Lyapunov Approach to Consensus Problems

Angelia Nedić

angelia@illinois.edu

Industrial and Enterprise Systems Engineering Department
and Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

joint work with Ji Liu (ECE, UIUC)

Motivation

Say we are given an $m \times m$ row-stochastic matrix A, think of it as a transition matrix of a Markov chain over m states.

Suppose that the chain is ergodic, so that

$$\lim_{t \to \infty} A^t = \mathbf{1}\phi',$$

where ϕ is the stationary distribution of the chain and every state has positive charge, i.e.,

$$\phi_i > 0 \qquad \text{for all } i.$$

Convergence rate for $\|A^t - \mathbf{1}\phi'\|$ is known to be exponential

$$\|A^t - \mathbf{1}\phi'\|^2 \leq c\left(1 - \frac{\alpha_0}{f(m)}\right)^t \qquad \text{for all } t \geq 0,$$

where c, α_0 are some universal constants that do not depend on m.

What is the best known convergence rate (in terms of the dependence on m)?

- For deterministic graphs? $f(m) \approx O(m^2)$
- For random graphs? $f(m) \approx O(m \log_2 m)$

Back to Weighted Averaging Models for Consensus

- The literature is vast: Tsitsiklis 1984, Jadbabie Morse and Lin 2003 and the stream of publications that followed

- Unconstrained consensus problem - can be used as a mechanism for information diffusion in networks to deal with optimization, tracking, estimation, learning in networks

- Understanding of the convergence rate is critical for making use of the consensus as a building block for other algorithms over the network AN and Ozdaglar 2007, 2008, 2009, 2010

- Convergence rate results did not explicitly capture the network structure

- Alternative approach by using a different Lyapunov function to measure the progress of the consensus-dynamics allows us to capture the rate in terms of network structure

Unconstrained Consensus Problem

- Consider a system consisting of m agents (nodes, sensors, robots, etc), represented by a set $[m] = \{1, \ldots, m\}$

- We assume that a sequence $\{G_t\}$ of directed graphs is given externally, where each graph G_t represents communication structure among the agents, where $G_t = ([m], E_t)$

We consider the unconstrained (scalar) consensus problem, formalized as follows.

[Unconstrained Consensus] *Design a distributed algorithm obeying the communication structure given by graph G_t at each time t and ensuring that, for every set of initial values $x_i(0) \in \mathbb{R}$, $i \in [m]$, the following limiting behavior emerges:*

$$\lim_{t\to\infty} x_i(t) = c \qquad \text{for all } i \in [m] \text{ and some } c \in \mathbb{R}.$$

Literature: VAST

Roughly speaking there are three approaches

- Push-sum or Ratio Consensus Algorithm (discussed yesterday)

- Laplacian-Based Algorithm:

$$x(t+1) = \left(I - \frac{1}{\gamma}L(t)\right)x(t)$$

  where $x(t) = [x_1(t), \ldots, x_m(t)]'$ and $\gamma \geq m$ (see Jadbabaie et al. 2003)

- Weighted-Averaging Algorithm

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

  where $A(t)$ is a row-stochastic matrix with sparsity pattern matching the graph G_t structure (Tsitsiklis 1984)

Most of the Laplacian-based algorithms require that each $L(t)$ is also *symmetric* which implicitly require bidirectional communication links.

Weighted-averaging algorithm gets around this limitation.

Weighted-Averaging Algorithm

$$x(t+1) = A(t)x(t),$$

where the weight matrices $A(t)$ are row-stochastic and compliant with the graph G_t structure (to be discussed soon)

The existing analysis of the weighted-averaging is based on studying the behavior of the left-matrix products

$$x(t) = A(t)A(t-1)\cdots A(s+1)A(0)x(0) \qquad \text{for } t \geq 0,$$

In particular, when the matrices $A(t)A(t-1)\cdots A(1)A(0)$ converge to a rank one matrix, the iterates $x(t)$ converge to a consensus

Some conditions on the graphs G_t and the matrices $A(t)$ that convergence are:
(Tsitsiklis 1984, Nedić and Ozdaglar 2009, Nedić, Olshevsky, Ozdaglar, Tsitsiklis 2009)

Assumption Let $\{G_t\}$ be a graph sequence and $\{A(t)\}$ be a sequence of $m \times m$ matrices that satisfy the following conditions:

(a) Graph Compliance Each $A(t)$ is a stochastic matrix that is compliant with the graph G_t, i.e., $A_{ij}(t) > 0$ when $(j,i) \in E_t$, for all t.

(b) Aperiodicity The diagonal entries of each $A(t)$ are positive, $A_{ii}(t) > 0$ for all t and $i \in [m]$.

(c) Uniform Positivity There is a scalar $\beta > 0$ such that $A_{ij}(t) \geq \beta$ whenever $A_{ij}(t) > 0$.

(d) Irreducibility Each G_t is strongly connected.

Convergence Rate

Under the preceding Assumption we have

$$\lim_{t\to\infty} A(t)\cdots A(k+1)A(k) = \mathbf{1}\phi'(k) \qquad \text{for all } k \geq 0,$$

where each $\phi(k)$ is stochastic vector and $\mathbf{1} = [1, \ldots, 1]'$.

Furthermore, the convergence rate is geometric: for all $t \geq k \geq 0$,

$$\|A(t)\cdots A(k+1)A(k) - \mathbf{1}\phi'(k)\|^2 \leq Cq^{t-k},$$

where the constants $C > 0$ and $q \in (0,1)$ depend only on m and β.

When the matrices $A(t)$ are doubly stochastic, we have for all $t \geq k \geq 0$,

$$\left\|A(t)\cdots A(k+1)A(k) - \frac{1}{m}\mathbf{1}\mathbf{1}'\right\|^2 \leq \left(1 - \frac{\beta}{2m^2}\right)^{t-k}.$$

Refs. Tsitsiklis 1984, AN and Ozdaglar 2009; AN, Olshevsky, Ozdaglar, and Tsitsiklis 2009

These and other existing rate results *are not explicitly capturing the structure of the graph* G_t such as the longest shortest path or the maximum node degrees for example.

Convergence Rate: Lyapunov Approach

New rate results are possible by adopting dynamic system point of view and applying Lyapunov approach

This approach allows us to characterize the convergence of the weighted-averaging algorithm with a more explicit dependence on the graph structure than the existing results

In particular, we work with a quadratic Lyapunov comparison function proposed by Touri 2011, and Touri and AN 2014

In this approach, *an absolute probability sequence of matrices $A(t)$ play a critical role* in the construction of a Lyapunov comparison function and in establishing its rate of decrease along the iterates of the algorithm.

Definition For a chain of row-stochastic matrices $\{A(t)\}$ we say that the vector sequence $\{\pi(t)\}$ is an absolute probability sequence of the chain if each $\pi(t)$ is a probability vector and

$$\pi'(t) = \pi'(t+1)A(t) \qquad \text{for all } t \geq 0.$$

NOTE The absolute probability sequence is an adjoint dynamic for the weighted-averaging dynamic

Assumptions

An undirected tree $\mathcal{T}$ is weakly spanning tree of a directed graph $G = ([m], E)$, if for every $\{i, j\} \in \mathcal{T}$ we have either $(i, j) \in E$ or $(j, i) \in E$

Assumption 1

Let $\{G_t\}$ be a graph sequence and $\{A(t)\}$ be a matrix sequence such that:

(a) (*Partial Irreducibility*)

Each graph G_t contains a weakly spanning tree and each $A(t)$ is a stochastic matrix that is compliant with a weakly spanning tree T_t of G_t, i.e., $A_{ij}(t) > 0$ whenever $(j, i) \in \mathsf{T}_t$ for all $t \geq 0$.

(b) (*Aperiodicity*)

The diagonal entries of each $A(t)$ are positive, $A_{ii}(t) > 0$ for all t, and $i \in [m]$.

(c) (*Partial Uniform Positivity*)

There is a scalar $\beta > 0$ such that $A_{ii}(t) \geq \beta$ and $A_{ij}(t) \geq \beta$ for all $(j, i) \in \mathsf{T}_t$ and for all $t \geq 0$.

(d) (*Adjoint Dynamics / Absolute Probability Sequence*)

The matrix sequence $\{A(t)\}$ has an absolute probability sequence $\{\pi(t)\}$ that is uniformly bounded away from zero, i.e., there is $\delta \in (0, 1)$ such that $\pi_i(t) \geq \delta$ for all i and t.

Basic Result

Theorem Under Assumption 1, for the iterates $\{x(t)\}$ generated by the weighted-averaging algorithm with any initial vector $x(0) \in \mathbb{R}^m$, we have for any $t \geq k \geq 0$,

$$\sum_{i=1}^{m} \pi_i(t) \left(x_i(t) - \pi(0)'x(0)\right)^2 \leq \left(1 - \frac{\delta\beta^2}{p^*}\right)^{t-k} \sum_{j=1}^{m} \pi_j(k) \left(x_j(k) - \pi(0)'x(0)\right)^2,$$

where

- $\beta > 0$ is the uniform lower bound on the entries in $A(t)$ (Assumptions 1(c))
- $\delta > 0$ is the uniform lower bound on the entries of the absolute probability vectors $\pi(t)$ (Assumption 1(d)),
- while
  $$p^* = \max_{s \geq 0} p^*(s)$$
  where $p^*(s)$ is the longest shortest path in the tree T_s of Assumption 1(a).

Proof Sketch

Proof: Based on the use of Lyapunov Function

$$V(x,\pi) = \sum_{i=1}^{m} \pi_i (x_i - \pi' x)^2$$

and the relation

$$V\left(x(t+1), \pi(t+1)\right) = V\left(x(t), \pi(t)\right) - D(t),$$

where

$$D(t) = \frac{1}{2} \sum_{i=1}^{m} \pi_i(t+1) \sum_{j=1}^{m} \sum_{\ell=1}^{m} A_{ij}(t) A_{i\ell}(t) \left(x_j(t) - x_\ell(t)\right)^2.$$

Abound for the decrement is obtained (using Assumption 1)

$$D(t) \geq \frac{\delta \beta^2}{p^*(t)} \max_{j,\ell \in [m]} \left(x_j(t) - x_\ell(t)\right)^2 \quad \text{for } t \geq 0,$$

$p^*(t)$ is the maximum number of links in any of the shortest paths in the tree T_t of Assumption 1(a).

We also use

$$\pi'(t) x(t) = \pi'(0) x(0)$$

Implications of Theorem 1: Doubly Stochastic

Let Assumption 1 hold, and assume also that the weight matrices $A(t)$, $t \geq 0$, are doubly stochastic. Then, we have $\pi(t) = \frac{1}{m}\mathbf{1}$ and the relation of Theorem 1 reduces to:

$$\|x(t) - \bar{x}(0)\mathbf{1}\|^2 \leq \left(1 - \frac{\beta^2}{mp^*}\right)^{t-k} \|x(k) - \bar{x}(0)\mathbf{1}\|^2, \qquad (1)$$

with $\bar{x}(0) = \frac{\mathbf{1}'x(0)}{m}$. Since the maximum path length in a tree is no larger than $m-1$, i.e., $p^*(s) \leq m-1$, it follows that

$$\|x(t) - \bar{x}(0)\mathbf{1}\|^2 \leq \left(1 - \frac{\beta^2}{m(m-1)}\right)^{t-k} \|x(k) - \bar{x}(0)\mathbf{1}\|^2.$$

Thus, when β does not depend on m, the convergence rate has dependency of $O(m^2)$ in terms of the number m of agents, which is the same as the rate result in our earlier work (see slide 4)

Think of $\beta = \frac{1}{\max_{i,t}|N_i(t)|}$.

Given m, what graphs on m nodes will be the best?

Implications of Theorem 1: New Bound

Suppose now that we want to construct the graphs G_t such that Assumption 1 holds and we want to get the most favorable rate dependency on m. In this case, the following result is valid.

Theorem 1 *There is a sequence* $\{G_t\}$ *of regular undirected graphs such that for all* $x(0) \in \mathbb{R}^m$ *and all* $t \geq k \geq 0$,

$$\|x(t) - \bar{x}(0)\mathbf{1}\|^2 \leq q^{t-k} \|x_j(k) - \bar{x}(0)\mathbf{1}\|^2 ,$$

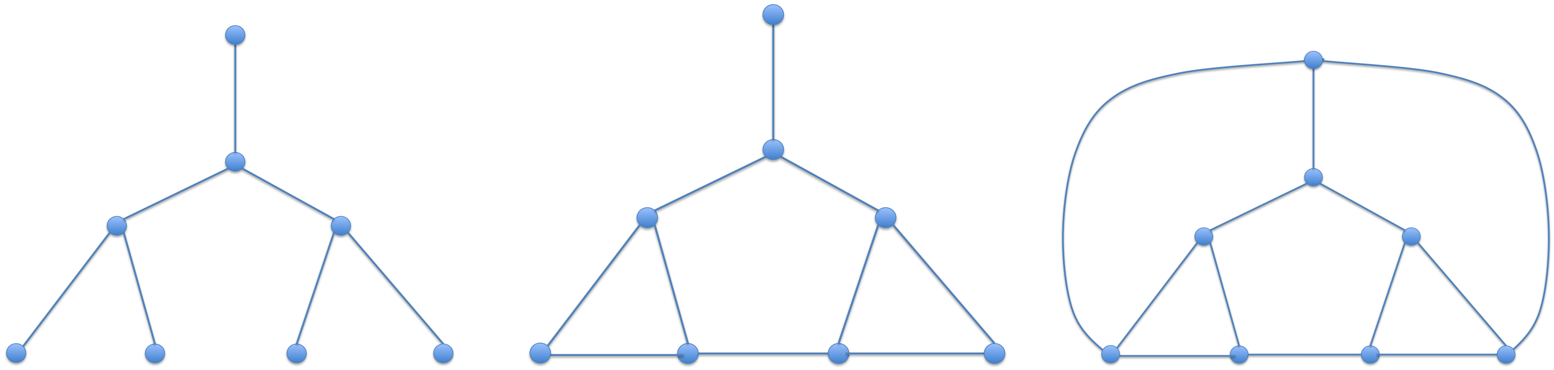
with $\bar{x}(0) = \frac{\mathbf{1}' x(0)}{m}$ *and*

$$q = 1 - \frac{1}{4^3 m \lceil \log_2 m \rceil}$$

The rate of this order is provided in Diaconis and Stroock 1991 (Proposition 3 and Example 2.3) for a *static ergodic Markov Chain*.

Proof

We will construct an undirected graph sequence $\{G_t\}$ that satisfies Assumption 1.
Let $m = 2^d$ for some integer $d \geq 1$. Let t be arbitrary but fixed time. Select $2^d - 1$ agents and construct an undirected binary tree with these agents as nodes. Next, add one extra agent as a root with a single child (see left plot in the Figure). Thus, each agent i except for the root and the leaf agents has the degree equal to 3.
Consider, now connecting all leaf-nodes with undirected edges (see middle plot in Figure). Now, all leaf-agents have degree equal to 3 except for the far most left and far most right agents, each of which has the degree equal to 2. Connect these two agents to the root node (see right plot in Figure). In this way, the far most left and far most right leaf agents, as well as the root agent have degree 3. In the resulting regular undirected graph, we let $A_{ij}(t) = \frac{1}{4}$ for all $j \in N_i(t) \cup \{i\}$ and for all i, so that $\beta = \frac{1}{4}$. The shortest path from the root agent to any other agent in the graph is at most $\lceil \frac{d}{2} \rceil$ (going down from the root of the tree to the nodes at the depth $\lceil \frac{d}{2} \rceil$, and going through the leaf nodes to reach those that are the depth larger than d).
Using the same construction, for all times t, we have that $\{A(t)\}$ is a sequence of doubly stochastic matrices, and therefore $\pi(t) = \frac{1}{m}\mathbf{1}$ for all t. Thus, Assumption 1 is



satisfied, and the estimate in (1) reduces to

$$\|x(t) - \bar{x}(0)\mathbf{1}\|^2 \leq \left(1 - \frac{1}{4^2 md}\right)^{t-k} \|x(k) - \bar{x}(0)\mathbf{1}\|^2 .$$

The result follows by noting that $d = \log_2 m$.

Theorem 1 shows that the exponential convergence rate with the ratio of the order $1 - O(\frac{1}{m \log_2 m})$ is achievable for consensus on some tree-like regular undirected graphs. This improves the best known bound with the ratio of the order $1 - O(\frac{1}{m^2})$ for undirected graphs and doubly stochastic matrices (see slide 4).

Implication for Matrix Sequence

We next consider an implication of Theorem 1 for the convergence of matrix products

$$A(t:k) \triangleq A(t)\cdots A(k+1)A(k) \qquad \text{for all } t \geq k \geq 0,$$

where $A(t:k) \triangleq A(k)$ whenever $t = k$.

Theorem 2 *If Assumption 1 holds, then for all $t \geq k \geq 0$,*

$$\left\|A(t:k) - \mathbf{1}\pi(k)'\right\|^2 \leq \frac{1}{\delta}\left(1 - \frac{\delta\beta^2}{p^*}\right)^{t-k}\left\|I - \mathbf{1}\pi(k)'\right\|^2,$$

where I is the identity matrix.

Corollary 3 *Under Assumption 1, the sequence $\{A(t)\}$ is ergodic:*

$$\lim_{t\to\infty} A(t)\cdots A(k) = \mathbf{1}\pi(k)'$$

for all $k \geq 0$.

Constrained Consensus

Assumption 2 We are given a collection of sets $X_i \subseteq \mathbb{R}^n$ which closed and convex, and their intersection is nonempty, i.e., $X \triangleq \cap_{i=1}^m X_i \neq \emptyset$.

The constrained consensus problem is as follows.

[Constrained Consensus] *Assuming that each agent i knows only its set X_i, design a distributed algorithm obeying the communication structure given by graph G_t at each time t and ensuring that, for every set of initial values $\mathbf{x}_i(0) \in \mathbb{R}^n$, $i \in [m]$, the following limiting behavior emerges:*

$$\lim_{t\to\infty} \mathbf{x}_i(t) = c \qquad \text{for all } i \in [m] \text{ and some } c \in X.$$

We have considered this problem in AN, Ozdaglar, and Parrilo 2010, proposing the following algorithm

$$\mathbf{w}_i(t+1) = \sum_{j=1}^m A_{ij}(t)\mathbf{x}_j(t),$$

$$\mathbf{x}_i(t+1) = \mathbb{P}_{X_i}[\mathbf{w}_i(t+1)], \qquad (2)$$

with a restriction to the use of doubly stochastic matrices

Convergence

The following result proves that the iterates of the algorithm converge to a common point in the set X.

Theorem 4 *Let Assumption 1 and Assumption 2 hold. Then, the sequences* $\{\mathbf{x}_i(t)\}$, $i \in [m]$ *of the algorithm* (2) *are bounded, i.e., there is a scalar* $\rho > 0$ *such that*

$$\|\mathbf{x}_i(t)\| \leq \rho \qquad \text{for all } i \in [m] \text{ and all } t \geq 0,$$

and they converge to a common point $x^* \in X$*:*

$$\lim_{t\to\infty} \mathbf{x}_i(t) = x^* \qquad \text{for some } x^* \in X \text{ and for all } i \in [m].$$

Some notes on the determining ρ

Boundedness of iterates: for all $y \in X$ and all $t \geq 0$,

$$\sum_{i=1}^{m} \pi_i(t)\|\mathbf{x}_i(t) - y\|^2 \leq \sum_{j=1}^{m} \pi_j(0)\|\mathbf{x}_j(0) - y\|^2$$

yields (under Assumption 1) for any $y \in X$, all t and i,

$$\|\mathbf{x}_i(t)\| \leq \|y\| + \frac{1}{\sqrt{\delta}} \max_j \|\mathbf{x}_j(0) - y\|$$

Convergence Rate

We have results for the sets X_i that satisfy a certain regularity condition which relates the distances from a given point to the sets X_ℓ with the distance from the point to the intersection set $X = \cap_{i=1}^m X_i$.

In particular, since $X \subseteq X_i$ for all i, it follows that

$$\mathrm{dist}(x, X_i) \le \mathrm{dist}(x, X) \text{ for all } x \in \mathbb{R}^n \text{ and } i \in [m]. \tag{3}$$

In our analysis, we need an upper bound on $\mathrm{dist}(x, X)$ in terms of the distances $\mathrm{dist}(x, X_i)$, $i \in [m]$.

We will use the following definition of set regularity.

Let $Z \subseteq \mathbb{R}^n$ be a nonempty set. We say that a (finite) collection of closed convex sets $\{Y_i, i \in \mathcal{I}\}$ is regular (in Euclidian norm) with respect to the set Z, if there is a constant $r \ge 1$ such that

$$\mathrm{dist}(y, Y) \le r \max_{i \in \mathcal{I}} \{\mathrm{dist}(y, Y_i)\} \qquad \text{for all } y \in Z.$$

We refer to the scalar r as a *regularity constant*. When the preceding relation holds with $Z = \mathbb{R}^n$, we say that the sets $\{Y_i, i \in \mathcal{I}\}$ are *uniformly regular*. In view of relation (3) it follows that the regularity constant r must satisfy $r \ge 1$.

Theorem 5 *Let Assumption 1 and Assumption 2 hold. Assume further that the sets* $\{X_i, i \in [m]\}$ *are regular, with a regularity constant* $r \geq 1$*, with respect to a ball* $B(0, \rho)$ *which contains all the iterates* $\{\mathbf{x}_i(t)\}$ *generated by the algorithm* (2)*. Consider the following Lyapunov comparison function:*

$$\varphi(t, y) \triangleq \sum_{i=1}^{m} \pi_i(t) \|\mathbf{x}_i(t) - y\|^2 . \tag{4}$$

Then, the Lyapunov comparison function $\varphi(t, \mathbf{v}(t))$ *decreases at a geometric rate:*

$$\varphi(t+1, \mathbf{v}(t+1)) \leq \left(1 - \frac{\delta\beta^2}{p^*(r+1)^2}\right) \varphi(t, \mathbf{v}(t)) \qquad \text{for all } t \geq 0,$$

where

$$\mathbf{u}(t) = \sum_{i=1}^{m} \pi_i(t)\mathbf{x}_i(t), \quad \mathbf{v}(t) = \mathbb{P}_X[u(t)], \text{ for all } t \geq 0. \tag{5}$$

and the scalars $\delta, \beta \in (0, 1)$ *and the integer* $p^* \geq 1$ *are the same as in Theorem 1.*

Theorem 6 *Under the assumptions of Theorem 5, for all* $t \geq 0$,

$$\sum_{j=1}^{m} \text{dist}^2(\mathbf{x}_j(t), X) \leq \frac{1}{\delta}\left(1 - \frac{\delta\beta^2}{4p^*(r+1)^2}\right)^t \varphi(0, \mathbf{v}(0)),$$

where $\mathbf{v}(0) = \mathbb{P}_X[\mathbf{u}(0)]$ *with* $\mathbf{u}(0) = \sum_{j=1}^{m} \pi_j(0)\mathbf{x}_j(0)$