Distributed Consensus: Convergence, Robustness, and Optimization

Lin Xiao
Microsoft Research

Mathematical Challenges and Opportunities in Sensor Networking
IPAM, UCLA
January 10, 2007
Distributed average consensus

- **basic algorithm**
  - initial values at each node $x_i(0)$
  - every node to compute $\frac{1}{n} \sum x_i(0)$

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

- design **local parameters** $w_{ij}$ to achieve **global performance**
  - convergence, robustness to unreliable links, noises, topology changes
  - optimization for fastest convergence, minimum communication costs
Applications of distributed consensus

- distributed coordination, synchronization, flocking, and load balancing


- information processing in sensor networks
  - distributed parameter estimation, distributed Kalman filtering
  - distributed detection/inference

Mathematical challenges and opportunities

• basic algorithm

\[ x_i(t + 1) = x_i(t) + \sum_{j \in \mathcal{N}_i} w_{ij} (x_j(t) - x_i(t)) \]

• challenges

convergence, robustness, optimization

• mathematical tools
  – linear algebra
  – spectral graph theory
  – convex optimization
  – duality

• connections: Markov chains, sensor localization, dimensionality reduction
Outline

- convergence on fixed graph
- optimization on fixed graph
- robustness to random link failures
- optimization of randomized gossip
Convergence on fixed graph

• basic algorithm

\[ x_i(t + 1) = x_i(t) + \sum_{j \in N_i} w_{ij} \left( x_j(t) - x_i(t) \right) \]

• vector form: \( x(t + 1) = W x(t) \)

\[
W_{ij} = \begin{cases} 
   w_{ij} & \{i, j\} \in \mathcal{E} \\
   0 & \{i, j\} \notin \mathcal{E} \\
   1 - \sum_k w_{ik} & i = j \end{cases} \quad (W = I - L)
\]

• theorem (XB'04): \( \lim_{t \to \infty} x_i(t) = \frac{1}{n} \sum_{j=1}^{n} x_j(0) \) for all \( x(0) \in \mathbb{R}^n \) iff

\[
1^T W = 1^T, \quad W 1 = 1, \quad \rho \left( W - \left(1/n\right)11^T \right) < 1
\]
Optimal design for fastest convergence

• for symmetric weights, define convergence factor

$$\mu(W) = \rho \left( W - \frac{1}{n}11^T \right) = \| W - \frac{1}{n}11^T \|$$

• convergence bound

$$\| x(t) - \bar{x}1 \|_2 \leq \mu(W)^t \| x(0) - \bar{x}1 \|_2$$

• optimization with symmetric weights

minimize \( \mu(W) \)
subject to \( W = W^T, \quad W1 = 1 \)
\( W_{ij} = 0, \quad \{ i, j \} \notin E \)

a convex optimization problem, globally and efficiently solved
A small example

**FMMC**

\[ W^* = \arg \min_{W \in \mathcal{C}} \mu(W) \]

\[ \mu^* = 0.72 \]

**Metropolis**

\[ W_{ij} = \frac{1}{\max\{d_i, d_j\}} \]

\[ \mu = 0.77 \]
A larger example

- randomly generated network with 50 nodes, 200 edges

<table>
<thead>
<tr>
<th></th>
<th>Metropolis</th>
<th>best constant</th>
<th>optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho(W - 11^T/n)$</td>
<td>0.949</td>
<td>0.947</td>
<td>0.902</td>
</tr>
<tr>
<td>$\tau = 1/\log(1/\rho)$</td>
<td>19.104</td>
<td>18.363</td>
<td>9.696</td>
</tr>
</tbody>
</table>
Sparse network design

find most sparse subgraph with guaranteed convergence factor \( \mu_0 \)

- a hard combinatorial optimization problem; use \( \ell_1 \) heuristic:

\[
\begin{align*}
\text{minimize} \quad & \sum_{\{i,j\} \in \mathcal{E}} |w_{ij}| \\
\text{subject to} \quad & \mu(W) \leq \mu_0 \\
& W = W^T, \quad W \mathbf{1} = \mathbf{1} \\
& w_{ij} = 0, \quad \{i, j\} \notin \mathcal{E}
\end{align*}
\]

- best possible rate: \( \mu^* = 0.902 \)
- set design target: \( \mu_0 = 0.910 \)

- total number of edges: 200
  - used by sparse design: 96
Distributed consensus with additive noise

- example: distributed load balancing with new jobs arriving randomly

\[ x_i(t + 1) = x_i(t) + \sum_{j \in \mathcal{N}_i} w_{ij} (x_j(t) - x_i(t)) + v_i(t) \]

\( v_i(t) \) i.i.d., zero mean & unit variance (consider \( v_i(t) = u_i(t) - c \))

- instantaneous average: \( a(t) = \frac{1}{n} \sum_{i=1}^{n} x_i(t) \), does random walk
Mean-squares deviation

• \( x(t + 1) = W x(t) + v(t) \)

• consider mean-square deviation

\[
\text{MSD}(W) = \lim_{t \to \infty} E \sum_{i=1}^{n} (x_i(t) - a(t))^2
\]

(Cybenko, 1989)

• optimize \( W \) to minimize MSD(\( W \))
Least-mean-square (LMS) distributed consensus

- **Theorem (XBK'06):** explicit expression of MSD($W$)

\[
\lim_{t \to \infty} \mathbb{E} \sum_{i=1}^{n} (x_i(t) - a(t))^2 = \sum_{i=2}^{n} \frac{1}{1 - \lambda_i(W)^2}
\]

Eigenvalues: \(1 = \lambda_1(W) > \lambda_2(W) \geq \cdots \geq \lambda_n(W) \geq -1\)

- Optimal design for least-mean-square deviation

\[
\text{minimize } \sum_{i=2}^{n} \frac{1}{1 - \lambda_i(W)^2}
\]

subject to \(W = W^T, \quad W1 = 1\)

\(W_{ij} = 0, \quad \{i, j\} \notin \mathcal{E}\)

Again, a **convex optimization** problem (convex, and differentiable)
Examples of LMS load balancing

- **2-D grids**

  \[ n = 5 \]

- **hypecubes**

  \[ d = 3 \]
Outline

- convergence on fixed graph
- optimization on fixed graph
- robustness to random link failures
- optimization of randomized gossip
Distributed consensus with unreliable links

- communication links may work or fail at random (due to mobility, fading, power constraints); network topology changes with time

- some notations
  - $G(t) = (\mathcal{E}(t), \mathcal{V})$ time-varying communication graph
  - $\mathcal{N}_i(t) = \{j \in \mathcal{V} | \{i, j\} \in \mathcal{E}(t)\}$ instantaneous neighborhood
  - $\{G(t)\}_{t=0}^{\infty}$ can be either deterministic or stochastic

- distributed average consensus (same form, time-varying weights)

\[
    x_i(t + 1) = x_i(t) + \sum_{j \in \mathcal{N}_i(t)} w_{ij}(t) (x_j(t) - x_i(t))
\]

what conditions on $\{G(t), W(t)\}_{t=0}^{\infty}$ guarantee convergence?
Metropolis weights

• distributed average consensus

\[ x_i(t + 1) = x_i(t) + \sum_{j \in N_i(t)} w_{ij}(t) (x_j(t) - x_i(t)) \]

• time-varying Metropolis weights

\[ w_{ij}(t) = \frac{1}{1 + \max\{d_i(t), d_j(t)\}} \quad \{i, j\} \in \mathcal{E}(t) \]

\[ d_i(t) = |N_i(t)|, \text{ number of neighbors at time } t \]

• only use local information, suitable for distributed implementation
Robustness of convergence

- **Theorem (XBL:05):** If the communication graphs that occur infinitely often in \( \{G(t)\}_{t=0}^{\infty} \) are **jointly connected**, then the iteration

\[
x(t + 1) = W(t)x(t)
\]

converges to the average for any \( x(0) \in \mathbb{R}^n \)

- A finite set of graphs with common vertex set \( \mathcal{G}_i = (\mathcal{E}_i, \mathcal{V}), i = 1, \ldots, r \), are **jointly connected** if their **union graph** \( \mathcal{G} = (\bigcup_{i=1}^{r} \mathcal{E}_i, \mathcal{V}) \) is connected

- **Intuition:** Convergence happens if graphs “**connected in a long run**”
Outline

- convergence on fixed graph
- optimization on fixed graph
- robustness to random link failures
- optimization of randomized gossip
• randomized gossip on complete graph arranged as $5 \times 5$ grid

• communication costs

$$C_{ij} \propto d_{ij}^\alpha, \quad d_{ij}: \text{Euclidean distance between node } i \text{ and } j$$

$\alpha = 2.1$  \quad $\alpha = 2.0$  \quad $\alpha = 1.9$
Summary

• distributed consensus in sensor networks
  – convergence
  – robustness
  – optimization

• right level of abstraction leads to interesting yet tractable problems

• mathematical tools used
  – linear algebra
  – spectral graph theory
  – convex optimization
  – duality