A stochastic opinion dynamics model with multiple contents

Julio Cesar Louzada Pinto¹, Tijani Chahed¹ and Jérémy Jakubowicz¹

Abstract— We introduce a new model of opinion dynamics in which agents interact with each other about several distinct opinions/contents. In most of the literature about opinion dynamics, agents perform convex combinations of other agents’ opinions. In our framework, a competition between opinions takes place: agents do not exchange all their opinions with each other, they only communicate about the opinions they like the most. Our model uses scores to take this competition into account: each agent maintains a list of scores for each opinion held. Opinions are selected according to their scores (the higher its score, the more likely an opinion is to be expressed) and then transmitted to neighbors. Once an agent receives an opinion it gives more credit to it, i.e., a higher score to this opinion. Under this new framework, we derive a convergence result which holds under mild assumptions on the way information is transmitted by agents and leads to consensus in a particular case. We also provide some numerical results illustrating the formation of consensus under different topologies (complete and ring graphs) and different initial conditions (random and biased towards a specific content).

I. INTRODUCTION

Consider a network of agents sharing opinions about some contents. These contents might be movies, books, political leaders, etc. The network could consist of a small group of agents chatting in the same room or a large social network. Literature on opinion dynamics [1], [2], [4], [6], [7], [8] propose simple models at the agents’ level and analyze their impact at the network level: do these models lead to consensus, spontaneous clustering, etc.? A celebrated model was proposed by Hegselmann and Krause [8], and several other models have also been proposed. Interestingly enough, just like simple statistical physics models can lead to deep technical issues (see e.g. [3]), very simple agent behaviour can hide mathematical challenges too, leading to fascinating conjectures [8].

In most opinion dynamics models, the "opinion" is either a scalar value [8] or a vector [5], [10], [14]. In gossip models, for instance, agents perform linear combinations of their own opinion with those of their neighbors [15]. In the Hegselmann and Krause model, agents only perform averages with agents sharing a close opinion, in which case the network topology can be seen as state-dependent. These models have in common the underlying assumption that agents are concerned about some unique, possibly complex, "quantity" which form their opinion: it might be a scalar score [8], a velocity [4], a momentum, etc.

In this work, we assume that agents opinions deal with multiple distinct contents. This setting deeply differs from the abovementioned ones in the sense that, since resources are limited, competition naturally arises. Let us consider two friends chatting about movies: if they do not have a large amount of time, they will probably talk about movies that made the strongest impression on them. The model we propose here has two separate building blocks. First, each agent has a probability distribution over the contents reflecting its tendency to communicate about them. Second, this probability distribution comes from scores attributed to each content as in the Luce-Plackett model [12]. The dynamics takes place at the score level: the more an agent is exposed to a given content, the higher is this content score. The competition between contents is automatically taken into account by the fact that probability distributions must sum to one.

The contributions of this paper are the following. First, to the best of our knowledge, the model we put forward is new. It can be linked to recent works on gossip, where gossip takes place only over the k largest components of a vector items [13], but it significantly differs from top-k gossip in that, instead of taking averages on the k strongest opinions, scores are used to poll agents opinions, then incremented. A direct byproduct of this distinction is that the score process is non-decreasing, hence it cannot reach a consensus. However, and this is the second contribution of this paper, we show that the probability distributions attached to the score process do converge to a consensus under mild assumptions.

The rest of the paper is organized as follows. Section II introduces notations and the proposed opinion dynamics model. Section III details the main results of the paper, regarding the convergence of the opinion dynamics algorithm. In Section IV numerical experiments are performed and Section V concludes the paper.

II. MODEL

A. Notations

For a matrix $X \in \mathcal{M}_{N \times K}(\mathbb{R})$, we define $|X|$ as the diagonal matrix with entries $|X|_{ii} = \sum_{k=1}^{K} |X_{ik}|$, $\forall i \leq N$ and $||X||$ the norm of $X$ such that $||X|| = \max_{i \leq N} \max_{k \leq K} |X_{ik}|$. Also we define $||x||$ as the euclidian norm of a vector $x \in \mathbb{R}^{d}$, i.e.: $||x|| = (\sum_{i=1}^{d} x_{i}^{2})^{1/2}$.

For an undirected graph $G = (V, E)$ with $N$ nodes, we define its adjacency matrix $A$ as the $N \times N$ matrix such that $A_{ij} = 1$ if node $i$ has an edge in $E$ to node $j$ and $A_{ij} = 0$ if not, and $N(i) = \{ j \mid A_{ij} = 1 \}$ the set of neighbours of agent $i$. Also we define $D$ as the diagonal matrix with entry $D_{ii} = \sum_{j=1}^{N} A_{ij}$ the degree of node $i$. 

¹Institut Mines-Telecom, Telecom SudParis, UMR CNRS 5157, France
julio.louzada.pinto@telecom-sudparis.eu
tijani.chahed@telecom-sudparis.eu
jeremie.jakubowicz@telecom-sudparis.eu
We define the \((K - 1)\)-dimensional simplex \(\Delta_K = \{x \in \mathbb{R}_+^K \mid \|x\| = \sum_{k=1}^K x_k = 1\}\) and \((\Delta_K)^N\) the space of \(N \times K\) real matrices such that every row is in \(\Delta_K\) (the space of \(N \times K\) real stochastic matrices), i.e.: \((\Delta_K)^N = \{M \in \mathcal{M}_{N \times K}(\mathbb{R}_+) \mid \sum_{k=1}^K M_{ik} = 1, \forall i \leq N\}\).

We denote \(T\) the uniform vector in \(\mathbb{R}^d\), i.e.: \(T^i = 1, 1 \leq i \leq d\) and \(I_d\) the identity matrix in \(\mathbb{R}^d\). We also say that a sequence \(x_n \in \mathbb{R}^d\) converges to a set \(A \subset \mathbb{R}^d\) if \(\lim_{n \to \infty} \inf_{y \in A} \|x_n - y\| = 0\).

**B. Agents’ contents**

Let us consider an opinion dynamics scenario where several agents exchange information about \(K\) contents. Each content is denoted by an integer \(1 \leq k \leq K\). Each agent is not necessarily able to communicate with all the other agents. We assume that there is a communication undirected graph \(G = (V, E)\), where \(V\) is \(\{1, \ldots, N\}\) is the set of agents and \(E\) is a subset of all possible pairs \(\{i, j\}\) such that \(\{i, j\} \in E\) if and only if agent \(i\) is able to communicate with agent \(j\) (note that in this model, whenever agent \(i\) is able to communicate with agent \(j\), then agent \(j\) is also able to communicate with agent \(i\)).

**C. Time model**

We follow the standard discrete model stemming from individual Poisson clocks ([16], [17]): each agent \(i\) can exchange information with its peers when its clock ticks. The \(n^{th}\) tick is denoted by \(T^i_n \in \mathbb{R}_+\) and is such that, with an abuse of language, \(T^i_n\) follows a Poisson process with constant intensity \(\lambda > 0\) (we assume all clocks have the same intensity, for the sake of simplicity). In this case, it is equivalent to consider a common Poisson clock \(T_n\) obtained by merging all the individual clocks on a common timescale that ticks with intensity \(N\lambda\). From now on, time is thence supposed to be discrete and represented by \(\mathbb{N}\); and we think about time \(t \in \mathbb{N}\) as the global \(t^{th}\) event that has occurred in the network.

**D. Scores**

Each agent \(i\) has a specific score \(X^i_k \in \mathbb{R}_+\) for content \(k\) at time \(t\). The higher \(X^i_k\), the higher is the appreciation for content \(k\) by agent \(i\) at time \(t\). We denote by \(X^i_t = (X^i_{t,k})_{1 \leq k \leq K}\) the collection of all scores of agent \(i\) at time \(t\).

**E. Opinion dynamics model**

All agents communicate about the contents. We assume that, for each agent \(i\), the higher the score for a specific content \(k\), the more likely agent \(i\) is to communicate about \(k\) with its peers. Specifically, we can define \(X_t \in \mathcal{M}_{N \times K}(\mathbb{R}_+)\) as the matrix of scores of all agents, such that \(X^i_{t,k}\) is the score of agent \(i\) over content \(k\) and we assume that at time \(t\), agent \(i\) chooses to communicate about content \(k\) according to a random variable \(u^i_t\), with distribution

\[
P(u^i_{t+1} = k) = (f(P_t))_{ik}, \quad 1 \leq k \leq K, \quad (1)
\]

where \(P_t \in (\Delta_K)^N\) is the matrix of preferences defined as \(X_t = |X_t|P_t\) and \(f : (\Delta_K)^N \to (\Delta_K)^N\) is a measurable function that accounts for possible nonlinearity in the model.

For instance, \(f(P) = \frac{\exp(\beta P)}{\sum_{i=1}^N \exp(\beta P_{i1})}\), with \(\beta \gg 1\) gives a softmax-like behaviour, \(\beta \ll 1\) models the indifference of agents to contents or \(f(P) = P\) is the identity and each content \(k\) has a probability \(P^i_{t,k}\) of being broadcasted at time \(t\) by agent \(i\).

Scores are then updated as

\[
X^i_{t+1} = X^i_t + \sum_{j \in N(i)} \mathbb{I}_{\{u^i_{t+1} = k\}} \mathbb{I}_{\{n^j_{t+1} = j\}}, \quad (2)
\]

where \(n^i_{t+1}\) is a sequence of i.i.d. random variables with values in \(\{1, \ldots, N\}\) responsible for selecting the agent that broadcasts the information at each time \(t\). We can write the uploading mechanism (2) in matrix form as

\[
X_{t+1} = X_t + U_{t+1}, \quad (3)
\]

where \(U_{t+1} = AU_{t+1}\) with \(U_{t+1} = \sum_{i=1}^N \mathbb{I}_{\{n^i_{t+1} = i\}} = \mathbb{I}_{\{n^i_{t+1} = i\}}.\)

See Figure 1.

---

**III. CONVERGENCE RESULTS**

When studying the opinion dynamics algorithm (3), an important feature worth investigating is the formation of a consensus, i.e.: a state where all agents have the same scores. We cannot expect that scores converge to a stationary state, because at each time agents update their scores in an unbounded manner. On the other hand, we can look at the probabilities that agents broadcast information at each time instant, i.e.: their preferences \(P^i_{t,k}\). These quantities are
retrieved from the scores $X_t^{i,k}$ by a normalization procedure, and hence are bounded.

Thus, we can expect convergence of the quantities $P_t^{i,k}$ over time towards a stationary state (which of course is stochastic). It is worth mentioning that qualitatively $X^{i,k}_t$ and $P_t^{i,k}$ represent the same thing: tastes of agents over contents. The only difference is that $X^{i,k}_t$ represent absolute scores of agents while $P_t^{i,k}$ represent preferences, i.e. relative scores, of agents.

The consensus is formally defined as follows.

**Definition 1:** (Reaching Consensus) We say that the discrete stochastic dynamical system (3) reaches a consensus if:

$$\max_{0 \leq k \leq K} |P_t^{i,k} - P_t^{j,k}| \rightarrow 0, \forall i, j \leq N \text{ almost surely}. \quad (4)$$

We make the following assumptions:

**Assumption 1:** (i) Let $G = (V,E)$ be a connected and undirected graph with $N$ nodes.

(ii) Let $f : (\Delta_K)^N \rightarrow (\Delta_K)^N$, $P \mapsto f(P)$ be a smooth function defined by Eqn. (1).

(iii) $|X_0^{i,k}| > 0$, $\forall 1 \leq i \leq N$.

Assumption 1 (i) could be refined taking into account time-varying network topologies but is stated for the sake of simplicity. Assumption 1 (ii) is a technical assumption that ensures that the underlying ODE (Ordinary Differential Equation) is well defined. Assumption 1 (iii) ensures that all agents have at least one opinion before starting the dynamics.

**Theorem 1:** Under Assumption 1 (i), (ii) and (iii) we have that

$$P_t \xrightarrow{t \to \infty} A_\omega \text{ almost surely},$$

where $A_\omega$ is some limit set of the ODE (where $P \in \mathcal{M}_{N \times K}(\mathbb{R}_+)$)

$$\dot{P} = D^{-1} A f(P) - P,$$

with $D$ the diagonal matrix with entry $D_{ii} = \sum_{j=1}^N A_{ij}$ the degree of node $i$.

**Proof:** The reader can see appendix V-B for the proof.

However, the limit sets involved in Theorem 1 can be arbitrarily intricate depending on the function $f$. In the case where $f$ is the identity, a much simpler result is available as shown by the next theorem.

**Corollary 1:** We assume (i) and (iii) and let $f : (\Delta_K)^N \rightarrow (\Delta_K)^N$, $P \mapsto f(P) = P$ be the identity.

Then the algorithm (3) reaches a consensus, i.e.: it satisfies Eqn. (4).

**Sketch of proof:** Applying Theorem 1 we get that $V(P_t)$ converges almost surely to some limit set of equation

$$V(\dot{P}) = -(I_K \otimes D^{-1} \Delta) V(P), \quad (5)$$

where $V(P)$ is the vectorization of the matrix $P$, $\otimes$ is the Kronecker product and $\Delta = D - A$ the Laplacian of $G$.

The proof stems from the fact that $(I_K \otimes D^{-1} \Delta) x, x$ is a Lyapunov function associated to the ODE $V(\dot{P}) = -(I_K \otimes D^{-1} \Delta) V(P)$.

IV. Numerical Experiments

In this section we perform some numerical simulations under the hypothesis of corollary 1.

Figure 2 shows randomly generated initial scores of agents for the case of a complete graph: each line corresponds to one initial set of agent scores for some 12 contents shown on the x-axis. Figure 3 shows the corresponding obtained consensus in which case all the lines condensate to just one which shows the asymptotic preferences of the agents.

The same experiment is reproduced in Figures 4 and 5 for the case of a ring graph with random initial conditions, and in Figures 6 and 7 for the case of a ring graph but with initial conditions biased towards a specific content.

V. Conclusion

We introduced in this paper a new opinion dynamics model which is able to incorporate opinions about multiple contents and creates an intrinsic competition. We showed a general convergence result for the preferences of agents, based on the ODE method of stochastic approximations. We also showed a first result on consensus formation under mild assumptions on the way information is broadcasted by agents. The consensus in this case does not depend on the structure of the network, i.e.: every connected network converges to a consensus that depends, of course, on the initial scores of agents and on the topological properties of
the network itself. We believe that several extensions can stem from this first, simple model and would allow the study of opinion dynamics at a much finer scale.

APPENDIX

Let \((A_i)_{i \in I}\) be a union (maybe uncountable) of random variables. Then \(\sigma(A_i, i \in I)\) is the smallest \(\sigma\)-algebra generated by these random variables, when well defined. This notation also include filtrations, i.e.: let \(\mathcal{F}_t\) be a filtration (with respect to some probability space) and \((A_i)_{i \in I}\) be a union of random variables. Then \(\sigma(\mathcal{F}_t, A_i, i \in I)\) is the smallest \(\sigma\)-algebra generated by \(\mathcal{F}_t\) and \((A_i)_{i \in I}\), when well defined.

A. Results needed for convergence

We fix a probability space \((\Omega, \mathcal{O}, \mathbb{P})\). We define the following Stochastic Approximation Algorithm:

\[
\sum_{n=0}^{\infty} \epsilon_n = \infty, \quad \epsilon_n \geq 0, \quad \epsilon_n \to 0 \quad \text{for} \quad n \geq 0.
\]  

\[
P_{n+1} = P_n + \epsilon_n Y_n, \quad P_n \in \mathbb{R}^r.
\]

We define the interpolated and shifted processes:

\[
t_n = \sum_{i=0}^{n-1} \epsilon_i,
\]

\(m(t)\) is the unique value of \(n\) such that \(t_n \leq t < t_{n+1}\),

\[
P^0(\omega, t) = P_n(\omega), \quad \text{for} \quad t_n \leq t < t_{n+1}, \quad \omega \in \Omega,
\]

\[
P^n(\omega, t) = P^0(\omega, t_n + t), \quad \omega \in \Omega.
\]

The method used to prove the almost sure convergence is the ODE method, a classical tool in stochastic algorithms. We also state some technical conditions (hypotheses) for algorithm (7) to work (see [9]):

\[
\sum_n n^2 < \infty.
\]

\[
\sup_n \mathbb{E}[|Y_n|] < \infty.
\]

\[
\mathbb{E}[Y_n | P_0, Y_i, i < n] = g_n(P_n, \xi_{n+1}) + \beta_n,
\]

where \(\xi_{n+1}\) and \(\beta_n\) are random variables\(^1\) and \(g_n(P, \xi)\) are continuous functions in \(P\) for each \(\xi\) and \(n\).

There exists a continuous function \(g\) such that for each \(P\)

\[
\lim_n \mathbb{P} \left( \sup_{n, t \leq t} \left| \sum_{i=m(T)}^{m(T+1)-1} \epsilon_i g_i(P, \xi_{i+1}) - g(P) \right| \geq \mu \right) = 0
\]

\(^1\)Due to the vast generality of \(\xi_{n+1}\) and the reminders \(\beta_n\), we do not put any hypotheses over the measurability of these random variables. Suitable assumptions will come naturally as we proof our results.
for every $\mu > 0$ and some $T > 0$. For every $\mu > 0$ and some $T > 0$, we have
\[
\lim \mathbb{P} \left( \sup_{n \geq 2} \max_{t \leq T} \left| \sum_{i=1}^{m_1(T+t)} \varepsilon_i \delta M_i \right| \geq \mu \right) = 0, \tag{13}
\]
where $\delta M_i = Y_i - E[Y_i | P_0, Y_i, l < i]$ is a martingale difference. For every $\mu > 0$ and some $T > 0$, we have
\[
\lim \mathbb{P} \left( \sup_{n \geq 2} \max_{t \leq T} \left| \sum_{i=1}^{m_1(T+t)} \varepsilon_i \beta_{l,i} \right| \geq \mu \right) = 0. \tag{14}
\]
There are measurable and non-negative functions $\rho_3(P)$ and $\rho_{4}(\xi)$ such that
\[
\|g_n(P, \xi)\| \leq \rho_3(P)\rho_{4}(\xi), \tag{15}
\]
where $\rho_3$ is bounded on each bounded $P$-set, and for each $\mu > 0$,
\[
\lim \lim_{\tau \to 0} \mathbb{P} \left( \sup_{j \geq n} \sum_{i=m(j,\tau)}^{m(j+\tau-1)} \varepsilon_i \rho_{4}(\xi_{i+1}) \geq \mu \right) = 0. \tag{16}
\]
There are non-negative measurable functions $\rho_1(P)$ and $\rho_{2}(\xi)$ such that $\rho_1$ is bounded on each bounded $P$-set and
\[
\|g_n(P, \xi) - g_n(y, \xi)\| \leq \rho_1(P - y)\rho_{2}(\xi), \quad \rho_1(P) \overset{P \to 0}{\to} 0 \tag{17}
\]
and
\[
\lim \mathbb{P} \left( \sup_{j \geq m(t+\tau)} \sum_{i=j}^{m(t+\tau)} \varepsilon_i \rho_{2}(\xi_{i+1}) < \infty \right) = 1 \tag{18}
\]
for some $\tau > 0$.

**Theorem 2 (Theorem 1.1, pg. 166 of [9]):** Let (6) and (9)-(18) hold for algorithm (7). Then there is a set $N$ of probability zero such that for $\omega \notin N$, the set of functions $P^n(\omega, \cdot), n < \infty$ is equicontinuous. Let $P(\omega, \cdot)$ denote the limit of some convergent subsequence. If $\{P_n\}$ is bounded with probability one, then it satisfies the ODE
\[
\dot{P} = g(P), \tag{19}
\]
and $\{P_n(\omega)\}$ converges to some limit set of this ODE in some bounded invariant set.

**Lemma 1:** Let $X_i$ be a sequence of random matrices such that $\sup_{i} \|X_i\| \leq K$ almost surely. Then for every $T > 0$ we have
\[
\lim_{t \to \infty} \frac{\sum_{i=0}^{t} X_i}{t} = \lim_{t \to \infty} \frac{\sum_{i=0}^{t} X_i}{t} = TK \text{ implies}
\]
\[
\lim_{t \to \infty} \frac{\sum_{i=0}^{t} X_i}{t} = \lim_{t \to \infty} \frac{\sum_{i=0}^{t} X_i}{t} = \lim_{t \to \infty} \frac{\sum_{i=0}^{t} X_i}{t}, \tag{19}
\]
and
\[
\lim_{t \to \infty} \frac{\sum_{i=0}^{t} X_i}{t} \geq \lim_{t \to \infty} \frac{\sum_{i=0}^{t} X_i}{t} = \lim_{t \to \infty} \frac{\sum_{i=0}^{t} X_i}{t}. \tag{19}
\]

**Lemma 2:** Define $\varepsilon^i_t = |X^i_t|$ with $|X^i_0| = K^i > 0$. We have that $\frac{1}{\varepsilon^i_t} \to \frac{1}{\sum_{j=1}^{N} A_{ij}}$ almost surely when $t \to \infty$. In matrix form, we have that
\[
(t + 1)|X_{t+1}|^{-1} \to ND^{-1} \text{ almost surely when } t \to \infty,
\]
where $D$ is a diagonal matrix with entry $D_{ii} = \sum_{j=1}^{N} A_{ij}$ the degree of node $i$.

**Proof:** We have that $\varepsilon^i_t = |X^i_t| = K^i + \sum_{l=1}^{t} Z^i_l$, where $Z^i_l = \sum_{j=1}^{N} A_{ij} \sum_{n=1}^{j} Z^i_n$ are i.i.d random variables with $E[Z^i_l] = \frac{1}{N} \sum_{j=1}^{N} A_{ij}$ by Eqn. (20). By the Strong Law of Large Numbers we have that $\frac{\varepsilon^i_t}{t} \to \frac{1}{N} \sum_{j=1}^{N} A_{ij}$ almost surely. By the continuity of $x \to \frac{1}{x}$ in $[K^i, \infty)$ we have that $\frac{1}{\varepsilon^i_t} \to \frac{1}{\sum_{j=1}^{N} A_{ij}}$ almost surely when $t \to \infty$. \hfill \blacksquare

**B. Proof of theorem 1**

Eqn. (3) can be written as
\[
|X_{t+1}|P_{t+1} = |X_t|P_t + AU_{t+1},
\]
where $P_t$ is a $N \times K$ matrix with agents preferences $P_{t,i}^k$ at time $t$ and $U_{t+1}$ is a random matrix accounting for the updating of the algorithm. Since at each time $t$ on only one agent broadcasts its information, we define a sequence of i.i.d random variables $n_{t+1}$ with values in $\{1, \cdots, N\}$ that is responsible for selecting the agent that broadcasts the information at each time $t$. Since each agent has a Poisson clock with the same intensity, a simple calculation gives us that, conditional to an event on the system (a "tick" of some Poisson clock), the probability that the event was triggered by agent $i$ is
\[
\mathbb{P}(n_{t+1} = i) = \frac{\lambda_i}{\sum_{j=1}^{N} \lambda_j} = \frac{1}{N}. \tag{19}
\]
Hence, the matrix $U_{t+1}$ has entries $(U_{t+1})_{ik} = \mathbb{I}_{n_{t+1} = i} \mathbb{I}_{n_{t+1} = k}$. We denote by $U_{t+1}^j$ the $j$th row of matrix $U_{t+1}$ and $F_t = \sigma(X_s, s \leq t)$ the filtration generated by the past of the system. We have then
\[
E[U_{t+1} | F_t] = E[E(U_{t+1} | \sigma(n_{t+1}, F_t)) | F_t]
\]
\[
= E[\frac{1}{N} \sum_{j=1}^{N} U_{t+1} | n_{t+1} = j] | F_t]
\]
\[
= \frac{1}{N} \sum_{j=1}^{N} E[U_{t+1}^j | F_t] = \frac{1}{N} f(P_t)
\]
by the independence of $n_{t+1}$ and $F_t$. Hence $\zeta_t = AU_{t+1} - \frac{1}{N} Af(P_t)$ is a bounded martingale difference. Eqn. (3) resolves to
\[
|X_{t+1}|P_{t+1} = |X_t|P_t + AU_{t+1} = |X_t|P_t + \zeta_t + \frac{1}{N} Af(P_t)
\]
\[
\Rightarrow |X_{t+1}|(P_{t+1} - P_t) = (|X_t| - |X_{t+1}|)P_t + \zeta_t + \frac{1}{N} Af(P_t)
\]
Recall that $D$ is a diagonal matrix with entry $D_{ii} = \sum_{j=1}^{N} A_{ij}$ the degree of node $i$. Then

\[
P_{t+1} = P_t + |X_t+1|^{-1}(|X_t| - |X_{t+1}|)P_t + \frac{1}{N} A f(P_t) + \zeta_t
\]

\[
= P_t + \frac{1}{t+1} \left( (t+1)|X_t+1|^{-1}(|X_t| - |X_{t+1}|)P_t + (t+1)|X_{t+1}|^{-1} \frac{1}{N} A f(P_t) + ND^{-1} \zeta_t + ((t+1)|X_{t+1}|^{-1} - ND^{-1}) \zeta_t \right)
\]

\[
= P_t + \frac{1}{t+1} \left( - (t+1)|X_t+1|^{-1}(|X_t| - |X_{t+1}|)P_t + D^{-1} A f(P_t) + ND^{-1} \xi_t \right.
\]

\[
+ ((t+1)|X_{t+1}|^{-1} - ND^{-1}) \left( \frac{1}{N} A f(P_t) + \zeta_t \right)
\]

\[
= P_t + \frac{1}{t+1} \left( - g_t(P_t, |X_{t+1}| - |X_t|) + ND^{-1} \xi_t + \beta_t \right).
\]

with $x \in \mathcal{M}_N \times \mathbb{R}_+$ and $\eta \in \mathcal{M}_N \times \mathbb{N}$. $g_t(x, \eta_t) = (t+1)|X_{t+1}|^{-1} \eta x - D^{-1} A f(x)$ and $\beta_t = ((t+1)|X_{t+1}|^{-1} - ND^{-1}) \left( \frac{1}{N} A f(P_t) + \zeta_t \right)$.

Firstly, note that $\eta_{t+1} = |X_{t+1}| - |X_t|$ are i.i.d. random variables such that

\[
E[(\eta_{t+1})^2] = \sum_{k=1}^{N} \sum_{j=1}^{N} A_{ij} E[w_{t+1}^j, 1] E[\eta_{t+1+j}] \tag{20}
\]

\[
= \sum_{j=1}^{N} A_{ij} E[\eta_{t+1+j}] = \frac{1}{N} \sum_{j=1}^{N} A_{ij} = \frac{1}{N} D_{ii}.
\]

We can write then

\[
E[|X_{t+1}| - |X_t|] = E[\eta_{t+1}] = \frac{1}{N} D.
\]

We need to prove that the hypothesis of theorem 2 are satisfied. Let us start with the reminder $\beta_t$.

**Lemma 3**: We have $\beta_t \to 0$ almost surely when $t \to \infty$.

**Proof**: Since $\zeta_t$ is bounded almost surely, $f$ is continuous and $P_t$ is bounded almost surely, the result stems from lemma 2.

We also need to prove that $g_t(x, \eta_{t+1})$ satisfies the hypothesis of theorem 2.

**Proposition 1**: We fix $x \in \mathcal{M}_N \times \mathbb{R}_+$ and we define

\[
\overline{g}(x) = D^{-1} \left( D x - A f(x) \right)
\]

where $D$ is a diagonal matrix with entry $D_{ii} = \sum_{j=1}^{N} A_{ij}$ the degree of node $i$. Then the sequence $g_t(x, \eta_{t+1})$ satisfies the Strong Law of Large Numbers to $\overline{g}(x)$, i.e.: $\lim_{T \to \infty} \frac{\sum_{t=0}^{T} g_t(x, \eta_{t+1})}{T} = \overline{g}(x)$ almost surely.

**Proof**: Fix a $x \in \mathcal{M}_N \times \mathbb{R}_+$ and $\varepsilon > 0$. Firstly notice that since $\eta_{t+1}x$ and $\frac{1}{N} A f(x)$ are bounded by some $K(x) < \infty$, lemma 2 guarantees the existence of a $T(\varepsilon) > 0$ such that

\[
\left| g_t(x, \eta_{t+1}) - (ND^{-1} \eta_{t+1}x - D^{-1} A f(x)) \right| < \varepsilon, \tag{21}
\]

for all $t \geq T(\varepsilon)$. Since $g_t(x, \eta_{t+1})$ are uniformly bounded almost surely, by lemma 1 we only need to prove the result for the sequence $2$ $T(\varepsilon)$ there is only a finite number of the sequence, and after the almost sure convergence of $\frac{\sum_{t=1}^{T(\varepsilon)}}{|X_{t+1}|}$ guarantees the boundedness.

Of notation, we continue to call this sequence $g_t(x, \eta_{t+1})$. Since $\eta_{t+1}$ are i.i.d.s, we have that

\[
\frac{\sum_{t=0}^{T} (ND^{-1} \eta_{t+1}x - D^{-1} A f(x))}{T} \to ND^{-1} \left( \mathbb{E}[\eta_{t+1}]x - D^{-1} A f(x) \right) = \overline{g}(x)
\]

almost surely when $T \to \infty$ by the Strong Law of Large Numbers. Eqn. (21) gives us

\[
\left| \frac{\sum_{t=0}^{T} g_t(x, \eta_{t+1})}{T} - \frac{\sum_{t=0}^{T} (ND^{-1} \eta_{t+1}x - D^{-1} A f(x))}{T} \right| < \varepsilon,
\]

which proves the desired result.

We now check the hypothesis of theorem 2: using the time step $\varepsilon_t = \frac{1}{t+1}$, we have hypothesis (6) and (9). One can see that hypothesis (14) is verified by lemma 3 and hypothesis (12) by proposition 1 (see [9]). To finish the proof of theorem 1, we can see that the rest of the technical hypothesis are satisfied because we are in a compact $\{ |P_t| \leq 1, \forall t \}$, thus $f$ is actually uniformly continuous and so are $g_t(x, \eta_{t+1})$ and $\overline{g}(x)$.

**REFERENCES**


