Distributed Optimization and Machine Learning

Prof. Mayank Baranwal

Computer Science & Engineering, Electrical Engineering, Mathematics

Indian Institute of Technology Bombay

Week-9

Lecture 31: Consensus Algorithms-Fixed time

Having looked at the standard consensus algorithm, so we are going to borrow the ideas of fixed time stability too. to design consensus algorithms which are probably fixed time convergent. So, regardless of what every agent's initial belief is, you are guaranteed to converge enough to the average consensus value in a fixed amount of time. So, before describing the scheme, so let us look at few results. that we have already looked at in the previous lectures, but we are going to be needing them in this proof. So, the first thing is, we derived this in the last lecture.

So, this was something that we had derived in the last lecture. Another result that we had derived was summation i 1 through n, summation through i e i transpose. ok and it is needless to say that we are assuming that the graph is connected otherwise there is no point talking about consensus in such graphs right. So, for now it is undirected unweighted graphs.

* FxTS Consensus scheme:

Results: (i) $\sum_{i=1}^{N}\sum_{j\in N_i} \text{sign}(x_i - x_j) = 0$

(ii) $\sum_{i=1}^{N}\sum_{j\in N_i} e_i^T\, \omega(x_{ij}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\in N_i} e_{ij}^T\, \omega(x_{ij})$

where, $\omega(\cdot)$ is an odd function

So, the third result that we are going to be using is let us say you have z i's scalars which are greater than 0. So, summation of i 1 through ok and we also have the converse version of it. So, usually by triangular inequality we would have that if p is greater than 1 then the sort of inequality flips, but if we still want to retain this inequality it actually turns out that equality is still holds true, but then you have to sort of multiply it with this particular factor. So, that is result number 3 and result number 4. So, if 1 transpose x is equal to 0 that means x is orthogonal to the consensus vector right.

(iii) $z_i > 0$

$$\sum_{i=1}^{N} z_i^{p} \geq \left(\sum_{i=1}^{N} z_i\right)^{p} \quad \text{if } p \in (0,1)$$

$$\sum_{i=1}^{N} z_i^{p} \geq N^{1-p}\left(\sum_{i=1}^{N} z_i\right)^{p} \quad \text{if } p > 1$$

(iv) if $\mathbb{1}^T x = 0$, $x^T L x \geq \Lambda_2(L)\,\|x\|^2$

When I say 1 transpose x is equal to 0 that means x is orthogonal to the consensus vector So that means that vector x does not have any components in the direction of the consensus vector right. So this is true then x transpose L of x this is greater than or equal to I mean x transpose L of x is always greater than equal to 0 it is a positive semi definite matrix. But it is not since this vector x has no component in the direction of the consensus vector. and the only eigenvalue 0 and 0 is the eigenvalue in the basically direction of the consensus vector right. So, this is going to be greater than the Fiedler eigenvalue times square that is the second smallest eigenvalue right.

So, smallest eigenvalue 0 second smallest eigenvalue is a Fiedler value and this is going to be true if 1 transpose x is 0 or x is orthogonal to the consensus vector ok. So, these are the 4 results that we are going to be using. So, one definition or one notation rather that we would be using is. So, x can be vector here and it is defined as x times ok. So, it is almost like signum function.

* **Defn:** $\text{sgn}^{\mu}(x) := x\,\|x\|^{\mu-1}$ $\quad \mu \geq 0$

when $\mu = 0$, $\text{sgn}^{0}(x) = \dfrac{x}{\|x\|} = \text{sign}(x)$

So, when mu is equal sine 0 of x is essentially x over norm of x, which is like a signum function. So, the sine of that particular. So, it is almost like when x is scalar, this almost acts like a signum function. So, we are going to be using this notation to denote this particular quantity. So, well you can have mu greater than equal to 0 that is fine yeah I mean the reason that we are going to be using this notation is because we know that in the context of fixed time stability we would have certain exponents coming in right and this is why we are like in order to make it more compact we are going to be using this notation ok.

So, the fixed time consensus scheme. is x dot rather x i dot. So, if I look at, if I look at this particular thing, right, this was simply x i dot is summation j over like basically the x i minus x j, right, j over the neighborhood set, this j over the neighborhood set. Now I am going to be using this new notation signum ok. So, again if If you remove the signum nu 1 and signum nu 2 part or the sign nu 1 sign nu 2 part, it is exactly a standard consensus algorithm.

FxTS consensus scheme:

$$\dot{x}_i = -\sum_{j \in N_i} \text{sgn}^{\mu_1}(x_i - x_j) - \sum_{j \in N_i} \text{sgn}^{\mu_2}(x_i - x_j)$$

$$\text{with} \quad \left.\begin{array}{l} \mu_1 \in (0,1) \\ \mu_2 > 1 \end{array}\right] - \text{odd}$$

$1/3 \qquad 5/3$

But because we want to use a fixed time stability result, I somehow need to incorporate these these exponents mu 1 and mu 2 with mu 1 being a number between 0 and 1 and mu 2 greater than 1 and you want both these numbers to be odd. So when I say odd, so when mu2 is greater than 1 and a choice could be let's say 5 third or this could be, mu1 could be 1 third. So this is what we want mu1 and mu2 to look like where the both numerators and denominators are odd. And the reason being, as I said this is true for any odd function and not just sine function. So we want when I raise it to the power which is something odd, then this would also be true.

ok and that is why we are actually going to be choosing v1 and v2 which have this odd behavior like odd like characteristic ok. So, this is my scheme. Let us see how this works and why does this converge in a fixed time. So, first thing is we are going to be summing this xi dot and what is this quantity converge to? So, what is this number? x i dot if I just sum it over i 1 through n sum of this right and using this particular property we know that this sum is going to be 0. Is everyone with me on this? So, if I just sum this from 1 through n this quantity is going to be 0 and this quantity is going to sum to 0.

So, this is going to be 0 and therefore, we know that summation i 1 through n x i t is constant ok. The summation is going to be constant always. No, mu 1 for odd numbers as well it is going to be. because it is an odd function, right. This becomes an odd function.

Proof:

$$\sum_{i=1}^{N} \dot{x}_i = 0 \implies \sum_{i=1}^{N} x_i(t) = \text{constant} = \sum_{i=1}^{N} x_i(0)$$

$$x_c = \frac{1}{N} \sum_{i=1}^{N} x_i(0)$$

$\tilde{x}_i = x_i - x_c \longrightarrow$ Want this to converge to 0 for all i

So, for any odd function this would be true. So, in order to show consensus, what kind of Lyapunov functions can we choose? So, this time we do not want it to converge to 0 or anything like in the previous case we want f minus f star. We want like we had worked with f minus f star or gradient of f of x norm square that those were the choices for the Lyapunov function, right. this time we do not want so what quantity do we want it like basically what quantity do we want it to converge to 0 not x star so if i define let's say let's say i define xc to be 1 over n this is my this is where i would i expect the consensus to happen right and if i define x tilde i to be x i minus xc this x tilde is what I wanted to converge to 0 for all i. So, what would be a good choice of Lyapunov candidate So we can choose v to be something like this.

Is this clear? Because v is going to be 0 when the agents arrive at consensus and otherwise it is going to be greater than 0. So what is x i tilde dot? This is going to be x i dot minus x c dot. and xc dot is something that that is going to be that is 0 we know right. So, this is nothing but xi dot. So, that is one thing that we should know.

So, derivative of this is same as xi dot. So, let us take the time derivative of Lyapunov function. Now, let us substitute the value of xi dot. So, let us also write what xi tilde transpose the sine mu 1 of xi minus xj. This is same as xi tilde minus xi xj tilde right.

Choice of Lyapunov candidate:

$$V = \frac{1}{2}\sum_{i=1}^{N} \tilde{x}_i^T \tilde{x}_i$$

$$\dot{\tilde{x}}_i = \dot{x}_i - \cancel{\dot{x}_c}^{\,0} = \dot{x}_i$$

Taking time-derivative of V:

$$\dot{V} = \sum_{i=1}^{N} \tilde{x}_i^T \dot{\tilde{x}}_i$$

$$= \sum_{i=1}^{N} \tilde{x}_i^T \dot{x}_i$$

Because we are just subtracting, adding and subtracting xc. So the difference between xi and xj is same as difference between xi tilde and xj tilde Now, because we have chosen mu 1 and mu 2 to be odd, so this function is odd, these functions are odd and this is where I can use my second result which is summation i 1 through n, j in n i, e i transpose some odd function of xij that is equal to this particular quantity which is now written in terms of eij. So, I can write this as summation. Is everyone with me so far? Any questions on this? So what is the definition of this nu defined signum function? By definition this is xi tilde j x tilde ij times norm of x tilde ij raised to the power mu 1 minus 1 right. So norm mu 1 minus 1 and you have additional x tilde ij that makes it norm nu square norm square.

So nu 1 minus 1 plus 2 which is nu 1 plus 1. So, let me know if you guys did not follow this. again how why because we had this particular definition. So, at x times norm raised to the power mu minus 1 and you have x and there was an additional x transpose setting there. So, that gives you x transpose x which basically becomes norm x square.

So, this basically becomes mu plus 1 ok. Is this clear? so aijs are just 1 so 0 right so summation j in n i i can write it summation j 1 through n and if aijs are 0 then anyway they don't contribute or aijs are 1 so then they contribute right and this is what i'm just i mean in equivalent way to write this okay so now what do we do so if i know that mu 1 is a number between 0 and 1 so is 1 plus mu 1 by 2 that is a number between 0 and 1 and if nu 2 is a number greater than 1 and so is 1 plus minu 2 by 2 right. So, therefore, now I can use this particular result yes ok. So, we because we brought everything in this form. So, let me define let me call this let us say So, what do we have? We have summation half i 1 through n ok.

$$= -\sum_{i=1}^{N} \tilde{x}_i^{T} \left[\sum_{j \in N_i} \underline{\operatorname{sgn}^{\mu_1}(\tilde{x}_i - \tilde{x}_j)} + \sum_{j \in N_i} \underline{\operatorname{sgn}^{\mu_2}(\tilde{x}_i - \tilde{x}_j)} \right]$$

$$= -\frac{1}{2} \sum_{i=1}^{N} \sum_{j \in N_i} \tilde{x}_{ij} \left[\operatorname{sgn}^{\mu_1}(\tilde{x}_{ij}) + \operatorname{sgn}^{\mu_2}(\tilde{x}_{ij}) \right]$$

$$= -\frac{1}{2} \sum_{i=1}^{N} \sum_{j \in N_i} \left[\|\tilde{x}_{ij}\|^{1+\mu_1} + \|\tilde{x}_{ij}\|^{1+\mu_2} \right]$$

$$= -\frac{1}{2} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} \left(a_{ij} \|\tilde{x}_{ij}\|^{2} \right)^{\frac{1+\mu_1}{2}} + \sum_{j=1}^{N} \left(a_{ij} \|\tilde{x}_{ij}\|^{2} \right)^{\frac{1+\mu_2}{2}} \right]$$

And in order to take the summation inside, we need to use these two results. So, this thing is less than or equal to For the other one, I need to introduce addition, but it is not n. How many terms are there? n square because every i, j pair is there. We are summing it.

Yeah, this is n. Yes. n raised to the power 1 minus p. So, p is 1 plus now all we want to now do is to be able to write eta ij in terms of my usual x tilde ij and if I am able to do that then this would be I mean I have basically written the Lyapunov function in terms of in fact what you can do is you can also absorb i inside it so let's do that and let me do one thing get rid of this part. So, we can absorb i as well inside it right and what is preventing us to nothing but half i j 1 through n. now if I include let us say all the terms here. So, now this is summation and the n square many terms right.

So, you can write this as less than equal to negative half you get ok this is everyone is ok with this. Now what is this term summation eta ij it is nothing but summation ij 1 through n by definition this is aij x tilde ij norm square right and we had seen this already that x tilde transpose Lx a summation ij 1 through a ij in the previous lectures x i minus x j square which basically is the same thing. So, that is something that we had already seen. So, this is nothing but x tilde transpose L x and we know that another property is 1 transpose x tilde that is equal to 0.

$$= -\frac{1}{2}\left(\sum_{i,j=1}^{N}\left(a_{ij}\|\tilde{x}_{ij}\|^2\right)^{\frac{1+\mu_1}{2}} + \sum_{i,j=1}^{N}\left(a_{ij}\|\tilde{x}_{ij}\|^2\right)^{\frac{1+\mu_2}{2}}\right]$$

$$\leq -\frac{1}{2}\left[\left(\underbrace{\sum_{i,j=1}^{N}\eta_{ij}}\right)^{\frac{1+\mu_1}{2}} + (N^2)^{\frac{1-\mu_2}{2}}\left(\sum_{i,j=1}^{N}\eta_{ij}\right)^{\frac{1+\mu_2}{2}}\right]$$

$$\sum_{i,j=1}^{N} a_{ij}\|\tilde{x}_{ij}\|^2 \qquad \tilde{x}^T L x = \sum_{i,j=1}^{N} a_{ij}\|\tilde{x}_{ij}\|^2$$

$$\hookrightarrow \tilde{x}^T L x \geq \lambda_2 \|\tilde{x}_{ij}\|^2$$

$$1^T \tilde{x} = 0$$

because it is x i minus x j right. So, 1 transpose x tilde is going to be 0 ok. So, that means x tilde is orthogonal to your consensus vector and therefore, this quantity is greater than equal to lambda 2 ok and with because it is a negative sign. So, what we can show is this thing is less than equal to minus half now you have still norm x tilde square let us say lambda 2 plus n square 1 minus nu 2 by 2 norm x tilde square and norm x tilde square is nothing, but your v 2 times v sorry not square 1 plus nu 1 by 2. square ok. And this is nothing but your 2 times Lyapunov function, 2 times Lyapunov function.

Now you have gotten everything in terms of v dot is less than equal to some c times v 1 raise to the alpha 1 minus c 2 raise to the times v raise to the alpha 2. and therefore, you are guaranteed to converge in a fixed time and therefore, this consensus scheme it runs in a fixed time. So, it converges in a fixed time. Lambda 2 would also have power.

$$\leq -\frac{\Lambda_2}{2}\left[\underbrace{(\|\tilde{x}\|^2)}_{2V}^{\frac{1+\mu_1}{2}} + (N^2)^{1-\frac{\mu_2}{2}}\underbrace{(\|\tilde{x}\|^2)}_{2V}^{\frac{1+\mu_2}{2}}\right]$$

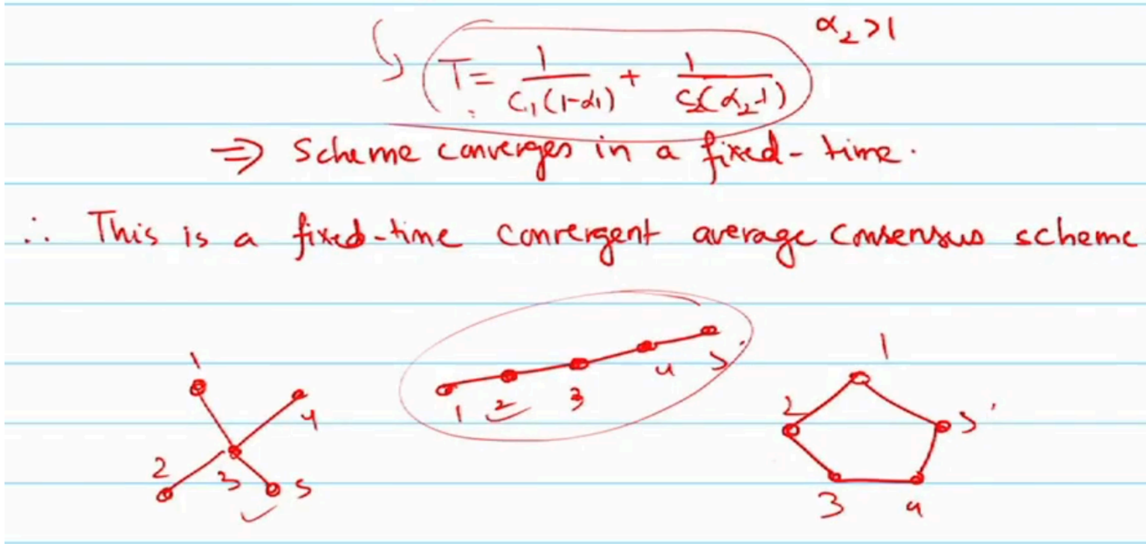
$$\Rightarrow \quad \dot{V} \leq -\underbrace{\frac{\Lambda_2(L)^{\frac{1+\mu_1}{2}}}{2}2^{\frac{1+\mu_1}{2}}}_{c_1} V^{\frac{1+\mu_1}{2}} - \underbrace{\frac{\Lambda_2(L)^{\frac{1+\mu_2}{2}}}{2}2^{\frac{1+\mu_2}{2}}}_{c_2} N^{1-\mu_2}$$

$$\dot{V} \leq -c_1 V^{\alpha_1} - c_2 V^{\alpha_2} \qquad \alpha_1 \in (0,1) \quad \alpha_2 > 1$$

Lambda 2 would also have power, yes. And yeah, lambda 2 absorbed here. So, let us call this c 1, this is c 2. So, you have v dot is less than c 1 v raise to the alpha 1 minus c 2 v raise to the alpha 2 with alpha 1 in a number between 0 and 1 and alpha 2 greater than 1 which implies scheme converges in a fixed time ok. So, this is a fixed time convergent consensus average consensus scheme we missed what we ok. So, fixed time convergent And again this result is pretty new, so I think in 20, you will find the genesis in 2017, 18 papers in 2017, 18 and beyond.

But again the idea is, so how do we design something like this? You know how you can potentially arrive at a particular result, right? What do you need to show? And that basically gives you some idea as to what kind of properties a dynamical system must have. And this is pretty much the recipe of designing new algorithms be in the context of optimization or now in the context of consensus. So, in the next set of lectures we are now going to be using this consensus scheme and an optimization scheme which is going to be gradient based combining the two schemes to guarantee not just the consensus, but this time consensus to the optimal solution of the team objective function or the global objective function. So, which is the distributed optimization problem that we are seeking solution to ok. you can I mean it is non I mean it is I would say it is non trivial and then here.

So, lot of it. Yeah, lot of in fact there are certain results where you do not need the graph to be connected at all times. But the graph can be connected over any consecutive period of L time points. So, it is called L connected. That means any continuous period of L time points, the graph is going to be connected or any continuous duration of L, the graph is going to be connected. So, in that case also you can extend these undirected case or even if it is directed.



strongly connected. For this result to follow, I agree for this result to follow you need the graph to be symmetric and in that case for directed graphs it is going to be somewhat challenging. So, one way to for instance when you work with directed graphs is you I mean it's first it's non-trivial extension of what in fact even the standard consensus algorithm it's a non-trivial extension of that to I mean like basically to come up with something which basically guarantees convergence in the context of directed graph. So, that is there for sure. So, one thing that you need to. So, is the Lyapunov function here v in this particular case when we chose v is v a function of the underlying topology.

Is v a function of underline? No right. Nowhere we in the structure. So, one thing that you can see here is in the convergence. So, what is the settling time? If you had something like this, c1 1 minus alpha 1 plus 1 over c2 alpha 2 minus 1 right. So, if you have the Fiedler value, that Fiedler value comes in the denominator. So, larger Fiedler value means smaller diameter, faster convergence.

So, larger Fiedler value means faster convergence and that is what you see over here. So, the settling time is first of all inversely proportional to the Fiedler value. Fiedler value will show up here. Fiedler value, that larger Fiedler value means faster consensus because it means a smaller diameter and faster consensus. And you see that in this case the settling time is also dependent on the Fiedler value.

So, if you have larger Fiedler value, the settling time becomes smaller and vice versa. So, that is one thing that you should note. The other thing, the underlying topology does not show up in the definition of V. So, it is independent of topology. So all these results hold true even if the graph is time varying.

What do we mean? So at any instant you can have the graph switch from one particular topology to another. Let's say right now it's a star topology and at time t equal to 10 it immediately switches to a ring topology. So as long as the graph is going to be connected, this result is going to hold true no matter what. So the consensus scheme is still going to be executed in a fixed time.

even if the graph is time varying. So, this scheme works for time varying graphs and in that case the settling time you just choose a Fiedler value for the graph which has the largest settling which has the which has the smallest Fiedler value and that would be that settling time of the overall algorithm even if the graph is time varying ok. So, this scheme is going to work even if you keep switching the topology every now and then this scheme is going to work as long as your network is connected. I can point you to my papers. So, we have done this for in the context of distributed optimization. but then the I mean the idea is like if you have a topology let's say five nodes like this or you have something like this.

So, if I mean if at different times it just keeps switching the topology keeps changing, but if you assume that the underlying graph is connected you are guaranteed to converge in a fixed amount of time and in that case you would use a settling time for the graph which which has the smallest Fiedler Eigen value or the largest diameter. No, so if you that is why if you choose the maximum one, the one with like the one or the one with the smallest Fiedler value, then you are fine. What do you mean? Yeah, I mean it's the same graph, just the network is changing. Your neighbors are changing essentially.

Yeah, then it's not. It's the same graph, just the connections are changing. Not a subgraph, just the connections are changing. Like right now you, so for instance, if I look at these, so initially one was a neighbor of 2, one was a neighbor of 3, now one is a neighbor of 2. So, you can be changing connections with your neighbors may be changing evolving, but then you may still as long as the underlying graph is connected or the network is connected this result would hold.

So, design for the maximum. Yeah, design for like. So, the guarantee that you are going to be providing on the settling time is going to be in terms of the graph with the smallest fielded value. So, in that case, so if you want to define like if let us say when x is a vector, so this x transpose instead of working with l, we work with the Kronecker product, but it I mean none of those results change, this is what we work. I did not want to complicate the proof too much. So, I did not really go into it, but if you have a vector valued x, then you essentially use the L chronicle product identity. Thank you.