Formation Control Problem

A **formation** is a collection of n points $p_1, \dots, p_n$ in $\mathbb{R}^d$, defined up to translation. Alternatively, a formation may be defined by specifying its offset vectors $r_{ij} = p_i - p_j$.

- We consider n nodes which begin at positions $x_1(0), \dots, x_n(0)$. Nodes can measure offsets from their neighbors, and they also know what those offsets should be.
- At time $t = 1, 2, \dots$, pairs of nodes that are neighbors in the undirected graph $G(t)$ can measure their offsets $x_i(t) - x_j(t)$.
- A natural approach: node i performs a step of gradient descent on the function

$$\sum_{j \in \mathcal{N}_i(t)} \left\| x_i(t) - x_j(t) - r_{ij} \right\|^2.$$

- This leads to the update law:

$$x_i(t+1) = x_i(t) + 2\epsilon \sum_{j \in \mathcal{N}_i(t)} (x_j(t) - x_i(t)) + 2\epsilon \sum_{j \in \mathcal{N}_i(t)} r_{ij}. \tag{16}$$

Theorem 65 (Undirected Consensus Theorem). *Suppose the graph sequence $G(t)$ is repeatedly connected with a self-loop at every node,[4] and $A(t)$ is a sequence of nonnegative, stochastic matrices that conforms with $G(t)$: if $(i,j) \notin G(t)$, then $a_{ji} = 0$. Moreover, assume that there exists some $\epsilon > 0$ such that if $(i,j) \in G(t)$ then $a_{ji} > \epsilon$. Then the iteration*

$$y(t+1) = A(t)y(t) \quad t = 0, 1, 2, \dots,$$

converges to a multiple of the all-ones vector from any initial vector $y(0)$.

Proof: We will prove the consensus theorem later in the course. □

Using the above above consensus theorem, we can show the following result regarding the formation control problem.

Theorem 66. *As long as $\epsilon < 1/(2n)$ and the graph sequence $G(t)$ is repeatedly connected, the variables $x_1(t), \dots, x_n(t)$ converge to a translate of the correct formation $p_1, \dots, p_n$.*

Proof: For simplicity, let us assume $x_i(t) \in \mathbb{R}$ without loss of generality, and we stack them into a vector $x(t)$. Then we may write Eq. (16) as

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

for appropriately chosen $A(t)$ and $b(t)$ (which depend on $G(t)$). The matrices $A(t)$ are nonnegative, stochastic, and symmetric (so they are doubly stochastic).

[4] Here, a graph sequence $G(t)$ is called *repeatedly connected* if $\bigcup_{t \geq s} G(t)$ is connected for every s.

Observe that for these $A(t)$ and $b(t)$, any translate z of the correct formation satisfies:

$$z = A(t)z + b(t),$$

and so by subtracting from the previous dynamics, we obtain

$$x(t+1) - z = A(t)(x(t) - z).$$

We now apply the consensus theorem (Theorem 65) to conclude that for some constant α^*,

$$\lim_{t\to\infty} x_i(t) - z = \alpha^* \quad \text{for all } i = 1, \ldots, n.$$

□

Coverage Problem on a Line

Consider n nodes $1, \ldots, n$ that start at initial positions $0 \le x_1(0) \le x_2(0) \le \cdots \le x_n(0) \le 1$. Their goal is to move to the positions $x_1^*, \ldots, x_n^*$, which minimize

$$\Phi(x_1, \ldots, x_n) = \max_{y\in[0,1]} \min_{i=1,\ldots,n} d(y, x_i),$$

where

$$d(a,b) = \int_{\min(a,b)}^{\max(a,b)} \phi(z)\, dz.$$

This must be done in a distributed way, with each node interacting only with its immediate neighbors to the right or to the left.

Simple Euclidean Case

Let $\phi(z) = 1$ for all $z \in [0,1]$ so that the distance metric reduces to the standard Euclidean distance, i.e., $d(a,b) = |a-b|$. Define:

$$d_0(t) = 2x_1(t), \quad d_i(t) = x_{i+1}(t) - x_i(t) \text{ for } i = 1, \ldots, n-1, \quad d_n(t) = 2(1 - x_n(t)).$$

Then, a geometric argument shows that the positions $x_1(t), \ldots, x_n(t)$ achieve optimal coverage if and only if:

$$d_0(t) = d_1(t) = d_2(t) = \cdots = d_{n-1}(t) = d_n(t).$$

This suggests developing a distributed update rule that relies on the consensus theorem.

Update Rule Let us consider an update rule where node i attempts to equalize $d_{i-1}(t)$ and $d_i(t)$. For $i = 2, \ldots, n-1$ (non-border cases), define:

$$y_i(t) = \frac{x_{i-1}(t) + x_{i+1}(t)}{2}.$$

The point $y_i(t)$ is the midpoint of the segment $[x_{i-1}(t), x_{i+1}(t)]$ and is the position between $x_{i-1}(t)$ and $x_{i+1}(t)$ that would equalize d_{i-1}, d_i. Every node will move halfway towards $y_i(t)$:

$$x_i(t+1) = \frac{x_i(t) + y_i(t)}{2}. \tag{17}$$

A slightly different definition of $y_i(t)$ is needed at the border cases:

$$y_1(t) = \frac{x_2(t)}{3}, \quad y_n(t) = \frac{2}{3} + \frac{1}{3}x_{n-1}(t). \tag{18}$$

Theorem 67. *Suppose at each time t, every node either updates via equations (17) or (18) (we will say the node updates in this case) or stays at the same place. If every node updates infinitely often according to the above distributed protocol, the system approaches optimal coverage:*

$$\lim_{t\to\infty} \Phi(x_1(t),\ldots,x_n(t)) = \inf_{x_1,\ldots,x_n\in[0,1]} \Phi(x_1,\ldots,x_n).$$

Remark 15. *The above theorem tells us that our protocol for coverage is correct. However, we also need to show that*

$$x_1(t) \le x_2(t) \le \cdots \le x_n(t), \tag{19}$$

so that our protocol only requires nodes to interact with their left and right neighbors.

Proof: It is easy to see that

$$y_i(t) - x_i(t) = \frac{d_i(t) - d_{i-1}(t)}{2},$$

so that

$$\begin{aligned} x_i(t+1) &= x_i(t) + \frac{y_i(t) - x_i(t)}{2} \\ &= x_i(t) + \frac{d_i(t) - d_{i-1}(t)}{4}. \end{aligned} \tag{20}$$

As a consequence,

$$|x_i(t+1) - x_i(t)| < \frac{\max(d_i(t), d_{i-1}(t))}{4}.$$

It follows that no node moves more than one-fourth of the distance to its left or right neighbor; this proves that the nodes never cross, so that Equation (19) holds indeed. Moreover, Equation (20) allows us to write down the dynamics in the space of the d_i variables. Let us work out how $d_i(t+1)$ depends on $d_i(t)$ and possibly one or more of the $d_{i-1}(t)$, $d_{i+1}(t)$. Clearly,

$$d_i(t+1) = d_i(t) + (x_{i+1}(t+1) - x_{i+1}(t)) - (x_i(t+1) - x_i(t)).$$

So we have three possibilities:

1. **Both i and $i+1$ update.** In that case,

$$\begin{aligned} d_i(t+1) &= d_i(t) + (x_{i+1}(t+1) - x_{i+1}(t)) - (x_i(t+1) - x_i(t)) \\ &= d_i(t) + \frac{d_{i+1}(t) - d_i(t)}{4} - \frac{d_i(t) - d_{i-1}(t)}{4} \\ &= \frac{1}{2} d_i(t) + \frac{1}{4} d_{i+1}(t) + \frac{1}{4} d_{i-1}(t). \end{aligned}$$

2. **Only $i+1$ updates.** In that case,

$$\begin{aligned} d_i(t+1) &= d_i(t) + (x_{i+1}(t+1) - x_{i+1}(t)) \\ &= d_i(t) + \frac{d_{i+1}(t) - d_i(t)}{4} \\ &= \frac{3}{4} d_i(t) + \frac{1}{4} d_{i+1}(t). \end{aligned}$$

3. **Only i updates.** In that case,

$$\begin{aligned} d_i(t+1) &= d_i(t) - (x_i(t+1) - x_i(t)) \\ &= d_i(t) - \frac{d_i(t) - d_{i-1}(t)}{4} \\ &= \frac{3}{4} d_i(t) + \frac{1}{4} d_{i-1}(t). \end{aligned}$$

Looking at the three cases above, we can conclude:

(a) The map from the vector $d(t)$ (which stacks up all the $d_i(t)$) to $d(t+1)$ can be written as

$$d(t+1) = B(t)d(t),$$

where $B(t)$ is a nonnegative stochastic matrix. Moreover, the positive entries of $B(t)$ are bounded away from zero—they are never less than $\frac{1}{4}$ in the above equations.

(b) The matrices $B(t)$ are undirected (i.e., if $b_{ij}(t) > 0$, then $b_{ji}(t) > 0$). Indeed, suppose $d_i(t+1)$ includes a positive coefficient on $d_{i+1}(t)$. Per the above cases, this occurs in case 1 or 2, meaning node $i+1$ updates. But this means that $d_{i+1}(t+1)$ will also include a positive coefficient on $d_i(t)$—thus confirming that the matrices are undirected.

(c) Whenever node i updates, the equations for both $d_i(t+1)$ and $d_{i-1}(t+1)$ place positive weights on $d_i(t)$ and $d_{i-1}(t)$. Since every node updates infinitely often, and because of existence of the self-loops, we obtain repeated connectivity.

Items (a), (b), and (c) above allow us to apply the consensus theorem to the update rule

$$d(t+1) = B(t)d(t),$$

and conclude that in the limit, every $d_i(t)$ is the same. Thus, as $t \to \infty$, the system achieves optimal coverage.[5] □

General Metric Case

- The case for a general ϕ is handled as follows. We define $m_\alpha(a,b)$ to be the point $c \in [\min(a,b), \max(a,b)]$ satisfying

$$\int_{\min(a,b)}^{c} \phi(z)\, dz = \alpha \int_{c}^{\max(a,b)} \phi(z)\, dz.$$

- Then the nodes set:

$$\begin{aligned} y_1(t) &= m_{1/2}(0, x_2(t)), \\ y_i(t) &= m_1(x_{i-1}(t), x_{i+1}(t)) \quad \text{for } i = 2, \ldots, n-1, \\ y_n(t) &= m_2(x_{n-1}(t), 1), \end{aligned}$$

and update according to:

$$x_i(t+1) = m_1(x_i(t), y_i(t)).$$

[5] To be fully rigorous, we should also analyze what happens in the border cases—we skip this (tedious) step here.

Remark 16. *To prove that this scheme achieves optimal coverage, we define the cumulative function:* $F(a) = \int_0^a \phi(z)\,dz$*, and observe that the transformed quantities:*

$$x_i'(t) = F(x_i(t)), \quad y_i'(t) = F(y_i(t))$$

follow the same dynamics as in the uniform (Euclidean) coverage algorithm. We then argue that the optimality conditions satisfied in the limit by the transformed variables x_i'*, namely,* $d_0' = d_1' = \cdots = d_n'$*, imply that we have achieved optimal coverage under the general metric* ϕ*.*

Ergodicity and Consensus Theorems

A nonnegative stochastic matrix $P \in \mathbb{R}^{n\times n}$ (i.e., each row sums to 1) is naturally associated with a time-homogeneous Markov chain $X(0), X(1), X(2), \ldots$ on the state-space $\{1, 2, \ldots, n\}$ for which

$$\mathbb{P}(X(t) = j \mid X(t-1) = i) = [P]_{ij}, \quad \text{for every } t.$$

Also, recall from the past:

$$\mathbb{P}(X(t+\ell) = j \mid X(t) = i) = [P^\ell]_{ij}.$$

A sequence of nonnegative stochastic matrices $P(1), P(2), P(3), \ldots$, all in $\mathbb{R}^{n\times n}$, can be associated with a time-inhomogeneous Markov chain $X(0), X(1), X(2), \ldots$ on the state-space $\{1, \ldots, n\}$ for which

$$\mathbb{P}(X(t) = j \mid X(t-1) = i) = [P(t)]_{ij}.$$

We then have

$$\mathbb{P}(X(t+\ell) = j \mid X(t) = i) = [P(t+1)P(t+2)\cdots P(t+\ell)]_{ij}.$$

We adopt the notation $p_i(t)$ for the row vector in $\mathbb{R}^n$ whose j-th component is $\mathbb{P}(X(t) = j \mid X(0) = i)$. Informally, $p_i(t)$ summarizes the system conditioned on starting at node i at time 0. As a direct consequence of the definition, $p_i(t)$ is the i-th row of the product matrix: $P(1)P(2)\cdots P(t)$.

Given a matrix $P \in \mathbb{R}^{n\times n}$, we define $G(P)$ to be the (directed) graph with vertex set $\{1, \ldots, n\}$ and edge set $\{(i,j) \mid [P]_{ij} > 0\}$. We say that the sequence $P(1), P(2), P(3), \ldots$ is *B-strongly connected* if for any $k = 0, 1, 2, \ldots$, the graph

$$\bigcup_{t=kB+1}^{(k+1)B} G(P(t))$$

is strongly connected (i.e., there is a directed path from every node to every other node).

Theorem 68 (Ergodicity Theorem). *Suppose* $P(1), P(2), P(3), \ldots$ *is a sequence of stochastic matrices such that:*

1. *The sequence* $G(P(t))$ *is* B*-strongly connected for some positive integer* B*.*
2. $[P(t)]_{ii} > 0$ *for all* i *and* t*.*
3. *There exists some* $\epsilon > 0$ *such that if* $[P(t)]_{ij} > 0$*, then* $[P(t)]_{ij} \geq \epsilon$*.*

Then, $\lim_{t\to\infty} \|p_i(t) - p_j(t)\| = 0 \;\forall i, j$*. Moreover, for any* $k = 1, \ldots, n$*, we have*

$$|p_{ik}(t) - p_{jk}(t)| \leq \left(1 - \epsilon^{nB}\right)^{\lfloor \frac{t}{nB} \rfloor}.$$

Intuition: This theorem states that, subject to a few natural conditions, a Markov chain forgets where it started: the distributions $p_i(t)$ and $p_j(t)$ become increasingly alike for all i, j. Note that we do not claim that $p_i(t)$ converges. Rather, we claim only that the differences $p_i(t) - p_j(t)$ converge to zero.

Consensus theorems can be derived from ergodicity theorems. Specifically, let us derive a consensus theorem from the above statement; it will be for directed graphs, so it will be slightly different from the form we have used previously. Specifically, let us now prove the following.

Theorem 69 (Directed Consensus Theorem). *Suppose*

$$x(t+1) = A(t)x(t) \quad t = 0, 1, 2, \ldots,$$

where each $A(t)$ is a stochastic matrix, and the sequence $\{A(t)\}$ satisfies the following:

1. *The sequence of graphs $G(A(t))$ is B-strongly-connected for some B.*
2. *$[A(t)]_{ii} > 0$ for all i and t.*
3. *There exists some $\epsilon > 0$ such that whenever $[A(t)]_{ij} > 0$, it holds that $[A(t)]_{ij} \geq \epsilon$.*

Then every $x_i(t)$ converges to the same value as $t \to \infty$; that is, the system reaches consensus.

Proof: Consider the product

$$A(t)A(t-1)\cdots A(1),$$

and let us apply the ergodicity theorem to it. Note that the indices of the matrices are in *reverse* order compared to the formulation of the ergodicity theorem, where the first matrix appears in the far-left position, while here $A(1)$ is in the far-right position.

It is easy to see that the sequence $G(A(t)), G(A(t-1)), \ldots, G(A(1))$ is the initial segment of a B-connected sequence, possibly after discarding at most B of the initial graphs. The second condition (positive diagonal entries) and the third condition (entries bounded below by ϵ) of the ergodicity theorem are satisfied due to the assumptions of the Directed Consensus Theorem. Therefore, we conclude the following: if $a_i(t)$ and $a_j(t)$ are the i-th and j-th rows of the product matrix $A(t)A(t-1)\cdots A(1)$, then

$$\|a_i(t) - a_j(t)\|_\infty \leq \left(1 - \epsilon^{nB}\right)^{\left\lfloor \frac{t-B}{nB} \right\rfloor}. \tag{21}$$

Now, for any $s > t$, the row $a_\ell(s)$ is a convex combination of the rows $a_1(t), a_2(t), \ldots, a_n(t)$. Hence, for all $s_1, s_2 \geq t$,

$$\|a_i(s_1) - a_j(s_2)\|_\infty \leq \left(1 - \epsilon^{nB}\right)^{\left\lfloor \frac{t-B}{nB} \right\rfloor}.$$

Applying this with $i = j$, we see that each $a_i(t)$ is a Cauchy sequence and thus converges. Since the difference between any two rows $a_i(t)$ and $a_j(t)$ tends to zero by Eq. (21), they all converge to the same row vector; call it $a^\top$. It follows that the i-th entry of $x(t)$ converges to $a^\top x(0)$ for all i. Therefore, the system reaches consensus. □

Remark 17. *Note that we are requiring the sequence $G(A(t))$ to be B-strongly-connected, which is a somewhat more stringent requirement than the previous notion of repeated connectivity. However, there is no requirement here that the matrices $A(t)$ be undirected. As an exercise, you can use a similar argument to prove the consensus theorems for repeatedly connected and undirected graph sequences.*

Proof (Ergodicity Theorem 68): Define $S_i(t)$ to be the support of $p_i(t)$, i.e., the set of indices k for which $p_{ik}(t) > 0$:

$$S_i(t) := \{k \in \{1,\dots,n\} \mid p_{ik}(t) > 0\}.$$

Since each $P(t)$ has positive diagonal entries, the supports $S_i(t)$ are non-decreasing in the sense that if $j \in S_i(t)$, then $j \in S_i(t+1)$. Moreover, because the sequence $P(t)$ is B-strongly-connected, we have that if $S_i(kB) \neq \{1,\dots,n\}$, then $S_i((k+1)B)$ is strictly larger than $S_i(kB)$. The previous two points imply that after nB steps, we must have:

$$S_i(nB) = \{1,\dots,n\}, \quad \text{for all } i = 1,\dots,n.$$

Since every positive transition probability is at least ϵ, it follows that:

$$\mathbb{P}(X_i(nB) = k) \geq \epsilon^{nB}, \quad \text{for all } i,k \in \{1,\dots,n\}, \tag{22}$$

where $X_i(t)$ denotes the state of the random walk at time t that began at node i at time 0.

Define $I_{ij}(t)$ to be the event that $X_i(m) = X_j(m)$ for some $m \in \{0,1,\dots,t\}$, i.e., two independent random walks starting at nodes i and j have intersected by time t. Then, we have:

$$\mathbb{P}(I_{ij}(nB)) \geq \epsilon^{nB} \quad \forall i,j,$$

because at time nB, the walker starting at node i is at some location (say, node i'), and the probability that the other walker starting at j is also at i' is at least ϵ^{nB} by (22). As a result, for any integer k, it follows that:

$$\mathbb{P}(I_{ij}(knB)) \geq 1 - (1-\epsilon^{nB})^k. \tag{23}$$

Lemma 70. *Define a random process $X_i^j(1), X_i^j(2), X_i^j(3), \dots$ as follows:*

- *If the walks starting at i and j have not intersected by time t, then set $X_i^j(t) = X_j(t)$.*
- *Otherwise, set $X_i^j(t) = X_i(t)$.*

Then, $X_i^j(t)$ has the same distribution as $X_j(t)$.

Lemma 71. *If X and Y are discrete random variables, then*

$$|\mathbb{P}(X=a) - \mathbb{P}(Y=a)| \leq \mathbb{P}(X \neq Y).$$

To complete the proof of the ergodicity theorem, we use the above lemmas to establish the following sequence of inequalities:

$$\begin{aligned}
|p_{ik}(t) - p_{jk}(t)| &= |\mathbb{P}(X_i(t) = k) - \mathbb{P}(X_j(t) = k)| \\
&= |\mathbb{P}(X_i(t) = k) - \mathbb{P}(X_i^j(t) = k)| \\
&\leq \mathbb{P}(X_i(t) \neq X_i^j(t)) \\
&= \mathbb{P}(I_{ij}(t)^c),
\end{aligned}$$

where $I_{ij}(t)^c$ is the complement of the event $I_{kl}(t)$, meaning that the independent random walks started at nodes i and j have not intersected by time t. Finally, by choosing $k = \lfloor \frac{t}{nB} \rfloor$ in (23), we know that:

$$\mathbb{P}(I_{ij}(t)^c) \leq \left(1 - \epsilon^{nB}\right)^{\lfloor \frac{t}{nB} \rfloor},$$

which completes the proof. □

A Quadratic Convergence Time in the Consensus Theorem

How Good Are Our Bounds for Consensus?

Recall the structure of the last few lectures: we have started from formation control and coverage problems and provided control protocols for them which relied on the consensus theorem. We now turn to the question of convergence times. If you go back to the previous lectures, you see that the convergence time bounds for coverage and formation control are just the convergence time bounds for the associated consensus process. But what sort of bounds have we derived? Let us revisit the previous lecture and see. To bound the convergence speed of the consensus iteration

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

we have considered the products

$$A(t)A(t-1)\cdots A(1).$$

Subject to some conditions given by the directed consensus theorems, we have proven that for all $i, j = 1, \ldots, n$, the i-th and j-th row of the above matrix, denoted by $a_i(t)$ and $a_j(t)$, satisfy

$$\|a_i(t) - a_j(t)\| \leq \left(1 - \epsilon^{nB}\right)^{\frac{t-B}{nB}},$$

for some constant B depending on the graph structure. But what does this mean for the vector $x(t)$ which satisfies $x(t+1) = A(t)x(t)$? Well, using the inequality $|x^T y| \leq \|x\|_1 \|y\|_1$, we have

$$|x_i(t) - x_j(t)| = |a_i(t)x(0) - a_j(t)x(0)| \leq \|a_i(t) - a_j(t)\|_1 \|x(0)\|_1,$$

or

$$|x_i(t) - x_j(t)| \leq (1 - \epsilon^{nB})^{\frac{t-B}{nB}} \sum_{j=1}^{n} |x_j(0)|.$$

This is discouraging. Indeed, consider: how long does it take for this last expression to fall below δ? Using some standard analysis, the answer is: $t = O\left(nB(\frac{1}{\epsilon})^{nB} \log \frac{\|x(0)\|_1}{\delta}\right)$ iterations for this to shrink below δ. Recall, for example, the matrices from on formation control with offsets: the entries were as small as $1/n$. Plugging that into here gives a bound that grows at least as n^{nB}, i.e., exponentially in n. This gets huge very quickly, for example, if $n = B = 10$, this is 10^{100}. Our goal here is to give a more refined version of the consensus theorem, which shows that on undirected and connected graphs, consensus can happen much faster if the nodes are smart about how they pick coefficients.

So suppose all the graphs $G(t)$ are undirected and connected. We define the diagonally-dominant *Metropolis matrices* $M(t)$ as:

$$[M(t)]_{ij} = \begin{cases} \frac{1}{2\max(d_i(t), d_j(t))} & \text{if } i \neq j \text{ and } \{j, i\} \in G(t) \\ 0 & \text{if } i \neq j \text{ and } \{j, i\} \notin G(t) \end{cases},$$

with

$$[M(t)]_{ii} = 1 - \sum_{j \neq i} [M(t)]_{ij},$$

where $d_i(t)$ denotes the degree of node i in the graph $G(t)$. Note that $M(t)$ is always stochastic, nonnegative, symmetric, and its diagonal entries are at least $1/2$.

Theorem 72 (Metropolis Consensus Theorem). *For the iterations*

$$x(t+1) = M(t)x(t) \quad t = 0, 1, 2, \ldots,$$

assuming all the graphs $G(M(t))$ are undirected and connected, the time until $|x_i(t) - x_j(t)| \leq \delta$ scales as

$$t = O\left(n^2 \log\left(\frac{n\|x(0)\|_1}{\delta}\right)\right).$$

Clearly this bound is much better than the exponential convergence bound given before and only scales quadratically (up to a log factor) with respect to the number of agents.

Distributed Optimization of a Sum

We consider here the problem of optimizing a sum of convex functions in a distributed way. We have n nodes, which we will label $1, \ldots, n$, each of which knows a convex function $f_i(\cdot) : \mathbb{R}^m \to \mathbb{R}$. They are interconnected in a sequence of graphs $G(t) = (\{1, \ldots, n\}, E(t))$ which we assume to be undirected and B-connected for some $B > 0$. The agents would like to minimize the sum of these functions:

$$f(x) = \sum_{i=1}^{n} f_i(x)$$

based on only neighbor interactions in the time-varying graphs $G(t)$.

Let us assume that the set of global minima $X^* = \arg\min_{x \in \mathbb{R}^m} f(x)$ is not empty. We will analyze a simple scheme that interpolates between consensus and the subgradient method. This is the scheme:

$$x_i(t+1) = \sum_{j \in N_i(t)} a_{ij}(t) x_j(t) - \alpha(t) d_i(t)$$

where $d_i(t)$ is a subgradient of the function $f_i(x)$ at the point $x = x_i(t)$, $a_{ij}(t)$ are the symmetric Metropolis weights (i.e., $a_{ij}(t) = \frac{1}{2\max(d(i),d(j))}$ whenever $(i, j) \in E(t)$ and zero otherwise), and as usual $\alpha(t)$ is a nonnegative step size that satisfies:

$$\sum_{t=1}^{\infty} \alpha(t) = \infty, \quad \sum_{t=1}^{\infty} \alpha^2(t) < \infty.$$

We will call this the *consensus subgradient scheme*.

We will show that this scheme drives every $x_i(t)$ to a point in X^* under the usual assumptions of subgradient boundedness. We begin with a series of useful lemmas.

Lemma 1. Suppose $f : \mathbb{R}^m \to \mathbb{R}$ is a convex function such that f has subgradients at points x, y whose norm is bounded by L. Then

$$|f(y) - f(x)| \leq L\|y - x\|_2$$

Lemma 73. *Suppose*

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

where $A(t)$ are Metropolis matrices on some undirected B-connected graph sequence. If $p(t) \to 0$,

$$\max_{i,j} \|x_i(t) - x_j(t)\| \leq C\rho^t + \sum_{s=0}^{t-1} 2\rho^{t-1-s}\|p(s)\| \to 0,$$

where C is a constant and $\rho \in (0, 1)$ depends on the Metropolis matrices (e.g., $\rho = \tilde{O}(1 - \frac{1}{n^2})$).

Proof: Consider the update rule:

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

where each $A(t)$ is a Metropolis matrix corresponding to a connected undirected graph, and $p(t) \to 0$. We want to bound the disagreement:

$$\Delta(t) := \max_{i,j} |x_i(t) - x_j(t)|.$$

We first unroll the recursion from time 0 to time t:

$$x(t) = \Phi(t,0)x(0) + \sum_{s=0}^{t-1} \Phi(t,s+1)p(s),$$

where the transition matrix is defined as:

$$\Phi(t,s+1) := A(t-1)A(t-2)\cdots A(s+1).$$

We split the solution into two parts: the homogeneous part and the perturbation part:

$$x(t) = \underbrace{\Phi(t,0)x(0)}_{\text{I}} + \underbrace{\sum_{s=0}^{t-1} \Phi(t,s+1)p(s)}_{\text{II}}.$$

Then the disagreement satisfies:

$$\Delta(t) \leq \max_{i,j} |\Phi_i(t,0)x(0) - \Phi_j(t,0)x(0)| + \max_{i,j} \left| \sum_{s=0}^{t-1} (\Phi_i(t,s+1) - \Phi_j(t,s+1))p(s) \right|.$$

For the homogeneous part, from standard results on consensus over connected graphs with Metropolis weights, there exist constants $C > 0$ and $\rho \in (0,1)$ such that:

$$\max_{i,j} |\Phi_i(t,0)x(0) - \Phi_j(t,0)x(0)| \leq C\rho^t.$$

Next, we bound the perturbation term using:

$$\left| \sum_{s=0}^{t-1} (\Phi_i(t,s+1) - \Phi_j(t,s+1))p(s) \right| \leq \sum_{s=0}^{t-1} \left\| \Phi_i(t,s+1) - \Phi_j(t,s+1) \right\| \cdot \|p(s)\|.$$

Using known contraction properties of stochastic consensus matrices:

$$\left\| \Phi_i(t,s+1) - \Phi_j(t,s+1) \right\| \leq 2\rho^{t-1-s}.$$

So the bound becomes:

$$\sum_{s=0}^{t-1} 2\rho^{t-1-s} \|p(s)\|.$$

Combining both terms, we obtain:

$$\Delta(t) = \max_{i,j} |x_i(t) - x_j(t)| \leq C\rho^t + \sum_{s=0}^{t-1} 2\rho^{t-1-s} \|p(s)\|.$$

□

Theorem 74. *Under all the assumptions made previously (each f_i convex, X^* nonempty, $\alpha(t)$ is not-summable-but-square-summable), and the assumption that the subgradients of each f_i are all uniformly bounded in 2-norm by L, we have that*

$$\min_{l=1,\ldots,t} f(x_i(l)) \to f(x^*)$$

as $t \to \infty$ for every $i = 1, \ldots, n$.

Proof: For simplicity, we assume that $m = 1$, i.e., that the functions f_i are from $\mathbb{R}$ to $\mathbb{R}$. Then the consensus subgradient scheme can be written as:

$$x(t+1) = A(t)x(t) - \alpha(t)d(t)$$

where $x(t)$ is the vector stacking the $x_i(t)$ and $d(t)$ is the vector stacking the $d_i(t)$. Letting $y(t)$ be the average of the entries of $x(t)$, we have:

$$y(t+1) = y(t) - \alpha(t) \cdot \frac{1}{n} \sum_{j=1}^{n} d_j(t)$$

Let x^* be any point in X^*. Then:

$$(y(t+1) - x^*)^2 \leq (y(t) - x^*)^2 + \alpha^2(t)L^2 - \frac{2\alpha(t)}{n} \sum_{i=1}^{n} d_i(t)(y(t) - x^*)$$

Now observe:

$$\begin{aligned} d_i(t)(y(t) - x^*) &= d_i(t)(x_i(t) - x^* + y(t) - x_i(t)) \\ &= d_i(t)(x_i(t) - x^*) + d_i(t)(y(t) - x_i(t)) \\ &\geq f_i(x_i(t)) - f_i(x^*) - L|y(t) - x_i(t)| \\ &\geq f_i(y(t)) - f_i(x^*) - 2L|y(t) - x_i(t)|, \end{aligned}$$

where we used convexity and Lemma 1 in the last steps. It follows that:

$$(y(t+1) - x^*)^2 \leq (y(t) - x^*)^2 + \alpha^2(t)L^2 - 2\frac{\alpha(t)}{n}(f(y(t)) - f(x^*)) + 4L\alpha(t)\max_i |y(t) - x_i(t)|.$$

Summing over time steps, we get

$$\frac{2}{n} \sum_{l=1}^{t} \alpha(l)(f(y(l)) - f(x^*)) \leq (y(0) - x^*)^2 + L^2 \sum_{l=1}^{t} \alpha^2(l) + 4L \sum_{l=1}^{t} \alpha(l) \max_i |y(l) - x_i(l)|.$$

Therefore,

$$\min_{l=1,\ldots,t} f(y(l)) - f(x^*) \leq \frac{(y(0) - x^*)^2 + L^2 \sum_{l=1}^{t} \alpha^2(l) + 2Ln \sum_{l=1}^{t} \alpha(l) \max_i |y(l) - x_i(l)|}{2 \sum_{l=1}^{t} \alpha(l)}$$

Note that the right-hand side of the above expression goes to zero as $t \to \infty$ because: $\sum \alpha(t) = \infty$, $\sum \alpha^2(t) < \infty$, and $\max_i |y(t) - x_i(t)| \to 0$ by Lemma 73. That gives us

$$\min_{l=1,\ldots,t} f(x_i(l)) - f(x^*) \to 0,$$

for every $i = 1, \ldots, n$. □