\infty)}$.

Bounded Intercommunication Intervals If i communicates to j an infinite number of times, then there is some $T \geq 0$ such that $(i, j) \in \mathcal{A}(t) \cup \mathcal{A}(t+1) \cup ... \cup \mathcal{A}(t+T-1)$ $\forall t \in \mathbb{N}$.
Although the previous assumption assures the necessary communication, the dynamics of the sequence of graphs could avoid consensus yet. Let us suppose that the global connection switches on infinite times with increasing size intervals between them; asymptotically, it is similar to absence of connection. In order to avoid this situation, it is necessary to force the connectivity to repeat itself in finite intervals. Let us show that with the following formal example.

▷ **Example** The agreement algorithm can generate nonconvergent dynamics when the communication intervals become unbounded. Let us consider a real-values system with three agents, with initially vector state $x(0) = (0, 1, 1) \in \mathbb{R}^3$.
Let us consider the sequence of arc sets
$\mathcal{A}_a = \{(1, 2)\}$
$\mathcal{A}_b = \{(1, 2), (2, 1)\}$
$\mathcal{A}_c = \{(3, 2)\}$
$\mathcal{A}_d = \{(2, 3), (3, 2)\}$

Defining the concatenation

$$B_k = \underbrace{\mathcal{A}_a, ..., \mathcal{A}_a}_{\text{2k times}}, \mathcal{A}_b, \underbrace{\mathcal{A}_c, ..., \mathcal{A}_c}_{\text{2k+1 times}}, \mathcal{A}_d \qquad k \in \{0\} \cup \mathbb{N}$$

let us consider as arc sets sequence $(\mathcal{A}(t))$ for the system 1.1 the concatenation

$$B_0, B_1, B_2, B_3, ...$$

and let us consider the sequence of time instants $t_1, t_2, t_3, ... \to \infty$ determined by

$$\begin{cases} t_{p+1} - t_p = p + 1 \quad \forall p > 1 \\ t_1 = 2 \end{cases}$$

Let us observe that the union of the arc sets $\mathcal{A}(t)$ over any interval of the form $[t_0, \infty)$ is given by $\mathcal{N} \times \mathcal{N}$, and then the assumption about global asymptotic connectivity is respected.

Let us suppose the system uses the consensus algorithm informally explained in the Introductory Example 1.2 (*"do the mean"*); it is easy to prove that the local convexity assumption is respected too (see Subsection 2.2).

Instead, the assumption about bounded intercommunication intervals is not respected: it is impossible to find such a $T \geq 0$, because of the increasing of intervals $t_{p+1} - t_p$.

Let us show that the three components of x do not converge to a common value when $t \to \infty$.
In order to show that, let us evaluate the *difference* $d(p) = x_3(t_p) - x_1(t_p)$ *at the time-instants* $t_1, t_2, t_3, ... \to \infty$. It can be proven that:

$$\begin{cases} d(p+1) = d(p) - \frac{1}{2^{p+1}} d(p) = \frac{2^{p+1}-1}{2^{p+1}} d(p) \quad \forall p > 1 \\ d(1) = \frac{1}{2} \end{cases}$$

Thus, $0 < d(p) < 1$ and $d(p)$ is decreasing with p (convergence as $p \to \infty$). The *total accumulative decrease of d satisfies*

$$\sum_{p=1}^{\infty} (d(p) - d(p+1)) = \sum_{1=1}^{\infty} \frac{1}{2^{p+1}} d(p) < \sum_{p=1}^{\infty} \frac{1}{2^{p+1}} = \frac{1}{2}$$

Since $d(1) = \frac{1}{2}$, we conclude that

$$\lim_{p \to \infty} \sup(x_3(t_p) - x_1(t_p)) \geq \lim_{p \to \infty} d(p) = d(1) - \sum_{p=1}^{\infty} (d(p) - d(p+1)) > 0$$

On the other hand, if there is *simmetry*, the bounded intercommunication interval assumption is unnecessary and can be relaxed.

1.5 The Linear Case

Let us define a particular case regarding consensus, where each agent i has a **scalar value** $x_i(t) \in \mathbb{R}$ as state, $\forall t \in \mathbb{N}$, so that the state vector satisfies $[x_1(t), x_2(t), ..., x_n(t)]^T \in \mathbb{R}^n \ \forall t \in \mathbb{N}$.
In the linear case of the consensus problem, considering again system 1.1, the consensus algorithm updates the state vector in this way:

$$x(t+1) = W(t)x(t) \tag{1.2}$$

where $x(t) = [x_1(t), x_2(t), ..., x_n(t)]^T \in \mathbb{R}^n$ is the *state vector* of the n agents and $W(t)$ is an $n \times n$ real matrix.
In detail,

$$x_i(t+1) = \sum_{k=1}^{n} w_{ik}(t) x_k(t)$$

where $w_{ij}(t)$ are entries of $W(t)$.

For compatibility motivation with the previous graph interpretation, whenever $w_{ij}(t) \neq 0$ agent j communicates its current value $x_j(t)$ to agent i, and so $j \in Neighbors(i, t)$; instead, whenever $w_{ij}(t) = 0$ two situations are possible:

- agent j does not communicate its current value $x_j(t)$ to agent i, and so $j \notin Neighbors(i, t)$
- agent j communicates its current value $x_j(t)$ to agent i, and so $j \in Neighbors(i, t)$, but i does not take account about it

This treatment corresponds to consider a **sequence of weighted graphs**

$$\mathcal{G}_w(t) = (\mathcal{N}, \mathcal{A}(t), W(t))$$

2

Two Different Approaches to Consensus

2.1 Approach Based on Set-Valued Lyapunov Theory

The first approach we consider is based on [1]. This paper is centered around the notion of convexity, and gives a stability analysis based upon the existence of a Lyapunov set-valued function.

In the first subsection we will give notions and theory fundaments of set-valued Lyapunov characterization; in the second subsection we will see how to use Lyapunov Theory in consensus.

2.1.1 Set-Valued Lyapunov Theory

We consider the discrete-time system (1.1) $x(t+1) = f(t, x(t))$ introduced in the previous chapter, with $f : \mathbb{N} \times X^n \to X^n$ a continuous map and X a finite-dimensional Euclidean space. We refer to Chapter 1 for an introduction and for examples to the systems that this class of models can describe.

We are interesting in studying the agents' states converging to a common constant value, and we expect this value to depend on the initial states: we are thus considering a *continuum of equilibrium points.*

Hence, a stability theory focusing only on isolated equilibria is not suitable for our setting, and the *stability of the set of equilibria* should be considered instead. Beside the need for stability of sets of equilibria, there is another aspect worth emphasizing: while these techniques may be used to assert that the individual agents' converge towards a common value, this value is not guaranteed to be constant in time.

▷ **Example** Consider the discrete-time real system

$$x_1(t) = \frac{x_1(t)}{\sqrt{1 + x_1^2(t)}} \qquad x_2(t) = x_1(t) + x_2(t)$$

There is a continuum of equilibrium points $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 = 0\}$ and this invariant set is stable and attracts the system solutions. Nevertheless, the solution of this system starting in $(x_1, x_2) = (1, 0)$ is given by

$$x_1(t) = \frac{1}{\sqrt{t}} \qquad x_2(t) = \sum_{s=1}^{t-1} \frac{1}{\sqrt{s}} \quad \forall t > 1$$

The second component diverges to infinity as $t \to \infty$, and this is not the convergence property we are aiming at.

In order to exclude situations with a drift along the set of equilibrium points we need to give a suitable definition of stability and attractivity.

Consider a continuous map

$$f : \mathbb{N} \times \mathcal{X} \to \mathcal{X}$$

with $\mathcal{X}$ a finite-dimensional Euclidean space, and the discrete-time system:

$$x(t+1) = f(t, x(t)) \tag{2.1}$$

Consider now a collection of equilibrium solutions of the system (2.1).

Definition 4 *With respect to a collection of equilibrium solutions* Φ *, the system (2.1) is called* **globally asymptotically stable** *if:*

1) is stable, *namely*
$\forall \phi_1 \in \Phi$, $\forall c_2 > 0$, $\forall t_0 \in \mathbb{N}$, $\exists \quad c_1 > 0$:
$\forall$ *solution* ξ *of (2.1)* *if* $|\xi(t_0) - \phi_1| < c_1$ *then* $\exists \quad \phi_2 \in \Phi : |\xi(t) - \phi_2| < c_2 \quad \forall t \geq t_0$

2) is bounded, *namely*
$\forall \phi_1 \in \Phi$, $\forall c_1 > 0$, $\forall t_0 \in \mathbb{N}$, $\exists \quad c_2 > 0$:
$\forall$ *solution* ξ *of (2.1)* *if* $|\xi(t_0) - \phi_1| < c_1$ *then* $\exists \quad \phi_2 \in \Phi : |\xi(t) - \phi_2| < c_2 \quad \forall t \geq t_0$

3) is globally attractive, *namely*
$\forall \phi_1 \in \Phi$, $\forall c_1, c_2 > 0$, $\forall t_0 \in \mathbb{N}$, $\exists \quad T \geq 0$:
$\forall$ *solution* ξ *of (2.1)* *if* $|\xi(t_0) - \phi_1| < c_1$ *then* $\exists \quad \phi_2 \in \Phi : |\xi(t) - \phi_2| < c_2 \quad \forall t \geq t_0 + T$

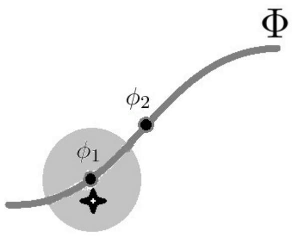
If the number c_1 (respectively c_2 and T) can be chosen *independently of* t_0, than the system is called **uniformly globally asymptotically stable** with respect to the considered collection of equilibrium solutions.
The previous definition may be interpreted as follows:

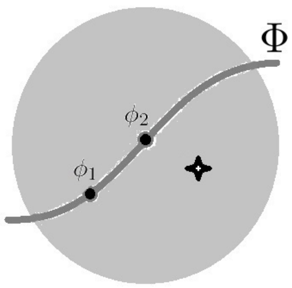
1) stability means that considering an equilibrium ϕ_1 and a certain constant c_2, exists a constant c_1 such that any solution of 2.1 remains at a distance smaller than c_2 from ϕ_2 provided it initially was at a distance smaller than c_1 from ϕ_1.

There is *stability* if there is assurance to remain *arbitrary* close to ϕ_2 starting close *enough* to ϕ_1.

2) boundedness means that considering an equilibrium ϕ_1 and a certain constant c_1, exists a constant c_2 such that any solution of 2.1 which is initially at a distance smaller than c_1 from ϕ_1 will remain at a distance smaller than c_2 from ϕ_2.

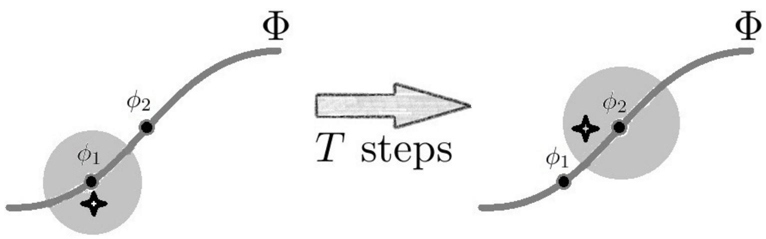
There is *boundedness* if there is assurance to remain close *enough* to ϕ_2 starting *arbitrary* close to ϕ_1.





3) attractivity means that considering an equilibrium ϕ_1 and two certain constants c_1 and c_2, exists a period T such that any solution of 2.1 which is initially at a distance smaller than c_1 from ϕ_1 will necessarilly remain at a distance smaller than c_2 from ϕ_2 after T steps.

There is *attractivity* if there is assurance to remain *arbitrary* close to ϕ_2 starting *arbitrary* close to ϕ_1, after a certain number T of steps.

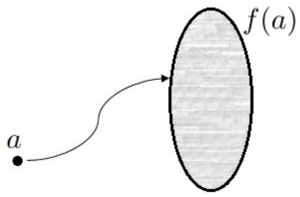


Subsequentially, stability and boundedness require that any solution initially close to Φ remains close to one of the equilibria in Φ, and this excludes the possibility of drift along the set Φ; attractivity implies that every solution converges to one of the equilibria in Φ.

After having precisely defined these notions, we can now enunciate a theorem regarding Lyapunov Characterization, which provides *sufficient conditions for uniform global asymptotic stability* in terms of the existence of a set-valued Lyapunov function.

We need a definition about *set-valued functions*, functions which maps an element of a set A to a subset of the set B.

Definition 5 *A* **set-valued function** $f : A \rightrightarrows B$ *is a map* $f : A \to \mathcal{P}(B)$, *where* $\mathcal{P}(B) = \{\beta \subseteq B\}$ *is the power set of* B, *namely the set of all the subsets made of elements of* B.



Another necessary notion for the theorem is the next:

Definition 6 *Consider two finite-dimensional Euclidean spaces* A *and* B. *A set-valued function* $f : A \rightrightarrows B$ *is called* **upper semicontinuous** *if* $\forall a \in A$ *and* $\forall \epsilon > 0$ $\exists \delta > 0$ *such that* $f(x) \in \mathcal{B}(f(a), \epsilon)$ *whenever* $x \in \mathcal{B}(a, \delta)$.

If K is a subset of a finite-dimensional Euclidean space A, then $\mathcal{B}(K, c)$ $(c > 0)$ is defined as the set of points in A whose distance to K is strictly smaller than c.

Moreover, a set valued function $f : A \rightrightarrows B$ with A and B finite-dimensional Euclidean spaces, is said *bounded* if exists a compact set $\beta \subset B$ such that $f(a) \in \beta \quad \forall a \in A$.

Let us enunciate the theorem:

Theorem 1 (Lyapunov Characterization) *Consider the system (2.1). If it is* possible to **individuate** an upper semicontinuous **set-valued function** $V : \mathcal{X} \rightrightarrows \mathcal{X}$ *satisfying:*

1) $x \in V(x)$, $\forall x \in \mathcal{X}$

2) $V(f(t, x)) \subseteq V(x)$, $\forall t \in \mathbb{N}$, $x \in \mathcal{X}$

then we have **uniform stability** *if* $V(\phi) = \{\phi\}$ $\forall \phi \in \Phi$
and we have **uniform boundedness** *if* $V(x)$ *is bounded* $\forall x \in \mathcal{X}$.

In addition, if it is possible to individuate a function $\mu : Image(V) \to \mathbb{R}_{\geq 0}$, and a function $\Delta : \mathcal{X} \to \mathbb{R}_{\geq 0}$ satisfying:
3) $\mu \circ V : \mathcal{X} \to \mathbb{R}_{\geq 0}, x \mapsto \mu(V(x))$ is bounded on bounded subsets of $\mathcal{X}$
4) Δ is positive definite with respect to Φ, $\quad$ *($\Delta(\phi) = 0\ \forall \phi \in \Phi$*
$\Delta(x) > 0$ otherwise)
5) $\mu(V(f(t,x))) - \mu(V(x)) \leq -\Delta(x),\ \forall t \in \mathbb{N}, x \in \mathcal{X}$
then we have **uniform global attractivity***, and so the system (2.1) is* **uniformly globally asymptotically stable***.*

Instead of give a mathematical proof of these statements, with the relative heavy notations, let us give an informal interpretation of the functions mentioned, to understand in a better way the meaning of the theorem.

The set-valued function V is the Lyapunov function, which is *decreasing* along the solutions of (2.1).

The function μ serves as a measure for the size of the values of V: it maps $V(x) \subset \mathcal{X}$ to a real positive number, and we can imagine $\mu(V(x))$ as the *diameter* of the set $V(x)$.

The function Δ characterizes the *decrease* of V along the solutions of (2.1) as measured in terms of μ: the two functions μ and Δ give a more quantitative aspect to the decreasing of V.

Let us consider conditions 1) and 2) and an equilibrium solution ξ such that $\xi(t_0) = x_0 \in \mathcal{X}$.

- From 1), $\xi(t) \in V(\xi(t))$ because $\xi(t) \in \mathcal{X}$.
- From 2), $V(\xi(t)) \subseteq V(x_0)$.
- By upper semicontinuity of V, given an equilibrium ϕ_1:
  - if $V(\phi) = \{\phi\}$, $\forall c_1 > 0 \quad \exists \quad c_2 > 0$ such that we can have $V(\xi(t)) \subseteq \mathcal{B}(\phi_1, c_1) \forall t > t_0$ provided that $V(\xi(t_0)) \subseteq \mathcal{B}(\phi_1, c_2)$;
  - if $V(x)$ is bounded, $\forall c_2 > 0 \quad \exists \quad c_1 > 0$ such that we can have $V(\xi(t)) \subseteq \mathcal{B}(\phi_1, c_1) \forall t > t_0$ provided that $V(\xi(t_0)) \subseteq \mathcal{B}(\phi_1, c_2)$.

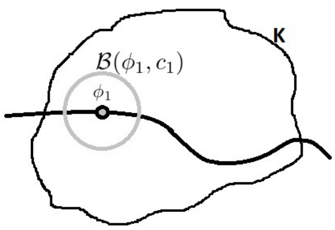
The conclusion in the previous two cases, regarding $V(\phi) = \{\phi\}$ or $V(x)$ *bounded*, is the same:

$$\xi(t) \in V(\xi(t)) \subseteq V(x_0) \subset \mathcal{B}(\phi_1, c_2)$$

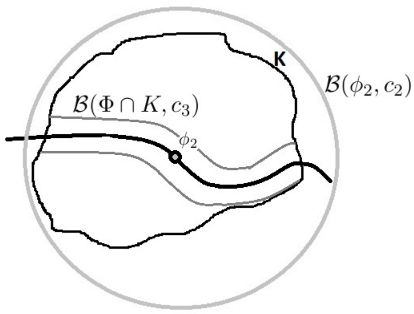
and this means *stability* in the first case (we have proved the existence of c_2) and *boundedness* in the second one (we have proved the existence of c_1).

Let us briefly see the *attractivity* aspect:

- boundedness implies that $\forall \phi_1 \in \Phi$ and $c_1 > 0$, a solution in $\mathcal{B}(\phi_1, c_1)$ remains in a certain compact set $K \supset \mathcal{B}(\phi_1, c_1)$;
- stability implies that $\forall c_2 > 0$ exists $\phi_2 \in \Phi$ and $c_3 > 0$ such that a solution in $\mathcal{B}(\Phi \cap K, c_3)$ remains in $\mathcal{B}(\phi_2, c_2)$.



Boundedness $\Rightarrow$ a solution in $\mathcal{B}(\phi_1, c_1)$ remains in K



Stability $\Rightarrow$ a solutions remains in $\mathcal{B}(\phi_2, c_2)$ provided it starts in $\mathcal{B}(\Phi \cap K, c_3)$

In order to prove attractivity (and so *asymptotic stability*), it can be proved that necessarily from 3) – 5) a solution in K falls in $\mathcal{B}(\Phi \cap K, c_3)$ after T steps: obviously to prove the existence of such a T, the quantitative aspect of the decreasing of V introducted by μ and Δ is fundamental (see Appendix I of [1], for details).

2.1.2 Results

Now, we can connect the consensus system 1.1 with the system 2.1 studied in the previous section just replacing $\mathcal{X}$ with X^n.
We have already seen in Section 1.4 the assumption about *local convexity*. In this context, we can give the next formulation about the system (1.1):

Convexity Assumption: Associated to each graph $(\mathcal{N}, \mathcal{A})$, with $\mathcal{A} \subseteq \mathcal{N} \times \mathcal{N}$, each agent i and each state $x \in X^n$ there is a compact set $e_i(x) \subset X$ satisfying:

1) $f_i(t, x) \in e_i(x) \quad \forall t \in \mathbb{N}, x \in X^n$

2) If the state of agent x_i and its neighbors are all equals, $e_i(x) = \{x_i\}$

3) If the state of agent x_i and its neighbors are not all equals, $e_i(x)$ is contained in the relative interior of the convex hull of the states of agents i and its neighbors

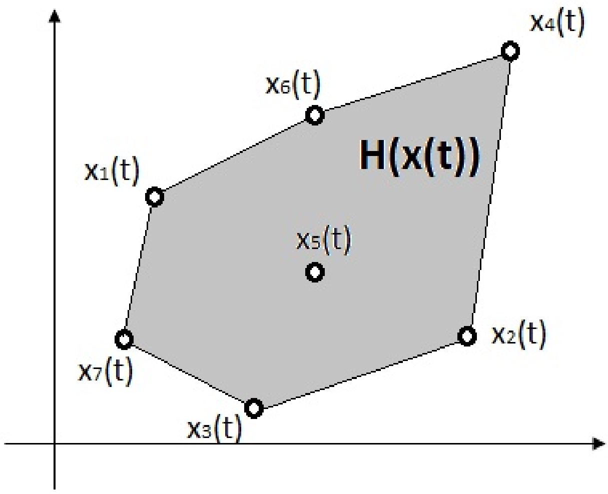
4) The set-valued function $e_i(x) : X^n \to X$ is continuous

The following result is very important:

Lemma: Consider the sequence of graphs $(\mathcal{N}, \mathcal{A}(t))$ associated to the system (1.1), satisfying the previous Convexity Assumption.
The **convex hull** *of the individual agents' states,* $H(x) = conv\{x_1, ..., x_n\} \subset X$*, does not grow along the solution of the discrete-time system (1.1)*:

$$H(f(t,x)) \subseteq H(x) \quad \forall t \in \mathbb{N}, x \in X^n$$

▷ **Example** For example, let X be $\mathbb{R}^2$, that is an Euclidean space. Let us suppose that the system is composed of seven agents, and that in instant t the situation is shown by the next figure:



where the convex hull is the grey area. Necessarily, from Convexity Assumption, at step $t + 1$ the agents' states will still be in the grey area too, and this is a noncreasing situation which suggests to take the convex hull as Lyapunov set-valued function.

It is interesting to observe that *the convex hull of the individual agents' states serves as a measure of disagreement*; the previous Lemma states that, in terms of this measure, the level of disagreement cannot increase with time.

Under the Convexity Assumption, the system is **uniformly stable and bounded** *with* **respect to the collection of equilibrium solutions** $x_1(t) = x_2(t) = ... = x_n(t) = constant$.

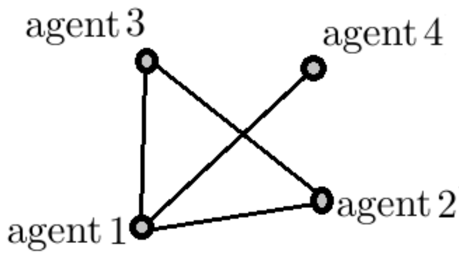
This is a direct consequence of the Lemma and the Lyapunov Characterization Theorem: the convex hull is a decreasing Lyapunov set-valued function whose existence we are looking for, because it respects conditions 1) and 2) of the theorem, and the Convexity Assumption ensures the properties which implies stability and boundedness.

A little technical precisation is in order: in the proof, the Lyapunov set-valued function $V : X^n \rightrightarrows X^n$

$$V(x_1, ..., x_n) = (conv\{x_1, ..., x_n\})^n$$

is considered, instead of the convex hull $H(x) = conv\{x_1, ..., x_n\}$. Nevertheless, the decreasing of V derives from the decreasing of H given by the Lemma.

▷ **Example** Let us suppose $X = \mathbb{R}$. Let us suppose there are four agents, with $x_1(t) = 10$, $x_2(t) = 20$, $x_3(t) = 30$, $x_4(t) = 40$. Let the communication graph topology at $t \in \mathbb{N}$ be:



where the arcs are bidirectionals, and then

$$\mathcal{A}(t) = \{(1,2), (1,3), (1,4), (2,1), (2,3), (3,1), (3,2), (4,1)\}$$

Then, the convex hull of $x(t) = [x_1(t), x_2(t), x_3(t), x_4(t)]$ is
$H(x(t)) = [10, 40] \subset \mathbb{R}$, while $V(x(t)) = [10, 40] \times [10, 30] \times [10, 30] \times [10, 40]$

Evidently, agreement is reached when the disagreement is nulled, namely *the convex hull reduces to a singleton* (that implies a common state value for all agents).

In order for this to happen, additional assumptions need to be imposed ensuring that the convex hull approaches a singleton.

Theorem 2 (Uniform Global Attractivity) *The discrete-time system (1.1), where the sequence of graphs $\mathcal{G}(t)$ has* unidirectional communication*, satisfying the Convexity Assumption is* **uniformly globally attractive** *with* **respect to the collection of equilibrium solutions** $x_1(t) = x_2(t) = \ldots = x_n(t) =$ *constant if and only if*
there is $T \geq 0$ such that $\forall t_0 \in \mathbb{N}$ there is a node connected to all other nodes across $[t_0, t_0 + T]$.

This is the major theorem about asymptotic consensus, a necessary and sufficient condition which not surprisingly involves a connectivity requirement on the sequence of graphs.

The Only-If part is intuitive: if consensus will be reached, we guess the existence of such a node and such a connectivity for all initial instant t_0.
The If part applies the second part of the Lyapunov Theorem, considering the Lyapunov set-valued function $V(x_1, ..., x_n) = (conv\{x_1, ..., x_n\})^n$ and showing that from the existence of such a connectivity node, the functions μ and Δ can be found.

The study of global attractivity is considerably simplified when bidirectional communication is assumed:

Theorem 3 (Global Attractivity - Bidirectional case) *The discrete-time system (1.1), where the sequence of graphs $\mathcal{G}(t)$ has* bidirectional communication*, satisfying the Convexity Assumption is* **globally attractive** *with* **respect to the collection of equilibrium solutions** $x_1(t) = x_2(t) = \ldots = x_n(t) =$ *constant if and only if*
$\forall t_0 \in \mathbb{N}$ there is a node connected to all other nodes across $[t_0, \infty)$.

This specific case, does not require the special kind of "periodicity" about connectivity, and that assumption can be relaxed.

2.2 Asymptotic Consensus in the Linear Case

In Section 1.5 we have introduced the linear system 1.2 as a particular case of the formulation 1.1, where

$$x_i(t+1) = f_i(t, x_1(t), x_2(t), ..., x_n(t)) = w_{i1}(t)x_1(t)+w_{i2}(t)x_2(t)+...+w_{in}(t)x_n(t)$$

The assumption regarding global connectivity and the assumption of bounded intercommunication intervals can be imposed in the same manner in the linear case too, but we need to particularize the assumption regarding local convexity of the consensus algorithm.
Let us first introduce the following algebraic definition

Definition 7 (Row-Stochastic Matrix) *A square-matrix $W(t)$, $n \times n$, with nonnegative entries $w_{ij}(t)$, is said to be* **row-stochastic** *if*

- $w_{ij}(t) \geq 0 \quad \forall i, j \in \mathcal{N}$
- $\sum_{k=1}^{n} w_{ik}(t) = 1 \quad \forall i \in \mathcal{N} \quad \forall t \in \mathbb{N}$

Local Convexity in Linear Case The matrix $W(t)$ associated to the linear system 1.2 is *row-stochastic*, and $w_{ii}(t) \neq 0 \quad \forall i \in \mathcal{N}$ (an agent always keeps some memory of its previous state).
Furthermore, there exist real numbers $\epsilon_{min}, \epsilon_{max}$, $0 < \epsilon_{min} \leq \epsilon_{max} \leq 1$, such that

$$\epsilon_{min} \leq w_{ij}(t) \leq \epsilon_{max} \quad \forall t \in \mathbb{N}, \quad \forall (i, j) \in \mathcal{A}(t)$$

This means that, for each time $t \in \mathbb{N}$, each agent i updates its own value, by forming an average value of its own value and the values of its neighbors, weighted on a certain matrix $W(t)$.
Let us observe that it is implicitly assumed that $(i, i) \in \mathcal{A}(t) \quad \forall i \in \mathcal{N}, t \in \mathbb{N}$

It is very important to emphasize the last condition of the previous assumption: it says that

whenever the arc (j, i) exists, the entry $w_{ij}(t) > 0$;

we can interpret this statement saying that *each agent takes account of its neighbors' states*, moving in the relative interior of the convex hull of them.

This implies that if in a certain time an agent has no neighbors, necessarily it must mantain its own state; in fact it can be considered the only neighbor of itself, so the convex hull of its neighbors' states is its state yet.

Local convexity does not regard the whole system, and the assumption does not imply directly a strictly decreasing of the convex hull of all agents' states; it only assures that whenever there is communication between two agents, they tend to bring closer their states. The next step, about necessary communications, is provided by the other assumptions.

Thus we have the next formulation about the fundamental consensus theorem

Theorem 4 (Consensus Criteria in Linear Case) *Under the* Local Convexity Assumption in the Linear Case *the system 1.2 is guaranteed to reach* asymptotic consensus *if and only if there is* $T \geq 0$ *such that for each* $t_0 \in \mathbb{N}$ *there is a node connected to all other nodes across* $[t_0, t_0 + T]$.

Let us briefly see the correspondence between the Convexity Assumption in the linear case and the Convexity Assumption in the general case: let us consider the agent $i \in \mathcal{N}$, with $t \in \mathbb{N}$

$$x_i(t+1) = \sum_{k=1}^{n} w_{ik}(t) x_k(t) \quad with \quad \sum_{k=1}^{n} w_{ik}(t) = 1$$

Let $r, s \in \mathcal{N}$ be two nodes such that

$$x_r(t) = \min_{j \in \mathcal{N}}(x_j(t)) = m \qquad x_s(t) = \max_{j \in \mathcal{N}}(x_j(t)) = M$$

with $w_{ir}(t), w_{is}(t) \neq 0$, namely $(i, r), (i, s) \in \mathcal{A}(t)$.
The *convex hull* of the states of agent i and its neighbors is the set

$$\mathcal{C}_{\mathcal{H}} = [m, M]$$

Excluding the trivial case $w_{ii}(t) = 1$, we have

$$x_i(t+1) = \sum_{k=1}^{n} w_{ik}(t) x_k(t) < \sum_{k=1}^{n} w_{ik}(t) M = M \sum_{k=1}^{n} w_{ik}(t) = M$$

and

$$x_i(t+1) = \sum_{k=1}^{n} w_{ik}(t) x_k(t) > \sum_{k=1}^{n} w_{ik}(t) m = m \sum_{k=1}^{n} w_{ik}(t) = m$$

Then, it is verified the Convexity Assumption, and

$$x_i(t+1) \in \overline{\mathcal{C}_{\mathcal{H}}}$$

Finally, let us observe that in the symmetric case, *asymptotical consensus will be reached if and only if for all* $t_0 \in \mathbb{N}$ *there is a node connected to all other nodes across* $[t_0, \infty)$.
we can observe that since the graphs are assumed to be bidirectional, the statement "*there is a node connected to all other nodes across* $[t_0, \infty)$" means that all nodes are connected to all other nodes across $[t_0, \infty)$.

2.3 Approach Based on Birkhoff Theorem

This approach comes from [2]; the Lyapunov function is shown to be the Hilbert distance to consensus in log coordinates, and Birkhoff theorem, which proves contraction of the Hilbert metric for any positive homogeneous monotone map, provides us a general convergence result for the consensus algorithms. Let us anticipate how.

Birkhoff theorem provides an important result for homogeneous monotone positive maps defined on closed cones of Banach spaces; the key idea is to use a metric introduced by Hilbert as *contraction measure*: in the positive orthant, Hilberts distance between two vectors x and y is $d_H(x,y) = \max\log(x_i/y_i) - \min\log(x_i/y_i)$, with x_i and y_i components of x and y, and this is invariant by scaling (Hilbert's metric is projective indeed).

We will observe that taking $y = \mathbf{1} = [1,1,1,,1]^T$, $d_H(x,\mathbf{1})$ is a natural *distance to consensus.*

Considering row-stochastic matrices (associated to consensus algorithms, as we saw in Section 2.2) and observing that they define linear positive maps in the positive orthant, we conclude that Birkhoff theorem can be applied and so it gives a convergence result for consensus algorithms.

2.3.1 Birkhoff Theorem

In this section we will see what Birkhoff theorem states.
The application of this theorem falls into *Banach spaces*: these are defined as *complete normed vector spaces.*

We are therefore considering a *vector space* with a *norm*, and an associated distance d and topology induced by that metric; completeness means that *every Cauchy sequence* (with respect to d) *in the space has a limit in the space* (with respect to the topology induced).
Easily, $\mathbb{R}^n$ is a Banach space.

We need the following definition:

Definition 8 *Let $\mathcal{X}$ be a real Banach space, and $\mathcal{K}$ be a closed subset of $\mathcal{X}$. $\mathcal{K}$ is a* **closed solid cone in $\mathcal{X}$** *if:*
1) int$\mathcal{K} \neq \emptyset$;
2) $\mathcal{K} + \mathcal{K} \subset \mathcal{K}$;
3) $\lambda\mathcal{K} \subset \mathcal{K}\ \forall\lambda \geq 0$;
4) $\mathcal{K} \cap (-\mathcal{K}) = \{0\}$.

The partial order induced $x \preceq y$ means $y - x \in \mathcal{K}$.

We also need to define the *Hilbert metric* in $\mathcal{K}_0 = \mathcal{K} - \{0\}$:

Definition 9 *Given $x, y \in \mathcal{K}_0$, define:*

$$M(x, y) = \inf\{\lambda : x - \lambda y \preceq 0\}$$
$$m(x, y) = \sup\{\lambda : x - \lambda y \succeq 0\}$$

Hilbert metric *d is defined in $\mathcal{K}_0$ as*

$$d(x, y) = \log \frac{M(x, y)}{m(x, y)}$$

Hilbert metric is **projective** *because of the invariance property:*

$$d(\alpha x, \beta y) = d(x, y) \qquad \forall \alpha, \beta > 0$$

that is a distance between the equivalence classes $[x] = \{\alpha x : \alpha > 0\}$ and $[y] = \{\beta y : \beta > 0\}$.

In this sense, it is similar to a measure of the angle between the vectors x and y.

The last definition we need to enunciate the Birkhoff theorem is about maps on $\mathcal{K}$:

Definition 10 *A map $A : \mathcal{K} \to \mathcal{K}$ is said to be* non-negative*;*
a map $A : int\mathcal{K} \to int\mathcal{K}$ is said to be positive*.*

Given A positive map, we define:
the **projective diameter** *of A:*

$$\Delta(A) = \sup\{d(A(x), A(y)) : x, y \in int\mathcal{K}\}$$

the **contraction ratio** *of A:*

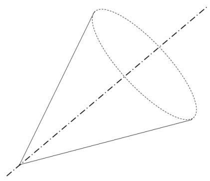
$$k(A) = \inf\{\lambda : d(A(x), A(y)) \leq \lambda d(x, y) \quad \forall x, y \in int\mathcal{K}\}$$

Finally, we can give the theorem:

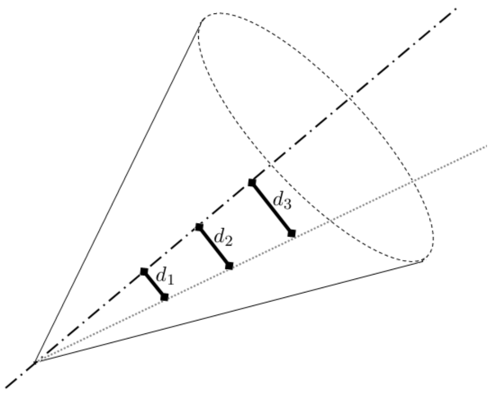
Theorem 5 (Birkhoff Theorem) *Let A be a map in $\mathcal{K}$ satisfying:*
1) A is positive;
2) A is homogeneous of degree p in $int\mathcal{K}$ ($A(\lambda x) = \lambda^p A(x)\ \forall \lambda > 0$);
3) A is monotone ($x \preceq y \Rightarrow A(x) \preceq A(y)$).
Then, the contraction ratio $k(A)$ does not exceed *p.*
If A is linear, $k(A) = \tanh \frac{1}{4}\Delta(A)$.

2.3.2 Graphical Visualizations of Projective Concepts

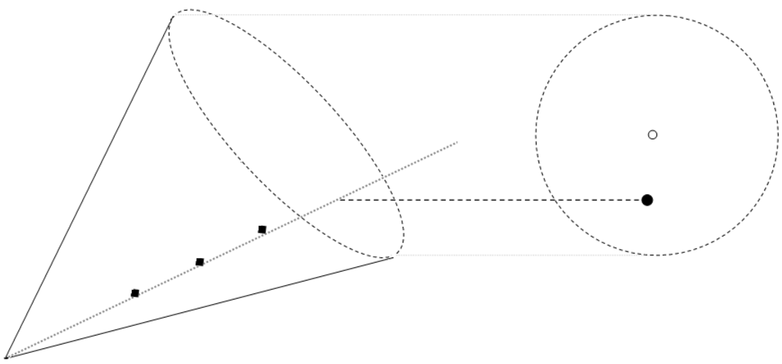
Let us give a graphical representation and interpretation about the notions introduced in the previous subsection.



First of all, we can think a *close solid cone* simply as an "infinite geometric cone", where the vertix corresponds to the zero element. We are looking for a distance from *span*$\{\mathbf{1}\}$, that can be represented as the cone bisector, an axis that corresponds to identity vectors.
Considering the *projective distance*, the distance from the origin does not matter, and only the distance from the bisector matters: it is like to consider the angle; in the figure below, $d_1 \equiv d_2 \equiv d_3$.

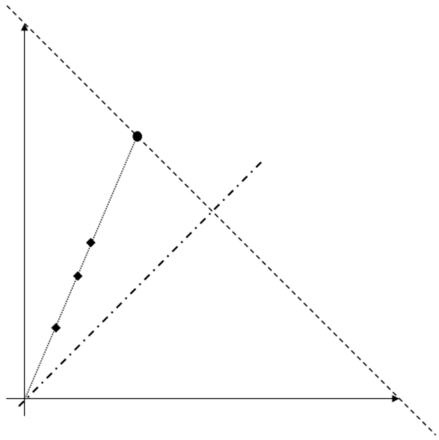


Another pictorial way of interpreting the concept of projective distance is that looking at the (infinite) cone base from an hole applied in the vertex, every ray can be projected to a single point.

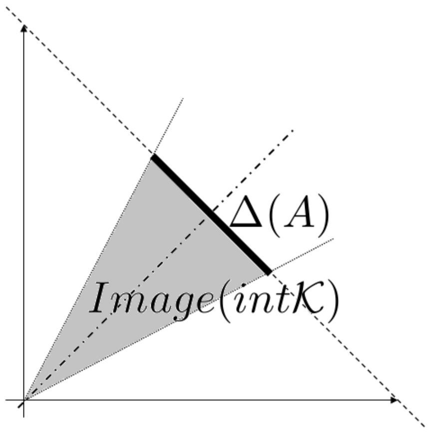


If A is *positive* then it maps the cone interior to the cone interior; the image of A can be figuratively observed from an hole in the vertex too. In this sense, the *projective diameter* $\Delta(A)$ of A is the size of the sphere in which the infinite base of the cone interior is mapped, while the *contraction ratio* k of A is the axis which allows the smaller contraction of the sphere.

In order to better illustrate this idea, let us consider the positive quadrant in the plane. In this case, $int\mathcal{K} = \{(x,y) \in {\mathbb{R}_{>0}}^2\}$, and the vectors are projected in the interior of the real line.



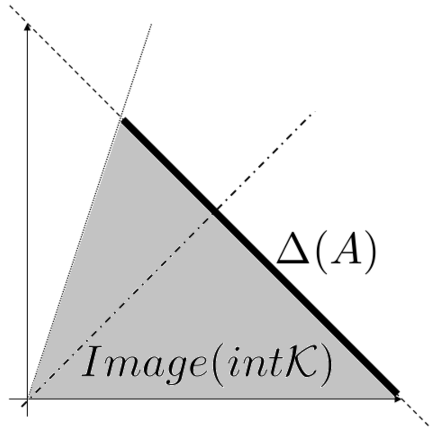
For example, if the map A has $Image(int\mathcal{K}) = \{(x,y) \in {\mathbb{R}_{>0}}^2 : \frac{x}{2} < y < 2x\}$



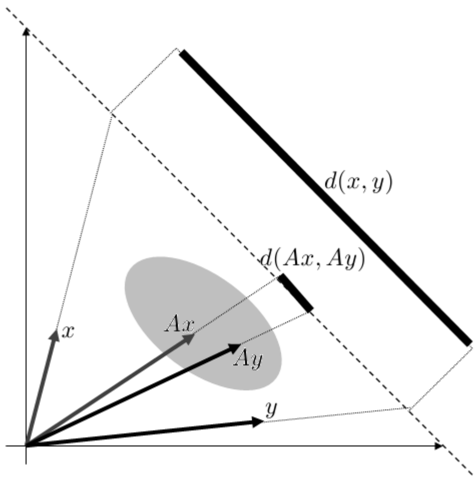
In this case, the projective diameter $\Delta(A)$ is finite; for example, this corresponds to consider the linear system

$$A\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = A \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

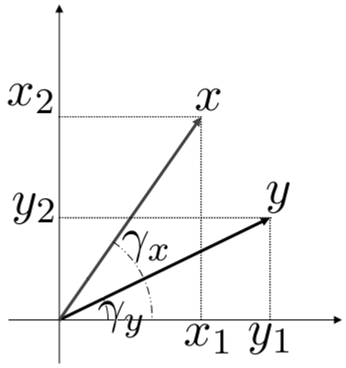
On the other hand, if $Image(int\mathcal{K}) = \{(x, y) \in \mathbb{R}_{>0}{}^2 : 0 \leq y < 3x\}$ the diameter is infinite:



In order to represent the contraction ratio $k(A)$, let us consider two different vectors $x, y \in \mathbb{R}_{>0}{}^2$ mapped in $A(x)$ and $A(y)$. Then, let us consider the ratio between the projective distances $d(A(x), A(y))$ and $d(x, y)$; the contraction ratio is the inferior of that ratio varying x and y.



Finally, let us try to understand better the appearance of the *Hilbert metric* d; observing that $d(x, y) = \log[M(x, y)/m(x, y)]$ we can focalize our attention to the ratio $M(x, y)/m(x, y)$



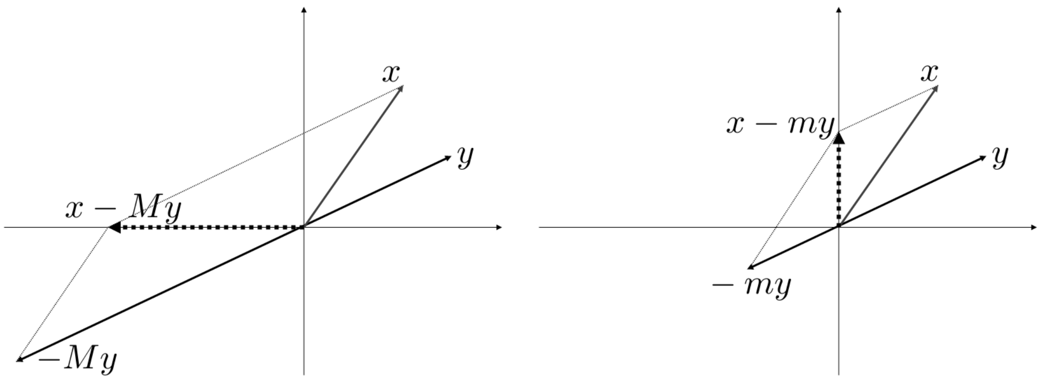
We can easily obtain

$$
\begin{aligned}
M(x,y) &= \inf\{\lambda : x - \lambda y \preceq 0\} = \inf\{\lambda : \lambda y - x \in \mathcal{K}\} = \\
&= \inf\{\lambda : \lambda y_1 - x_1 \geq 0 \quad \lambda y_2 - x_2 \geq 0\} = \\
&= \inf\{\lambda : \lambda \geq x_1/y_1 \quad \lambda \geq x_2/y_2\} = \max\{x_1/y_1, x_2/y_2\}
\end{aligned}
$$

and analogously

$$m(x,y) = \inf\{\lambda : \lambda \leq x_1/y_1 \quad \lambda \leq x_2/y_2\} = \min\{x_1/y_1, x_2/y_2\}$$

These concepts can be represented in the following manner:



In this representation, $x_1/y_1 < x_2/y_2$, and so $M = x_2/y_2$ while $m = x_1/y_1$. Furthermore,

$$\frac{M}{m} = \frac{x_2/y_2}{x_1/y_1} = \frac{x_2/x_1}{y_2/y_1} = \frac{\tan \gamma_x}{\tan \gamma_y}$$

where γ_x and γ_y are the angles pictorially defined above, and so

$$d(x,y) = |\log \tan \gamma_x - \log \tan \gamma_y|$$

This expression emphasizes a correspondence with the angles, and in particular the projective nature of this distance:

$$d(\alpha x, \beta y) = \log \tan \gamma_{\alpha x} - \log \tan \gamma_{\beta y} = \log \tan \gamma_x - \log \tan \gamma_y = d(x,y)$$

2.3.3 Results

To apply the previous Birkhoff Theorem in consensus, consider the **positive orthant** as $\mathcal{K}$.
In this case, $\mathcal{X} = \mathbb{R}^n$ and $\mathcal{K} = \{(x_1, ..., x_n) : x_i \geq 0, \quad 1 \leq i \leq n\}$. So,

$$M(x, y) = \inf\{\lambda : (x_1, ..., x_n) - (\lambda y_1, ..., \lambda y_n) \preceq 0\} = \max_i (x_i/y_i)$$

$$m(x, y) = \sup\{\lambda : (x_1, ..., x_n) - (\lambda y_1, ..., \lambda y_n) \succeq 0\} = \min_i (x_i/y_i)$$

Stochastic matrices $A(t)$, as we have introduced them in Definition 7, define *positive monotone maps in* $\mathcal{K}$, and by linearity they are also homogeneous of degree $p = 1$.

We observe that $\mathbf{1} = [1, 1, 1, ..., 1]^T$ is a *fixed point for all* $A(t)$: $A(t) \cdot \mathbf{1} = \mathbf{1}$.
Birkhoff Theorem implies that $k(A(t)) = \tanh \frac{1}{4}\Delta(A(t))$, and being $k(A) = \inf\{\lambda : d(Ax, Ay) \leq \lambda d(x, y)\}$, taking $y = \mathbf{1}$, $d(Ax, \mathbf{1}) \leq k(A) d(x, \mathbf{1})$ and so:

$$d\big(A(t)x, \mathbf{1}\big) \leq \Big(\tanh \frac{1}{4}\Delta(A(t))\Big) d\big(x, \mathbf{1}\big) \qquad \forall t \in \mathbb{N}$$

Let us notice that this implies that the distance from consensus is not increasing, because $0 < \tanh(a) \leq 1 \quad \forall a \in \mathbb{R}_{>0}$. The fundamental observation is that **when the diameter is finite, the contraction coefficient is strictly smaller than one**.

However, if $A(t)$ has not finite diameter, it is possible to aggregate a number of subsequent matrices $A(t)A(t+1)...A(t+T)$ requiring that this finite product has a finite diameter.

Let us formalize this argumentation:

Theorem 6 (Consensus Criteria on the Positive Orthant) *Consider the linear system 1.2 :*

$$x(t+1) = W(t)x(t)$$

with $W(t)$ sequence of row-stochastic $n \times n$ matrices.
If exists $T \geq 0$ such that the finite product $W(t+T)...W(t+1)W(t)$ has a finite diameter for all $t \in \mathbb{N}$, then the n components $\xi_i(t)$ of any solution ξ of the system converges to a common value as $t \to \infty$.

We can express the diameter like:

$$\Delta(A(t)) = \sup\Big\{ \log \frac{a_{ij}(t)a_{pq}(t)}{a_{iq}(t)a_{pj}(t)} \quad : \quad 1 \leq i, j, p, q \leq n \Big\}$$

$$A(t) = \left[\begin{array}{c|c|c|c|c} & \vdots & & \vdots & \\ \hline \cdots & a_{ij}(t) & \cdots & a_{iq}(t) & \cdots \\ \hline & \vdots & & \vdots & \\ & \vdots & & \vdots & \\ \hline \cdots & a_{pj}(t) & \cdots & a_{pq}(t) & \cdots \\ \hline & \vdots & & \vdots & \end{array}\right]$$

Infact,

$$Ax = \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j \end{bmatrix} \qquad Ay = \begin{bmatrix} \sum_{q=1}^n a_{1q}y_q \\ \sum_{q=1}^n a_{2q}y_q \\ \vdots \\ \sum_{q=1}^n a_{nq}y_q \end{bmatrix}$$

and so

$$M(Ax, Ay) = \max_i \frac{\sum_{j=1}^n a_{ij}x_j}{\sum_{q=1}^n a_{iq}y_q} \qquad m(Ax, Ay) = \min_i \frac{\sum_{j=1}^n a_{ij}x_j}{\sum_{q=1}^n a_{iq}y_q}$$

$$\Delta(A) = \sup\left(\log \frac{M(Ax, Ay)}{m(Ax, Ay)}\right) = \log\left(\sup_{ip} \sum_{j,q=1}^{n} \frac{a_{ij}x_j \cdot a_{pq}y_q}{a_{pj}x_j \cdot a_{iq}y_q}\right)$$

By choosing x and y to be suitable unit vectors $x_j = 1$ and $y_q = 1$ we can obtain the limit

$$\Delta(A) = \log\left(\sup_{i,p,j,q} \frac{a_{ij}a_{pq}}{a_{pj}a_{iq}}\right)$$

which cannot be exceeded since averaging (by positive weight factors $x_j y_q$) always makes ratios less extreme.

For each $x \in int\mathcal{K}$, Birkhoff Theorem provides the Lyapunov function

$$V(x) = d(x, \mathbf{1}) = \log \frac{\max_i(x_i)}{\min_i(x_i)} = \max \log(x_i) - \min \log(x_i)$$

namely the *projective distance from* $[x]$ *to* $[\mathbf{1}]$: this is a link with the previous approach, and in this sense the distance between $x(t+1)$ and $span\{\mathbf{1}\}$ can be seen as the Lyapunov function of Section 2.1 in *logarithmic coordinates.*

Let us anticipate that there are cases where, despite a uniform convergence to consensus, the diameter is infinite over any finite horizon. It will be shown in the Limit Cases studied in Chapter 3.

2.3.4 Generalization to Non-Commutative Consensus

The intent of [2] was not only an application of Birkhoff Theorem on linear consensus; infact, as we will illustrate with an example, sometimes this approach is less general then the first one, because the matrix diameter is required to be finite.

The authors of [2] have pursued an interesting generalization to the *non-commutative case*: instead of the positive orthant, they consider the *cone of positive definite matrices*, taking

- $\mathcal{X} = \{X = \overline{X} \in \mathbb{C}^{n \times n}\}$ instead of $\mathcal{X} = \mathbb{R}^n$
- $\mathcal{K} = \{X \succeq 0 : X \in \mathcal{X}\}$ instead of $\mathcal{K} = \{(x_1, ..., x_n) : x_i \geq 0, 1 \leq i \leq n\}$

Here, an analog of stochastic maps is given by the *dual maps of Kraus maps describing quantum channels.*

Definition 11 *The map* Ψ *is called* **Kraus map** *if admits the form*

$$\Psi(Z) = \sum_i V_i Z \overline{V_i} \quad \forall Z \in \mathcal{K}$$

with $\sum_i \overline{V_i} V_i = I$*; these maps are* positive *and* trace-preserving*, mapping positive matrices to positive matrices, and in quantum applications they operate on states (positive symmetric marices* X *with* $tr(X) = 1$*).*

The map Φ *is called* **dual map** *of* Ψ *if admits the form*

$$\Phi(X) = \sum_i \overline{V_i} X V_i \quad \forall X \in \mathcal{K}$$

These maps are positive *and* unital*, mapping* $\Phi(I) = I$*.*

In particular, let us observe that the *unital property* of Φ is the analog of $A(t)\mathbf{1} = \mathbf{1}$ for stochastic matrices.
This suggest the following non-commutative extension of consensus algorithm:

$$X(t+1) = \Phi_t(X(t)), \quad X(0) \in int\mathcal{K}$$

The details regarding this interesting generalization, however, are not included in the present work: we refer the reader to the original contribution [2], Section [VI] .

3

Analysis of Limit Cases

3.1 Similarities between the two Approaches

Let us summarize the previous analysis to give the correspondences between the two approaches we have seen.

The first approach (Section 2.1) uses in a direct way the Lyapunov theory, with set-valued functions. Lyapunov theory says that given a discrete-time system $x(t+1) = f(t, x(t)) \quad t \in \mathbb{N}$, defined on an Euclidean finite space, this system is convergent respect to a collection of equilibrium solutions Φ if there exists a set-valued function V such that $V(f(t, x(t))$ is *strictly decreasing* for $f(t, x(t)) \notin \Phi$.
In the consensus linear case, with n agents and $x(t) \in \mathbb{R}^n$, the collection of equilibrium solutions we are interested in is

$$\Phi = span\{\mathbf{1}\} = \{(\alpha, \alpha, ..., \alpha) \in \mathbb{R}^n\}$$

Taking the convex hull of $x_1(t), x_2(t), ..., x_n(t)$ as Lyapunov function, in general nothing can be said about its increase, or decrease, after the application of Φ. The *Convexity Assumption* is sufficient to ensure that it is not increasing, but it is still not sufficient to prove convergence. Finally, imposing the existence of an agent which enables a sort of *global connection*, there is *strict decreasing* of V.
In this sense V is a sort of representation of *disagreement*, which is strictly decreasing respect to consensus under these assumptions; we can think it as a time-varying compact set that contains the system dynamics, whose convergence to Φ implies the system convergence to an element of Φ.

The second approach (Section 2.2) is an application of Birkhoff theorem. Birkhoff theorem says that, given a function A, defined on a closed solid cone (as we have definied it), if it is (1) positive, (2) homogeneous of degree p, (3) monotone, then the contractio ratio of A is limited. Applying the theorem to the discrete-time linear system $x(t+1) = A(x(t)) \quad t \in \mathbb{N}$, the contraction ratio at time t is defined as the ratio of the distance between $x(t+1)$ and $span\{\mathbf{1}\}$ and the distance between $x(t)$ and $span\{\mathbf{1}\}$. If the map $A(t)$ satisfies (1)-(3) this ratio is smaller than one if the diameter of $A(t)$ is finite, and this means there is a *strict decreasing* of the distance from $span\{\mathbf{1}\}$.
The distance from $span\{\mathbf{1}\}$ can be interpreted again as a measure of *disagreement*; as already observed if the diameter of $A(t)$ is not finite, it is sufficient the existence of $T \in \mathbb{N}$ such that the diameter of $A(t)A(t+1)...A(t+T)$ is finite to have asymptotic consensus.

The distance between $x(t+1)$ and $span\{\mathbf{1}\}$ of Section 2.3 can be seen as the Lyapunov function of Section 2.1 in logarithmic coordinates: so, in both approaches, we require a *strict decreasing* of an opportune Lyapunov function.

3.2 Differences between the two Approaches

Let us analyze the differences regarding the consensus criteria formulations.

In both cases the update algorithm must have a convex nature. The main difference regards topology and connectivity.

- The first approach requires the existence of a node connected to all other nodes across a certain interval, that can be $[t_0, \infty)$ or $[t_0, t_0 + T]$ based on the context.
- The second approach requires that *the diameter of the concatenation of matrices* $W(t_0)W(t_0+1)...$ *is* $< \infty$.

We need to analyze better what a finite diameter implies to the row-stochastic matrix. Let us observe again the expression

$$\Delta(W(t)) = \sup\Big\{ \log \frac{w_{ij}(t)w_{pq}(t)}{w_{iq}(t)w_{pj}(t)} \quad : \quad 1 \leq i,j,p,q \leq n \Big\}$$

Obviously, $\Delta(W(t))$ is infinite whenever the logarithm argument is infinite too.
If $W(t)$ has infinite diameter, then it has at least a zero entry, otherwise that ratio could not be ∞. Let us suppose $w_{iq}(t) = 0$; if it is the only zero entry, then the diameter is infinite; instead, the diameter could be finite if one of following conditions is verified:

1. $w_{ij}(t) = 0$ and $w_{pj}(t) \neq 0$
2. $w_{pq}(t) = 0$ and $w_{pj}(t) \neq 0$
3. $w_{ij}(t) = 0$, $w_{pq}(t) = 0$ and $w_{pj}(t) = 0$

Taking into account the constraints on the dynamical matrix W, we have that *condition 1.* cannot be verified, because if $w_{iq}(t) = 0$ from row-stochasticity exists at least a not zero entry on the same row i; *condition 3.* cannot be verified, because $W(t)$ is no empty and we can always find at least a quartet $w_{iq}(t), w_{pq}(t), w_{ij}(t), w_{pj}(t)$ with a not zero entry; *condition 2.* could be verified only if there is at least an empty column q; the number of those columns must be less then n in order to preserve row-stochasticity.

An important conclusion is that *the projective diameter of a row-stochastic matrix is finite if and only if every zero entry belongs to a zero column.*

▷ **Example** Consider the row-stochastic matrix

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

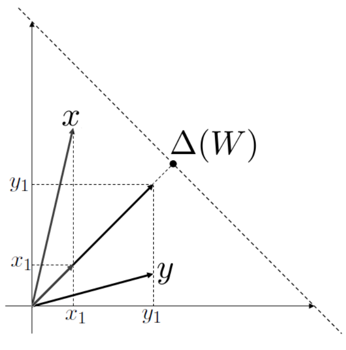
Considering two generic vectors $x = [x_1, x_2]^T$, $y = [y_1, y_2]^T$, we obtain:

$$Wx = \begin{bmatrix} x_1 \\ x_1 \end{bmatrix} \qquad Wy = \begin{bmatrix} y_1 \\ y_1 \end{bmatrix}$$

Computing the distance from these vectors, we obtain:

$$\Delta(W) = \log \frac{M\left(\begin{bmatrix} x_1 \\ x_1 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_1 \end{bmatrix}\right)}{m\left(\begin{bmatrix} x_1 \\ x_1 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_1 \end{bmatrix}\right)} = \log \frac{\max\{x_1/y_1, x_1/y_1\}}{\min\{x_1/y_1, x_1/y_1\}} = \log 1 = 0$$

The diameter becomes 0 because the vectors move to the same radius:



Let us observe that *Convexity Assumption* is not completely respected if we allow the existence of an empty column: in this case, exists $i \in \mathcal{N}$ such that $w_{ii} = 0$. The interpretation of the empty column is: "each agent j does not take account of agent i"; for $i \neq j$ it can be due to a lack of communication, and it does not directly avoid Convexity. The unusual aspect is that agent i does not take account of itself too.

The second approach seems to be more general in this case, because does not require a strict local convexity property. If we consider the previous example and a linear system with $W(t) = W \quad \forall t \in \mathbb{N}$, of course asymptotic consensus is guaranteed.

Let us observe that when $w_{ii}(t) = 0$, agent i becomes a beholder: it only "watches" other agents without any contribution to the system dynamics. So, this case could be reported to an $(n-1) \times (n-1)$ system where the first approach can be applied again.
A first conclusion is that:

Conclusion 1 *Whenever the* Local Convexity Assumption *does not hold, the second approach is more general than the first one.*
However, if that assumption does not hold because of a "beholder agent", the situation could be reported to the first approach ignoring that agent.

If we impose $w_{ii} \neq 0 \quad \forall i \in \mathcal{N}$, as in Section 2.2, then also condition 2. above cannot be verified (because *there are not empty columns*); so if the *Local Convexity Assumption* holds,

a row-stochastic matrix has finite diameter if and only if it is full

i.e. has no zero elements.
We can show with an example that in the following situation the first approach can be applied while the second one cannot.

▷ **Example** Consider the matrix

$$W = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

This is a row-stochastic matrix which respects the Convexity Assumption. Consider the system 1.2 having $W(t) = W \; \forall t \in \mathbb{N}$ as sequence of (constant) arc sets. Looking at W, there exists a node connected to all other nodes: this is the node 1, because the first column is nonempty; furthermore, the diameter of W is infinite, because

$$\Delta(W) = \sup\left\{ \log \frac{w_{ij}(t) w_{pq}(t)}{w_{iq}(t) w_{pj}(t)} \quad : \quad 1 \leq i, j, p, q \leq 3 \right\} \to \infty$$

Observe W is lower-triangular, and any finite power W^k is a lower-triangular row-stochastic matrix too, and has also an infinite diameter and the first column is nonempty.
On the other hand, the first approach can be applied, and in particular we can guess a convergence to $(x_1, x_2, x_3) \to (x_1(0), x_1(0), x_1(0))$; the second one cannot be applied.

The second approach requires that the matrices are full (in certain intervals); in those cases, the first approach can be applied too. In this way, we obtain the following conclusion:

Conclusion 2 *Whenever the* Local Convexity Assumption *holds, the first approach is more general than the second one.*
Furthermore, when the second approach can be applied, the first one can be applied too.

This conclusion could be guessed observing that Theorem 2–4 have an "if-and-only-if" formulation, while Theorem 6 is only a sufficient criteria.

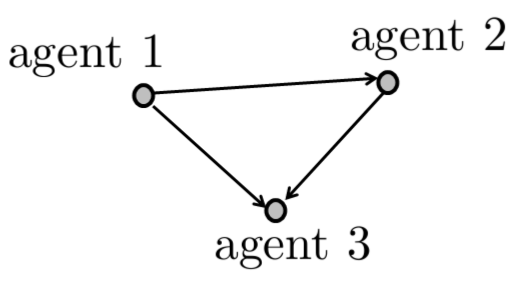
Finally, let us observe that there are cases where the first approach cannot be applied directly, but we can reduce those problems to it; this will be shown in the next example

---

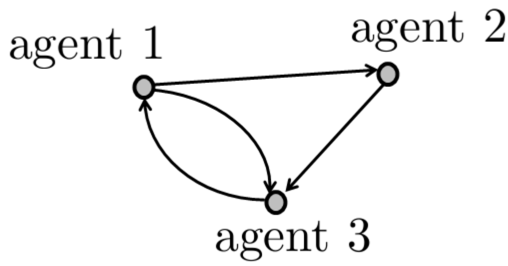
▷ **Example** Consider again the matrix

$$W = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

In the previous example, we have observed that it should respect the Convexity Assumption; we had implicitly assumed that the network topology of the multiagent system was



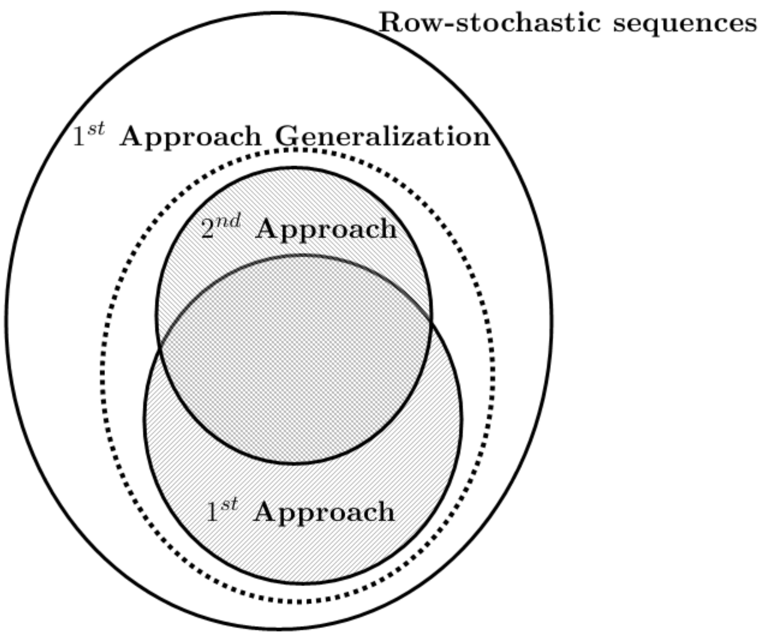
Suppose now that the network topology of the multiagent system is



the Convexity Assumption is not respected, because agent 1 updates its value without taking account of agent 3 although exists a link from 3 to 1. In other words, agent 1 does not update its value moving in the relative interior of the convex hull of its state and its neighbors states, but mantains itself on the border of the convex hull.

Although the first approach cannot be formally applied, the (constant) sequence of matrices is the same of the previous example, and so consensus will be reached.

The previous example shows that *cutting all the useless arcs*, the first approach could be applied again although the Convexity Assumption does not hold.



3.3 A Limit Case: the Headstrong Node

As limit case, let us see what happens when there is a *headstrong agent*, namely an agent which mantains its state in time. Obviously, if consensus will be reached all the agents will have the same state of the *headstrong agent*, because it does not change its state. Furthermore, we can observe that if there were two or more headstrong agents with different states, consensus would never been reached.

Without loss of generality, let the agent 1 be the headstrong one (the agents' order is irrelevant). The row-stochastic matrix relative to the system at instant $t \in \mathbb{N}$ has the next form:

$$A(t) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{21}(t) & a_{22}(t) & a_{23}(t) & \cdots & a_{2n}(t) \\ a_{31}(t) & a_{32}(t) & a_{33}(t) & \cdots & a_{3n}(t) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1}(t) & a_{n2}(t) & a_{n3}(t) & \cdots & a_{nn}(t) \end{bmatrix}$$

with $\sum_{j=1}^{n} a_{ij}(t) = 1 \quad \forall i \in [2, n]$.

Case n=2

Consider the case of two only agents, and let the first of them be headstrong. Let the initial state at time $t = 0$ be $x(0) = [x_1, x_2]^T$. At every instant $t \in \mathbb{N}$, the row-stochastic matrix associated to the system has the form:

$$A(t) = \begin{bmatrix} 1 & 0 \\ \alpha(t) & 1 - \alpha(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \alpha(t) & \beta(t) \end{bmatrix}$$

so, applying $A(0)$, $A(1)$, $A(2)$,... to $x(0)$ we have

$$x(1) = \begin{bmatrix} x_1(1) \\ x_2(1) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \alpha(0) & \beta(0) \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ \alpha(0)x_1 + \beta(0)x_2 \end{bmatrix}$$

and then applying $A(1)$ to $x(1)$:

$$x(2) = \begin{bmatrix} 1 & 0 \\ \alpha(1) & \beta(1) \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \alpha(0)x_1 + \beta(0)x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ [\alpha(0) + \alpha(1)\beta(0)]\, x_1 + \beta(0)\beta(1)x_2 \end{bmatrix}$$

and so on

$$x(3) = \begin{bmatrix} x_1 \\ [\alpha(0) + \alpha(1)\beta(0) + \alpha(2)\beta(0)\beta(1)]\, x_1 + \beta(0)\beta(1)\beta(2)x_2 \end{bmatrix}$$

With a simple induction, we observe that

$$x(n+1) = A(0)A(1)...A(n)x(0) = A_n x(0)$$

and

$$A_n = \begin{bmatrix} 1 & 0 \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \sum_{i=0}^{n} \alpha(i) \prod_{j=0}^{i-1} \beta(j) & \prod_{k=0}^{n} \beta(k) \end{bmatrix}$$

There is asymptotic consensus if and only if

$$A_n \to \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \qquad for \quad n \to \infty$$

It is necessary that

$$\lim_{n\to\infty} \prod_{k=0}^{n} \beta(k) = 0$$

We cannot ask that $\beta(k) = 0$ for a certain k, because in the consensus algorithm the second agent always considers its own value; so

$$0 < \beta(k) \leq 1 \quad \forall k$$

On the other hand, if $\beta(k) = 1 \quad \forall k$, the limit will not be verified, and consensus will not be reached; however, when $\beta(k) = 1$ the row-stochastic matrix is equal to the identity matrix, and it does not matter in the product.

Thus, the condition that allows the limit to be verified is that $0 < \beta(k) < 1$ an infinite number of times. Infact, under this assumption, exists $0 < \sigma < 1$ such that $\beta(k) < \sigma \quad \forall k \quad | \quad \beta(k) \neq 1$;

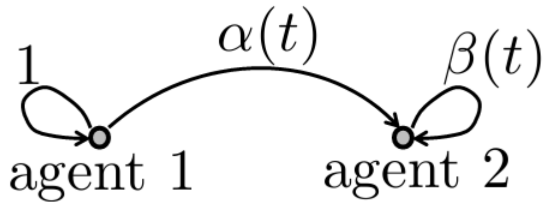
$$\lim_{n\to\infty} \prod_{k=0}^{n} \beta(k) = 0 \Longleftrightarrow \log(\prod_{k=0}^{\infty} \beta(k)) = \sum_{k=0}^{\infty} \log \beta(k) = -\sum_{k=0}^{\infty} \log \frac{1}{\beta(k)} \to -\infty$$

Operating the majorization

$$\sum_{k=0}^{\infty} \log \frac{1}{\beta(k)} > \sum_{k=0}^{\infty} \log \frac{1}{\sigma} \to \infty$$

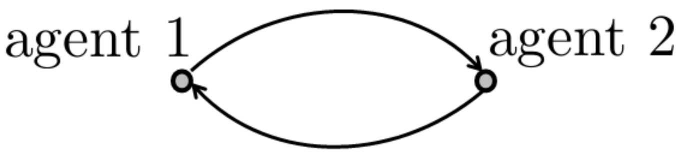
it is proved.

Interpretation
We can observe a correspondence with the first approach, which requires the existence of $T \geq 0$ such that in every interval of length T, a node (agent 1) is connected to all other nodes (agent 2).



The second approach, instead, cannot be applied because the system matrix has infinite diameter in each instant, as explained in the Example of Subsection 3.2.

Let us suppose that the network configuration were



The previous analysis remains unchanged, but the Local Convexity Assumption does not hold because agent 1 does not move into the *interior* of the convex hull. The first approach cannot be directly applied, but cutting the useless arc $(2, 1)$ we can do that (as we have seen in the previous subsection).

Conclusion

After introducing a dynamical model for n agents which communicate through time-dependent links and recalling what *convergence to consensus* means in this context, we have analyzed some results emerging from two different approaches to the consensus problem.

The first one employs set-valued Lyapunov theory, by exploiting the decreasing of a certain set-valued function under proper convexity assumptions, while the second one relies on geometric concepts, in particular the conclusions of a theorem by Birkhoff on positive maps onto cones.

We have compared these results in the *linear case*, where agents have a real-valued state and the system dynamics are associated to row-stochastic matrices.
The main difference is that the first approach requires a sort of *local convexity* (each agent updates its state according to a strict convex combination of its neighbors' states) while the second approach requires a *finite projective diameter* (i.e. a geometric condition on the systems matrices).

When the convexity assumption does not hold, the first approach cannot be directly applied; however, if the second approach can be applied, we have shown that we could resort to the first one by cutting some nodes that do not actually contribute to the dynamics.
When the convexity assumption holds, the first approach can be applied and is more general than the second one.
In some cases where convexity assumption does not formally hold, the first approach can be applied yet cutting some useless arcs.

The main result is that in the linear case the first approach is more general than the second one, although as it does not immediately provides for useful generalizations as the second one does, e.g. to study consensus on the non-commutative spaces of interest in the study of quantum channels.

Bibliography

[1] Luc Moreau, "Stability of Multiagent Systems With Time-Dependent Communication Links", *IEEE Transactions on Automatic Control*, vol.50, no. 2, February 2005

[2] Rodolphe Sepulchre, Alain Sarlette, Pierre Rouchon, "Consensus in Non-Commutative Spaces", *49th IEEE Conference on Decision and Control*, Atlanta, GA, USA, December 2010

[3] Vincent D. Blondel, Julien M. Hendrickx, Alex Olshevsky, John N.Tsitsiklis, "Convergence in Multiagent Coordination, Consensus, and Flocking", *44th IEEE Conference on Decision and Control*, Seville, Spain, December 2005

[4] Wei Ren, Randal W. Beard; *Distributed Consensus in Multi-vehicle Cooperative Control : Theory and Applications*, chapter 1 – Springer, 2008