Consensus with Random Asynchronous Updating Model

Joshua A. McGinnis

May 16, 2023

1 Introduction

Consensus protocol is an area of research seeking to understand how individuals or nodes sometimes referred to as agents may come to agreement. DeGroot's classical paper, "Reaching a Consensus" [1], is good starting point for understanding consensus models more generally. DeGroot seeks to model opinion dynamics and how agents' interactions with each other will eventually lead them toward agreement. This structure will be explored in section one. However, as can be seen in [7], consensus protocol models may also be useful in studying the synchronization of robotic movement, vehicle formation, flocking, and more generally agreement among nodes in a network. In fact, consensus protocol is often looked at through the lens of networks and graphs as can be seen in [3] where consensus protocol was studied over random graphs. The gist of consensus protocol in discrete time is, after specifying some number of agents say d-many, to update each agents' quantity at time t with some function depending on time $f(t) : \mathbb{R}^n \to \mathbb{R}^n$ yielding the $t+1$ agents' quantities. One major area of focus in the study of consensus protocol is finding sufficient conditions for achieving consensus in general settings. This is also the focus here. In this note the function $f(t)$ is always linear as we are studying a modification of the DeGroot model.

Section two is used to introduce notation and the DeGroot model, so that readers without background may have a firm example of what is being studied. The following subsection introduces the modification to the DeGroot model, the focus of our study here. The following section, section three, reviews results from two of the most modern papers that are applicable to our model. We find that proving consensus for our modified DeGroot model is still out of reach of the results of both these papers. Therefore in section four we go about proving consensus for the model directly given a certain condition. The method of proof to some extent mirrors that of the first paper we review in section three. Section five discusses how the condition yielding consensus is a necessary to guarantee the consensus property of a matrix closely related to the model, see section (2.3) below. The final section will be used to introduce further study ideas, and thereafter concluding remarks are given.

Our altered DeGroot model exhibits an example where consensus is reached with reasonable assumptions, but is not covered by the general results of [6] and [5]. This suggest something unknown in the theory, and it would be nice to find a generalization of the new results in this note. However, the proof of consensus of the modified DeGroot model is a bit unsatisfactory because it is unclear how it might be generalized. This is discussed more in the conclusion.

2 Basic Notation, the DeGroot Model, and Modification

2.1 Basic Notation

In this section we introduce basic notation. For this work the model is of finite dimension d. We use $[d]$ to denote the set of indices $\{1, .., d\}$. We use $\mathbf{e_i}$ to denote the vector with the i^{th} component as one and zero otherwise. We also use $\mathbf{1}$ to denote the vector which has one in every entry. The row vector $\mathbf{x^T}$ is the transpose of $\mathbf{x}$. We use $\min(\mathbf{x})$ and $\max(\mathbf{x})$ to denote the minimum and maximum component of $\mathbf{x}$. Components are labeled $\mathbf{x_i}$. We say $\mathbf{x} \in \mathcal{M}$ if $\mathbf{1^T x} = \mathbf{1}$ and $\mathbf{x}$ has non-negative entries. All matrices should be taken from the set of d by d stochastic matrices $\mathcal{S}$. A stochastic matrix is defined here as matrices whose rows are all vectors in $\mathcal{M}$. i.e. if A is a matrix, $A\mathbf{1} = \mathbf{1}$. We denote the i^{th} column of a matrix as $\mathbf{A^i}$ and the i^{th} row of a matrix as $\mathbf{A_i}$. Notice these are vectors, and therefore are bold, They should not be confused with powers of matrices. Given a matrix P, we denote its entries as $p_{i,j}$, or sometimes for matrix $X(t)$, $X(t)_{i,j}$, where $i,j \in [d]$. To express the i,j entry of a product of matrices we might put brackets around it like so $[AB]_{i,j}$. There will also be be matrices which are indexed by subscripts, like P_i, introduced later.

2.2 DeGroot

We are considering a special case of a discrete time linear dynamical model driven by stochastic matrices. For time t, we have a state vector $\mathbf{x(t)}$ and it is updated to its state at time $t+1$ via some stochastic matrix $X(t)$

$$\mathbf{x(t+1)} = X(t)\mathbf{x(t)}. \tag{1}$$

It is well known that if $X(t) = P$ for all t, and P is irreducible and aperiodic, (1) admits consensus [1]. This is the essential setup to the DeGroot model. In the DeGroot model the vector $\mathbf{x(t)}$ represents the opinions of the agents at time t in the following way. Given event space Θ with measurable sets $\mathcal{F}$, $\mathbf{x(t)_i} : \mathcal{F} \to \mathbf{[0,1]}$ is subjective probability measure of the i^{th} agent. We can regard these subjective probabilities as opinions. An application of P to the vector might represent a meeting in which agents get together to discuss their opinions about event θ. Having been influenced by each other, the individuals update their opinion by taking convex combinations of the opinions of others. The i^{th} row of P is how the i^{th} agent weights the opinions of others. For the time being, it is equivalent and perhaps simpler to think of $\mathbf{x(t)}$ to denote a vector of real numbers rather than measures. It should be noted that often it is said a matrix P admits consensus or reaches consensus. This means simply that if $X(t) = P$ for all t in (1), then consensus is reached. We can now define what consensus is.

Definition 1 *We say that that the system (1) admits consensus if for all* $\mathbf{z}$ *with* $\mathbf{x(0)} = \mathbf{z}$*, there exists a constant* $c(\mathbf{z})$ *such that*

$$\lim_{t\to\infty} ||\mathbf{x(t)} - c(\mathbf{z})\mathbf{1}|| = 0.$$

In (1), $X(t)$ might be chosen randomly. Let $\Omega_0 = \mathcal{S}$ and $\mathcal{B}$ be the Borel sigma algebra of Ω_0 and μ a probability measure. Define the product probability space as $(\Omega, \mathcal{F}, Pr) = \prod_{i=1}^{\infty}(\Omega_0, \mathcal{B}, \mu)$ where

$$\Omega = \{\omega = (\omega(1), \omega(2), ...) \mid \omega(k) \in \Omega_0\}$$

$$\mathcal{F} = (\mathcal{B} \times \mathcal{B} \times ...)$$

$$Pr = \mu \times \mu \times ...$$

The coordinates of ω are thus i.i.d stochastic matrices. To see this independence, let M and N be Borel. $Pr(\{\omega \mid \omega(k) \in M\} \cap \{\omega \mid \omega(j) \in N\}) = \mu(\Omega_0) \times ... \times \mu(M) ... \times \mu(N) \times \mu(\Omega_0) ... = Pr(\{\omega \mid \omega(k) \in M\})Pr(\{\omega \mid \omega(j) \in N\})$. This result can be extended to any number of coordinates thus proving independence. It should be clear that the coordinates are identically distributed since each coordinate has the same probability measure. Now we consider the probabilistic model

$$\mathbf{x}(k+1) = X(k)(\omega)\mathbf{x}(k). \tag{2}$$

Here $X(k)(\omega) = \omega(k)$. The above equation represents a generalization of the model that we deal with in this note. There is already much known about the model. In section (3.1) and (3.2) we review results as to when equation (2) reaches consensus and how this applies to the model introduced in this note. Whenever equation (2) is referred to it should be taken in conjunction with the related measure theoretic equations above it. Furthermore we define consensus in this random setting as

$$\lim_{t\to\infty} ||\mathbf{x(t)}(\omega) - c(\mathbf{z})\mathbf{1}|| = 0$$

where now of course c is a random variable associated with each ω. We will from now on drop ω in order to simplify notation.

2.3 The Random Asynchronous Update Model

In this section we introduce the modified DeGroot Model henceforth called the Random Asynchronous Update Model or RAUM for short. Suppose we have some aperiodic and irreducible matrix P. We wish to explore what happens if we assume that agents enumerated by indices $\{1, ..., d\} = [d]$ are updated one at a time instead of simultaneously and furthermore are chosen at random. To that end we define for $i \in [d]$

$$P_i = I + \mathbf{e}_i \mathbf{e}_i^T (P - I).$$

The matrix P_i retains the weights of the i^{th} row of P but acts as identity on all other agents. Since we need a matrix P in order to have a RAUM, we say a RAUM is induced by P and invoking (2), have that $\mu(P_i) > 0$ for all i and zero otherwise. What is eventually shown is that consensus is reached almost surely, whenever P admits consensus.

3 Convergence with Known Techniques

In this section we look at the results of two different papers [5] and [6]. We look at the results of [5] because the result gives a simple characterization of consensus in the special case when the diagonals of $X(k)$ are positive almost surely. The proof technique parallels the proof technique of theorem (2) below. In the second subsection we look at the primary result of [6], which generalizes the notion of positive diagonal entries with the notion of weak feedback. We do not provide any sort of proofs or the results in [6], because the proofs are long, arduous, and dizzying, but curious readers are encouraged to check the source. In both cases, we find that the results are not sufficient to prove consensus in certain RAUM.

3.1 Results with positive diagonal entries

We first make the assumption that for all i, $\mu(P_i) > 0$. It is fairly obvious that if $\mu(P_i)$ were positive for only one value of i that consensus would not be reached in any case where $d > 2$. Therefore we exclude any degenerate cases where some agents never update. Further we note that the expected value $E(X(t)) = \sum_{i=1}^{d} \mu(P_i) P_i$ has positive values in at least the same entries of that of P. Thus P being irreducible means $E(X(t))$ is also irreducible. The fact that each P_i acts as identity on most agents means $E(X(t))$ must also be aperiodic even when P is only irreducible. Therefore P is primitive i.e. there exists integer t_0 such that P^{t_0} has positive entries. Keeping these facts in find, we can now state the result of [5]. We state a slightly weaker version of the theorem in our own words here and prove it. The proof will skimp on details since it is very similar to the proof that the RAUM reaches consensus in section (4). Still we include the proof to take note of the similarity and the importance of positive diagonal entries. The complete version can of course be found in [5], but the most essential part of the proof for this note is what follows.

Theorem 1 *Let $X(t)$ in (2) have positive diagonal values almost surely. If $E(X(t))^{t_0}$ has positive entries, then (2) reaches consensus.*

Note that $E(X(t)...X(1)) = E(X(t))...E(X(1))$. Thus choosing t_0 sufficiently large guarantees that $E(X(t_0)...X(1))$ has all positive entries. Thus with positive probability $[X(t_0)...X(1)]_{i,j} > 0$ where i, j are arbitrary, fixed, and distinct. Since the diagonal entries are positive almost surely, if $[X(t_0)...X(1)]_{i,j} > 0$ then so is $[X(t_0 + k)...X(t_0)...X(1)]_{i,j} > 0$ for all integers k. Since i and j were arbitrary we can expect with positive probability that $[X(2t_0)...X(t_0 + 1)]_{k,i} > 0$. It then follows that for all i and fixed j the probability $[X((d-1)t_0)....X(1)]_{i,j} > 0$ is positive. Thus with we have shown that with positive probability there exists a column with non-zero entries. The remainder of the proof looks identical to the proof in section (4), so we will only outline it. We see below that a matrix with positive entries on some column must contract the diameter of any vector by a quantifiable amount. The Borel Cantelli Lemma can thus be applied since the system is i.i.d. to say that this contraction happens infinitely often, implying consensus. □

Again, the original result is actually slightly stronger than what is proven here and can be found in [5]. It can actually be shown that the expected matrix reaching consensus is equivalent to the model reaching consensus when diagonal entries are positive almost surely. The argument extends a little the argument above, and it is clear that having positive diagonal entries is essential. It is easy to exhibit a RAUM that reaches consensus and has missing diagonal entries. Therefore we move onto a result which generalizes the idea of positive diagonal entries.

3.2 Results with weak feedback

Now we use previously developed techniques in [6] which are very powerful and general, but ultimately we need to apply different tools to prove the desired result. In this case we consider a RAUM with uniform distribution on $\{P_1, ..., P_i\}$. The case with uniform distribution has nice properties worth exhibiting, even if not as general as we like. It is not difficult to observe that for all t

$$E(X(t)) = (1/d)\sum_{i=1}^{d} I + \mathbf{e}_i\mathbf{e}_i^T(P - I) = \frac{(d-1)I + P}{d}$$

The significance of this fact is that in expectation the system reaches consensus, if P is irreducible but not necessarily aperiodic. Theorem (7) in [6] tells us that a system reaches consensus if and only if the expected system reaches consensus, when the following three hypotheses are satisfied: (I) each $X(t)$ is independent, (II) the system has a common non-zero steady state in expectation, and (III) the system has the weak feedback property. (I) and (II) follow immediately from the model being i.i.d. We define the notion of weak feedback here for completeness, but it can be found in definition (2) of [6].

Definition 2 *The model has the weak feedback property if there exists a $\gamma > 0$ such that*

$$E[(\mathbf{X^i(t)})^T\mathbf{X^j(t)}] \geq \gamma(E[X(t)_{i,j}] + E[X(t)_{j,i}])$$

for all $t \geq 0$ and $i \neq j$.

One way we can think of weak feedback is as it being nearly a generalization of strictly positive diagonal entries $X(t)$ as seen in [5] and as seen in the previous subsection, except the diagonal entries must be bounded away from zero. We state this as a lemma and prove it here.

Lemma 1 *Considering equation (2), suppose that for all k and i, $X(k)_{i,i} \geq \gamma$ a.s. where $\gamma > 0$. Then (2) satisfies the weak feedback property.*

Since the model is i.i.d., we will write $X(k) = X$ for convenience. Then $X_{i,i}X_{i,j} \geq \gamma X_{i,j}$ almost surely. Also $X_{j,j}X_{j,i} \geq \gamma X_{i,j}$. It then follows almost surely that

$$\mathbf{X^i}^T\mathbf{X^j} \geq X_{i,i}X_{i,j} + X_{j,j}X_{j,i} \geq \gamma(X_{i,j} + X_{j,i})$$

Taking expectation yields the result. □

Proposition 1 *An i.i.d. RAUM model induced by a matrix P, satisfies the weak feedback property, if $\mathbf{e}_i^T P^T P\mathbf{e}_j$ is non-zero whenever $p_{i,j} + p_{j,i}$ is non-zero.*

For $i \neq j$, we calculate

$$E[\mathbf{X}^i(t))^T\mathbf{X}^j(t)] = (1/d)\sum_{s=1}^{d}\mathbf{e}_i^T P_s^T P_s\mathbf{e}_j = (1/d)\mathbf{e}_i^T(\sum_{s=1}^{d} P_s^T P_s)\mathbf{e}_j$$

$$= (1/d)\mathbf{e}_i^T[\sum_{s=1}^{d}(I + (P^T - I)\mathbf{e}_s\mathbf{e}_s^T)(I + \mathbf{e}_s\mathbf{e}_s^T(P - I))]\mathbf{e}_j$$

$$= (1/d)\mathbf{e}_i^T[\sum_{s=1}^{d} I + \mathbf{e}_s\mathbf{e}_s^T(P - I) + (P^T - I)\mathbf{e}_s\mathbf{e}_s^T + (P^T - I)\mathbf{e}_s\mathbf{e}_s^T(P - I)]\mathbf{e}_j.$$

Now summing over all the terms leaves

$$(1/d)\mathbf{e}_i^T(dI + P - I + P^T - I + (P^T - I)(P - I))\mathbf{e}_j.$$

$$= \mathbf{e}_i^T \frac{(d-1)I + P^T P}{d} \mathbf{e}_j = \frac{\mathbf{e}_i^T P^T P \mathbf{e}_j}{d}.$$

Now let $u = \min_{i,j}\{\mathbf{e}_i^T P^T P \mathbf{e}_j > 0\}$, let $v = \max_{i,j}(p_{i,j} + p_{j,i})$, and set $\gamma = u/v$. It thus follows that for all k

$$E[X^i(t))^T X^j(t)] = \frac{\mathbf{e}_i^T P^T P \mathbf{e}_j}{d} \geq u/d = \frac{uv}{dv} \geq \gamma \frac{p_{i,j} + p_{j,i}}{d} = \gamma(E[X(t)_{i,j}] + E[X(t)_{j,i}]),$$

and therefore shows that the model has the weak feedback property under certain conditions. □

We would like to demonstrate that the previous technique is not sufficient. Consider the example where we take

$$P = \begin{bmatrix} 0 & 1/3 & 2/3 \\ 0 & 1/3 & 2/3 \\ 1/2 & 0 & 1/2 \end{bmatrix}.$$

The matrix is clearly aperiodic and irreducible, and such a matrix admits consensus. However the i.i.d. RAUM induced by the matrix P will not have the weak feedback property.

4 Main Consensus Results

4.1 Consensus Result

In this section we build up to the result that, if P is irreducible, then the RAUM induced by P reaches consensus. It helps to note that aperiodicity will not help us in the proof and is not a necessary condition. An example of this can be seen, by taking P to be a two by two anti-diagonal stochastic matrix. We save for the second subsection, the proof that if P reaches consensus, so too does the RAUM. For now, we make the argument that consensus of the RAUM induced by P depends on a directed spanning tree, defined below, existing in P. An irreducible P will always have such a tree. As a warm up we will remind the reader of the total variation and diameter of a vector and Dobrushin coefficient. The latter describes the contractive nature of a matrix. First we begin with a definition of the diameter of a vector.

Definition 3 *Let $\mathbf{v} \in \mathbb{R}^d$, then $d(\mathbf{v}) = \max_i \mathbf{v}_i - \min_j \mathbf{v}_j$.*

The diameter will be one of our main tools for proving consensus. Next we define the total variation on stochastic vectors, which is just half the L_1 norm of the difference of the two vectors.

Definition 4 *Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors in $\mathcal{M}$. Then we define $||\mathbf{u}-\mathbf{v}||_{TV} = ||\mathbf{u}-\mathbf{v}|| = (1/2)\sum_{i=1}^{d} |\mathbf{u}_i - \mathbf{v}_i|$.*

The following lemma can be found in many intro to probability books such as [8]. There is also a proof of it in the appendix.

Lemma 2 *Let $\mathbf{v}$ and $\mathbf{w}$ be in $\mathcal{M}$, then $(1/2)||\mathbf{v}-\mathbf{w}|| = \sum_{i=1}^{d} |\mathbf{v}_i - \mathbf{w}_i| = \sum_{i\in(+)} (\mathbf{v}_i - \mathbf{w}_i) = -\sum_{i\in(-)} (\mathbf{v}_i - \mathbf{w}_i) \leq 1$, where $(+) = \{i \mid \mathbf{v}_i - \mathbf{w}_i \geq 0\}$ and $(-) = \{i \mid \mathbf{v}_i - \mathbf{w}_i < 0\}$.*

See the appendix for a proof. □

Now we define the Dobrushin coefficient $H(A)$ of a matrix A, which was known by Markov and others as a quantification for the contraction a stochastic matrix applies to a vectors diameter i.e. $d(A\mathbf{x}) \leq H(A)d(\mathbf{x})$ [4]. A proof of this can also be found in the appendix and relies on heavily lemma (2).

Definition 5 *Let A be a stochastic matrix and*

$$H(A) = (1/2) \max_{i,j \in \mathcal{S}} ||\mathbf{A_i} - \mathbf{A_j}||$$

Then H is what is known as the Dobrushin coefficient of A.

We are ready to define our concept of spanning tree. This spanning tree is what is essential to consensus. As a warning, the definition of directed spanning tree here might be different from those used in other literature.

Definition 6 *Suppose there exists an index i_0 such that for all indices $i_k \neq i_0$ in $[d]$, there exists indices $i_{k-1}, .., i_1$ s.t. $p_{i_k, i_{k-1}} > 0, p_{i_{k-1}, i_{k-2}} > 0, ...$ and $p_{i_1, i_0} > 0$. Then we say P contains a directed spanning tree with root i_0.*

Clearly P being irreducible implies we can find a directed spanning tree. Any index can be the root in this case. For the remaining arguments in this section, we are fixing some P with a spanning tree. Having found some index to be the root, we define subsets of the indexing set $[d]$ that we call the strata of the spanning tree for this fixed P. Note that rows and columns of P may be relabeled without affecting consensus of the RAUM. We can thus suppose without loss of generality that $i = 1$. Let $\mathcal{S}_1 = \{1\}$ and

$$\mathcal{S}_2 = \{i \in [d] \mid p_{i,1} > 0, i \neq 1\}$$

and recursively

$$\mathcal{S}_j = \{i \in [d] \mid \exists\, r \in \mathcal{S}_{j-1} \text{ s.t. } p_{i,r} > 0\} \cap \bigcap_{k=1}^{j-1} S_k^c.$$

Although we do not know a priori how many of these strata there are, we can be certain there are less than d of them. We next show that the set of sets $\{\mathcal{S}_1, ..., \mathcal{S}_N\}$ form a partition. We also call $S_j = \prod_{\text{inc } k \in \mathcal{S}_j} P_k$, where we are taking the product of increasing indices. It is arbitrary to choose increasing here so that we have some canonical form, but a product in any order would be permissible.

Lemma 3 *The set of strata $\{\mathcal{S}_1, ..., \mathcal{S}_N\}$ partition $[d]$.*

Let $j \in [d]$. Thus there exists distinct $i_1, .., i_r$ where $r < d$ such that $p_{j,i_1} > 0, p_{i_1,i_2} > 0, p_{i_2,i_3} > 0, ..., p_{i_r,1} > 0$ by the fact that 1 is the root of spanning tree in P. Thus all indices $j \in \bigcup_{k=1}^{N} \mathcal{S}_k$ and therefore must be in some $\mathcal{S}_k$. The sets are also clearly disjoint by construction. $\square$

We next work with the product $A = S_N \dots S_2$. One fact about this product is that the first row is $\mathbf{e_1}^T$. Also if $j \in \mathcal{S}_2$ then $A_{j,1} \geq p_{j,1}$. The first statement is easily seen since P_1 is not in the product. The second we show in the next lemma.

Lemma 4 *$A_{i,1}$ is positive for all $i \in [d]$.*

To show that $[S_2]_{j,1} > 0$, if $j \in \mathcal{S}_2$ we observe

$$\mathbf{e}_j^T S_2 \mathbf{e}_1 = \mathbf{e}_j^T P_j R \mathbf{e}_1 = (p_{j,1}\mathbf{e}_1^T + p_{j,2}\mathbf{e}_2^T + ... + p_{j,d}\mathbf{e}_d^T) R \mathbf{e}_1 \geq p_{j,1}$$

where the first equality is true since P_j occurs once in the product S_2. The second equality follows from the definition of P_j, and the final inequality from the fact that P_1 does not occur in the remaining product R.

The proof continues by induction. Let $j \in \mathcal{S}_{k+1}$. Then by hypothesis $[S_k \dots S_2]_{i,1} > 0$ for all $i \in \mathcal{S}_k$. In particular, there exists an i by construction such that $p_{j,i} > 0$. Therefore $[S_{k+1}]_{j,i} \geq p_{j,i}$. It follows that the $[S_{k+1} S_k ... S_2]_{j,1} \geq p_{j,i} [S_k ... S_2]_{i,1} > 0$. Finally it should be clear that since each index outside of one only occurs once in the product of A that $A_{1,i} > 0$ for all $i \in [d]$. $\square$

The lemma tells us the first column of A will be positive which means $H(A) < 1 - \delta$, where $1 \geq \delta > 0$. We are ready to give the final result.

Theorem 2 *Let P be a stochastic matrix that contains a directed spanning tree. Then the RAUM induced by P reaches consensus.*

Let $Y(1) = X(d-1)...X(1)$ and let $Y(k) = X(k(d-1))...X((k-1)(d-1)+1)$. Then the $\{Y(1), Y(2), ...\}$ are independent since $\{X(1), X(2), ...\}$ are. By the previous lemma we can find indices $i_1, ..., i_{d-1}$ such that $H(P_{i_{d-1}}...P_{i_1}) < 1-\delta$. Since convergence of the system is does not depend on labeling of the indices we can relabel the indices as $i_{d-1} = d, i_{d-2} = d-1, ..., i_1 = 2$ for convenience. Then for all k, we have that the probability $Pr[X(k(d-1)) = P_d, X(k(d-1)-1) = P_{d-1}, ..., X((k-1)(d-1)+1) = P_2] > 0$. Call $A = P_d...P_2$. Thus the $Pr(Y(k) = A) > 0$. Then $\sum_{k=1}^{\infty} Pr(Y(k) = A) = \infty$, therefore by the second Borel-Cantelli lemma [2], $Pr(Y(k) = A\ i.o.) = 1$. See the appendix for an explanation of the Borel-Cantelli lemma.

Fix ϵ positive. With probability one there exists $k_1 < k_2 < k_3...$ s.t. $H(Y(k_i)) < 1-\delta$. Therefore there exists k_j s.t. $(1-\delta)^j < \epsilon/d(x(0))$. Choose n large enough so that $k_j < n$. Then by by the contractive property of A, $d(Y_n...Y_1\mathbf{x(0)}) \leq H(Y(n))H(Y(n-1))...H(Y(1))d(\mathbf{x(0)}) \leq H(Y(k_j))H(Y(k_{j-1}))...H(Y(k_1))d(\mathbf{x(0)}) \leq (1-\delta)^j d(\mathbf{x(0)}) < \epsilon$. Therefore $d(\mathbf{x(0)})$ converges to 0 with probability one.

To show that the $\lim_{k\to\infty} X(k)...X(1)\mathbf{x(0)} = ce$, it needs to be shown that c is well defined. Note that since multiplication of a vector by a stochastic matrix is taking convex combinations of its components,

$$c \in [\min_{i\in[d]} \mathbf{x(k)}, \max_{i\in[d]} \mathbf{x(k)_i}] \subseteq [\min_{i\in[d]} \mathbf{x(k-1)}_i, \max_{i\in[d]} \mathbf{x(k-1)}_i].$$

Since the sets are compact and nested, their intersection must not be empty. We also know the diameter of the intervals is going to zero, and therefore must contain just one point i.e. c. $\square$

Thus we have that an irreducible P and more generally a P containing a directed spanning tree supplies us with a RAUM that reaches consensus.

4.2 Whenever P Admits Consensus

In this section, it is seen that any time the matrix P reaches consensus, the RAUM induced by P also reaches consensus. We have already seen in the previous section that if P contains a directed spanning tree, the RAUM induced by P will reach consensus. Here we show that a spanning tree is a necessary condition for the matrix P to reach consensus. There is a little bit of new notation introduced here. We are explicitly considering the directed graph $G =< V, E >$ induced by P. The vertices are given by the index set $[d]$ and edges are given by ordered pairs (i, j). Then $(i, j) \in E$ if $p_{i,j} > 0$. Let $i_1, i_k \in [d]$. We say $i_1 \rightsquigarrow i_k$ if there exists a directed path from i_1 to i_k by which we mean $(i_1, i_2), (i_2, i_3), ..., (i_{k-1}, i_k) \in E$. Clearly the notion of spanning tree can be fitted for this framework as well. Stated simply we say P contains a directed spanning tree if the graph induced by P contains a directed spanning tree and vice versa. If there exists a vertex $r \in V$ such that $i \rightsquigarrow r$ for all $i \in V$, then G contains a directed spanning tree. Clearly having a directed spanning tree in either P or G implies the other. If M is a set of vertices such that $M \rightsquigarrow i$, then $k \in M$ implies $k \rightsquigarrow i$. We will use the notation $i \not\rightsquigarrow j$ to mean there is no directed path from i to j.

Lemma 5 *Suppose P does not contain a directed spanning tree. There exist distinct vertices i and j such that $i \not\rightsquigarrow j$ and $j \not\rightsquigarrow i$. Furthermore, if vertex k is such that $i \rightsquigarrow k$, then $j \not\rightsquigarrow k$ and if vertex s is such that $j \rightsquigarrow s$ then $i \not\rightsquigarrow s$.*

Choose vertex r such that $I(r) := \{v \in V \mid v \rightsquigarrow r\}$ is maximal. Suppose $r \rightsquigarrow k$. It follows that $I(r) \subseteq I(k)$, but since $I(r)$ is maximal, $I(r) = I(k)$. Now since G is induced by a stochastic matrix P, $k \rightsquigarrow j$ where j is not necessarily distinct from k. Then $I(r) \subseteq I(k) \subseteq I(j)$ implying $I(r) = I(k) = I(j)$, since $I(r)$ is maximal. Since $k \in I(j)$, $k \in I(r)$. Thus $k \rightsquigarrow r$. Since G does not contain a spanning tree, there exists i such that $i \not\rightsquigarrow r$. By what was shown above $r \not\rightsquigarrow i$, or else $i \rightsquigarrow r$, a contradiction. By similar reasoning there cannot exist a k such that $i \rightsquigarrow k$ and $r \rightsquigarrow k$, since then $k \rightsquigarrow r$ so $i \rightsquigarrow r$. $\square$

Now we have shown that if P contains no spanning tree, then there must exist indices i and j s.t. $i \not\rightsquigarrow j$ and $j \not\rightsquigarrow i$ and furthermore there does not exist a k such that $i \rightsquigarrow k$ and $j \rightsquigarrow k$. We can always relabel indices in a fashion that suits our needs. This is done heavily in the next proof.

Theorem 3 *If matrix P does not contain a spanning tree, then P does not reach consensus.*

Since P does not contain a spanning tree, find i and j such that $i \not\rightsquigarrow j$ and $j \not\rightsquigarrow i$ and furthermore such that there does not exist a k such that $i \rightsquigarrow k$ and $j \rightsquigarrow k$. Let $M_i = \{v \in V \mid v \neq i,\ v \rightsquigarrow i\}$ and $N_i = (\{v \in V \mid i \rightsquigarrow v\} \cup \{i\}) \cap M_i^c$. Then the matrix must be of the form

$$\begin{array}{c} M_i \\ M_j \\ N_i \\ N_j \end{array} \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ 0 & 0 & * & 0 \\ 0 & 0 & 0 & * \end{bmatrix}$$

where M_i, M_j, N_i and N_j on the left side label how the nodes are labeled. The entries represent block matrices where 0 is the zero matrix and $*$ is some unknown matrix. It should be noted that N_i and N_j cannot be empty. The matrix above is formed by the hypothesis on i and j and definitions of M_i and N_i. $N_i \not\rightsquigarrow M_i$ by definition. $N_i \not\rightsquigarrow M_j$ because $i \not\rightsquigarrow j$ and finally $N_i \not\rightsquigarrow N_j$ since there does exist a directed path from i and j to a common k. Symmetric arguments work on N_j. It can be seen the vector $\mathbf{e}_d$ will not reach consensus. $\square$

In the previous subsection, it is shown that if P has a spanning tree, the induced RAUM model converges. Since it is necessary for P to have a spanning tree in order to reach consensus, we have the following corollary.

Corollary 1 *If P reaches consensus, so does the RAUM induced by P.*

It can also be noted that for any RAUM induced by P, if the expected matrix of the RAUM admits consensus, it must have a spanning tree. Furthermore it can easily be checked if the expected matrix has a tree, then P must as well giving us a final corollary.

Corollary 2 *The expected matrix of a RAUM model reaches consensus if and only if the RAUM model reaches consensus.*

The proof of the backwards direction of corollary (2) can be found in the appendix.

5 Further Study and Conclusion

5.1 Further Study

In this section we look at how the previous results may be applied to a similar problem in the non-linear regime. We consider the dynamics of a function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ and apply it iteratively to some initial vector. Ultimately we will want to consider more general idea of synchronization, but for now we stick with consensus. Therefore we suppose $...\mathbf{f}(\mathbf{f}(\mathbf{x}(\mathbf{0}))))$ reaches consensus. Assuming $\mathbf{f} = [f_1, f_2, .., f_n]^T$ define $\mathbf{f_i} = [I_1, .., I_{i-1}, f_i, I_{i+1}, .., I_d]^T$ where $I_j : \mathbb{R}^n \to \mathbb{R}$ simply projects onto the j^{th} coordinate of a vector. We define an update model in an analogous way to the RAUM with functions $\{\mathbf{f_i}\}_{i=1}^d$ instead of stochastic matrices. In the case where only stochastic matrices are dealt with, we say that consensus implied consensus. It would be interesting to find examples where this does not hold. Stability results are not necessarily global in the non-linear setting however; therefore we might ask if the domain of attraction usually shrinks or grows when we consider the RAUM induced by a non-linear function.

5.2 Conclusion

We were able to accomplish two major goals in this note. First, we were able to conclude that if, while modeling a consensus protocol using a matrix P, for some reason, it can no longer be assumed that agents update simultaneously or in any specific order, but rather update at any time step with some unchanging probability and independent of each other, then we can safely conclude that the property of consensus will not be lost. The second major goal was to expand the class of i.i.d. linear consensus protocol models for which consensus in expectation implies consensus almost surely. We have seen that positive diagonals and weak feedback define classes of models for which consensus in expectation implies consensus almost surely. The

RAUM model shows that these are not the only classes. It would be nice to know if there were necessary and sufficient conditions for when consensus can be implied from the expected matrix. Unfortunately, it could not be seen whether the method of a directed spanning tree used in these notes is able to generalize to a larger class of models that contained one or both of the others. Other open questions include rates of convergence and probability distribution of $c(\mathbf{z})$ given initially vector $\mathbf{z}$.Understanding the probability distribution might help understand how models could be altered in order that $c(\mathbf{z})$ ends within some desired range. Other future work might include how these results could be used to help prove consensus in a non-linear setting as seen in the previous subsection.

6 Appendix

6.1 Consensus implies Expected Consensus

Theorem 4 *Suppose that equation (2) reaches consensus, then the matrix $E(X(t))$ reaches consensus.*

Since the model reaches consensus, there exists a c depending on $\mathbf{x(0)}$ and ω such that $\lim_{t\to\infty} ||\mathbf{x(t)}-c\mathbf{1}|| = 0$ almost surely. We write this as $\lim_{t\to\infty} \mathbf{x(t)} \to c\mathbf{1}$ a.s. Note that $\mathbf{x(t)_j} \leq \max_i |\mathbf{x(0)_i}|$. Therefore applying the dominated convergence theorem yields

$$E(c)\mathbf{1} = \lim_{t\to\infty} E(\mathbf{x(t)}) = \lim_{t\to\infty} E(X(t)....X(1))\mathbf{x(0)} = \lim_{t\to\infty} E(X(t))^t\mathbf{x(0)}.$$

Note that c is not constant but depends upon ω. $\square$

6.2 Proof of Total Variation Lemma and Contraction Property

Lemma 6 *Let $\mathbf{v}$ and $\mathbf{w}$ be in $\mathcal{M}$, then $||\mathbf{v}-\mathbf{w}|| = (1/2)\sum_{i=1}^{d} |\mathbf{v}_i-\mathbf{w}_i| = \sum_{i\in(+)} (\mathbf{v}_i-\mathbf{w}_i) = -\sum_{i\in(-)} (\mathbf{v}_i-\mathbf{w}_i) \leq 1$, where $(+) = \{i \mid \mathbf{v}_i - \mathbf{w}_i \geq 0\}$ and $(-) = \{i \mid \mathbf{v}_i - \mathbf{w}_i < 0\}$.*

Since the sum of the the components of $\mathbf{v} - \mathbf{w}$ is zero

$$0 = \sum_{i=1}^{m} \mathbf{v}_i - \mathbf{w}_i = \sum_{i\in(+)} (\mathbf{v}_i - \mathbf{w}_i) + \sum_{i\in(-)} (\mathbf{v}_i - \mathbf{w}_i).$$

Thus $\sum_{i\in(+)} (\mathbf{v}_i - \mathbf{w}_i) = -\sum_{i\in(-)} (\mathbf{v}_i - \mathbf{w}_i)$ so

$$(1/2)\sum_{i=1}^{d} |\mathbf{v}_i - \mathbf{w}_i| = (1/2)\sum_{i\in(+)} (\mathbf{v}_i - \mathbf{w}_i) - (1/2)\sum_{i\in(-)} (\mathbf{v}_i - \mathbf{w}_i) = \sum_{i\in(+)} (\mathbf{v}_i - \mathbf{w}_i) = -\sum_{i\in(-)} (\mathbf{v}_i - \mathbf{w}_i)$$

and

$$(1/2)\sum_{i=1}^{d} |\mathbf{v}_i - \mathbf{w}_i| \leq \sum_{i=1}^{d} |\mathbf{v}_i| + |\mathbf{w}_i| \leq 1. \square$$

Lemma 7 *Suppose $H(A) \leq 1 - \epsilon$ where $0 < \epsilon \leq 1$. Then $d(A\mathbf{x}) \leq (1 - \epsilon)d(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}^d$.*

Let $i, j \in [d]$ be fixed. Define $(+) = \{s \mid a_{i,s} - a_{j,s} \geq 0\}$ and $(-) = \{s \mid a_{i,s} - a_{j,s} < 0\}$. Let $\mathbf{x} \in \mathbb{R}^d$. Since $\mathbf{x}_s \leq d(\mathbf{x}) + \min(\mathbf{x})$, we have $(a_{i,s} - a_{j,s})\mathbf{x}_s \leq (a_{i,s} - a_{j,s})(d(\mathbf{x}) + \min(\mathbf{x}))$ for all $s \in (+)$. Also $\mathbf{x}_s \geq \min(\mathbf{x})$ so we have $(a_{i,s} - a_{j,s})\mathbf{x}_s \leq (a_{i,s} - a_{j,s}) \min(\mathbf{x})$ for all $s \in (-)$. Therefore

$$[A\mathbf{x}]_i - [A\mathbf{x}]_j = \sum_{s=1}^{d}(a_{i,s} - a_{j,s})\mathbf{x}_s = \sum_{s\in(+)} (a_{i,s} - a_{j,s})\mathbf{x}_s + \sum_{s\in(-)} (a_{i,s} - a_{j,s})\mathbf{x}_s$$

$$\leq \sum_{s\in(+)} (a_{i,s} - a_{j,s})(d(\mathbf{x}) + \min(\mathbf{x})) + \sum_{s\in(-)} (a_{i,s} - a_{j,s}) \min(\mathbf{x})$$

$$= d(\mathbf{x}) \sum_{s\in(+)} (a_{i,s} - a_{j,s}) + \min(\mathbf{x})(\sum_{s\in(+)} (a_{i,s} - a_{j,s}) + \sum_{s\in(-)} (a_{i,s} - a_{j,s}))$$

$$= d(\mathbf{x}) \sum_{s\in(+)} (a_{i,s} - a_{j,s}) \leq d(\mathbf{x})H(A) \leq d(\mathbf{x})(1-\epsilon).$$

Since $\mathbf{x}_s \geq \min(\mathbf{x})$, we have $(a_{i,s} - a_{j,s})\mathbf{x}_s \geq (a_{i,s} - a_{j,s}) \min(\mathbf{x})$. for $s \in (+)$. Likewise, since $\mathbf{x}_s \leq d(\mathbf{x}) + \min(\mathbf{x})$, we get $(a_{i,s} - a_{j,s})\mathbf{x}_s \geq (a_{i,s} - a_{j,s})(d(\mathbf{x}) + \min(\mathbf{x}))$ for $s \in (-)$. Hence

$$[A\mathbf{x}]_i - [A\mathbf{x}]_j = \sum_{s=1}^{d}(a_{i,s} - a_{j,s})\mathbf{x}_s = \sum_{s\in(+)} (a_{i,s} - a_{j,s})\mathbf{x}_s + \sum_{s\in(-)} (a_{i,s} - a_{j,s})\mathbf{x}_s$$

$$\geq \sum_{s\in(+)} (a_{i,s} - a_{j,s}) \min(\mathbf{x}) + \sum_{s\in(-)} (a_{i,s} - a_{j,s})(d(\mathbf{x}) + \min(\mathbf{x}))$$

$$= \min(\mathbf{x})(\sum_{s\in(+)} (a_{i,s} - a_{j,s}) + \sum_{s\in(-)} (a_{i,s} - a_{j,s})) + d(\mathbf{x}) \sum_{s\in(-)} (a_{i,s} - a_{j,s})$$

$$= d(\mathbf{x}) \sum_{s\in(-)} (a_{i,s} - a_{j,s}) \geq -d(\mathbf{x})H(A) \geq d(\mathbf{x})(\epsilon - 1).$$

Thus $|[A\mathbf{x}]_i - [A\mathbf{x}]_j| \leq d(\mathbf{x})(1-\epsilon)$ for all i and j. Therefore $d(A\mathbf{x}) \leq d(\mathbf{x})(1-\epsilon)$. □

6.3 Borel-Cantelli Lemma

Lemma 8 *Let $A_1, A_2, A_3, ...$ be sequence of events that are independent. If $\sum_{n=1}^{\infty} Pr(A_n) = \infty$, then $Pr(A_n \text{ i.o.}) = 1$, where A_n i.o. means the set of ω that are in A_n infinitely often.*

First recall that if the events$\{A_n\}$ are independent, then so are the complements $\{A_n^c\}$. Also $1 - x \leq e^{-x}$ for all x.

$$Pr(\bigcap_{n=M}^{N} An^c) = \prod_{n=M}^{N} (1 - Pr(A_n)) \leq \prod_{n=M}^{N} e^{-Pr(A_n)} = e^{-\sum_{n=M}^{N} Pr(A_n)} \to 0$$

Therefore for all M, $Pr(\bigcup_{n=M}^{N} A_n^c) = 1$. Taking M to infinity and using continuity of measure from above we get that $Pr(\limsup A_n) = Pr(A_n \text{ i.o.}) = 1$.□

In relation to the work in section (4), suppose $Pr(\{\omega \mid H(Y(t)) < 1 - \delta\}) > \epsilon$ for all t. In this case our events are $A_t = \{\omega \mid H(Y(t)) < 1 - \delta\}$. Since we saw that $\sum_{n=1}^{\infty} Pr(A_n) = \infty$, it follows that $Pr(A_n \text{ i.o.} = 1)$. That is to say that the set $\{\omega \mid \exists\, k_1, k_2, .. \text{ with } \omega \in A_{k_1}, A_{k_2}, ..\}$ has measure one. Thus as an event it occurs almost surely allowing us to conclude consensus occurs almost surely.

7 References

References

[1] Morris H. DeGroot. "Reaching a Consensus". In: *Journal of the American Statistical Association* 69.345 (1974), pp. 118–121. DOI: 10.2307/2285509.

[2] Rick Durret. "Probability Theory and Examples". In: Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2010, p. 67. ISBN: 9780521765398.

[3] Y. Hatano and M. Mesbahi. "Agreement over random networks". In: *IEEE Transactions on Automatic Control* 50.11 (2005), pp. 1867–1872. DOI: 0.1109/TAC.2005.858670.

[4] Eugene Seneta. "Markov and the Creation of Markov Chains". In: *MAM2006: Markov Anniversary Meeting*. Ed. by Amy Langville and J. Stewart William. Releigh, North Carolina, 2006, pp. 1–20.

[5] Alireza Tahbaz-Salehi and Ali Jadbabaie. "A Necessary and Sufficient Condition for Consensus Over Random Networks". In: *IEEE Transactions on Automatic Control* 53.3 (2008), pp. 791–795. DOI: 10.1109/TAC.2008.917743.

[6] Behrouz Touri and Angelia Nedic. "On Ergodicity, Infinite Flow, and Consensus in Random Models." In: *IEEE Transactions on Automatic Control* 56.7 (2010), pp. 1593–1605.

[7] R.W. Beard Wei Ren and E.M. Atkins. "A Survey of Consensus in Multi -agent Coordinates". In: *Proceedings of the 2005, American Control Conference* (). DOI: 10.1109/ACC.2005.1470239.

[8] J. Sinai Yakov. "Probability Theory: An Introductory Course". In: trans. by D. Haughton. Springer-Verlag New York Berlin Heidelberg, 1992, pp. 54–66. ISBN: 0387533486.