Stochastic Modeling, Algorithms and Analysis for Consensus Seeking over Noisy Networks

Minyi Huang

School of Math. & Stats.
Carleton University
Ottawa

Detroit, April 2008
Contents

• Background: from animal behavior to engineering

• Existing research

• Consensus seeking in uncertain environment

• Stochastic algorithms

• Convergence and performance

• Concluding remarks
Animal Behavior: Birds

- A group of **birds** fly with coordination in speed and direction (**Flocking**)
Fish

- Huge number of fish cooperatively move (Schooling)-- Important for search for food or for protection from predators

Couzin et.al.  
*Nature*, 2005
Honeybees

• **Honeybees** select a new home from several candidate sites spotted by scout bees

• What is the mechanism for reaching **consensus**? (Visscher, *Nature*, 2003)

-- Important for avoiding population disintegration
From Birds to Bees: from Flocking/Swarming to Consensus

• Each agent has local information about neighboring agents

• and there is a key group objective (e.g., achieve accurate alignment during motion, or agree on a nest site, etc.)

Such coordination amounts to a form of consensus

Math theory for interpretation?
Applications in Technology

• Examples: a group of autonomous vehicles, or robot teams (formation control)

• In such distributed multi-agent control systems – coordination is critical for safety & the performance of tasks

(above: complex robots)

(below: simple robots)
Formation of Platoon of Vehicles

- Equalize velocity of different vehicles
- Maintain spacing
- Increase highway capacity and improve safety
The Consensus Issue

• For multi-agent coordination, it is usually important to maintain shared information between agents.

• This leads to the key issue of “Agreeing-on-something”. This agreement may

  (1) be the **objective of operation**

  (2) or a **condition** for proceeding to further operation

Hence, in this context, we study consensus problems.
What Is Consensus?

- By consensus seeking, we mean a mechanism whereby the agents adjust their individual values of an **underlying quantity** (e.g., a key state value – angle, velocity, etc.) so as to converge to a **common value**

- In general, **convergence is a primary objective**

- The actually reached limit may be of secondary importance

(small fish schooling)
Background: Models with Exact State Info

- Most existing research on consensus problems assumes **exact state information exchange**
  - Maintaining **certain connectivity** (which can be relaxed to different forms) is crucial for achieving consensus
- The most important analytical tools come from the theory of **stochastic matrices**
Background: Models with Noisy or Inaccurate Measurements

- In a distributed network, it may be impractical to have exact state exchange, for example, due to
  - receiver noise
  - quantization, etc. etc.

- Consensus models with additive noises have attracted the interest of many authors
  - (Ren, Beard and Kingston, ACC’05)
  - (Xiao, Boyd, and Kim, 2007)
  - (Huang and Manton, ACC’07, CDC’07, ACC’08, Preprint’06, Preprint’08)
  - More recent works by various authors …

- Related stochastic models for consensus
  - (Tsitsiklis, Bertsekas, and Athens, IEEE TAC’86) stochastic gradient based algorithms for distributed function optimization
Definitions

- **Definition 1 (weak consensus)** The agents are said to reach weak consensus if

\[
\lim_{t \to \infty} E|x_t^i - x_t^j|^2 = 0, \quad \forall i, j.
\]

- **Definition 2 (mean square consensus)** The agents are said to reach m.s. consensus if \( E|x_t^i|^2 < \infty, \forall i \in \mathcal{N}, t \) and there exists \( x^* \) such that

\[
\lim_{t \to \infty} E|x_t^i - x^*|^2 = 0, \quad \forall i \in \mathcal{N}
\]

- **Definition 3 (strong consensus)** The agents are said to reach strong consensus if there exists \( x^* \) such that

\[
x_t^i \to x^* \quad \text{with probability one for all } i.
\]
Graph Modeling of Networked Agents

- Consider directed graphs (i.e., digraphs) \( G = (\mathcal{N}, \mathcal{E}) \)
- Each agent is denoted by a node
- In a digraph, arrow indicates neighboring relationship & information flow (Example -- right top, agent 1 is a neighbor of agent 2)
- In undirected graph (special case), information is bidirectional
Network Topology Modeling

- For our further analysis: we assume---
  
The digraph contains a spanning tree (special case: connected undirected graphs)

- Implication: information may propagate across the network from one or more points

- In a deterministic model with fixed topology, Ren et. al. (2005) proved existence of a spanning tree is the weakest connectivity condition for consensus
The Measurement Model

- Each agent knows its own state $x_t^i$ exactly,

- and it has noisy observation $y_t^{ik}$ of its neighbors’ states, i.e.,

$$y_t^{ik} = x_t^k + w_t^{ik}, \quad t \in \mathbb{Z}^+, \quad k \in \mathcal{N}_i.$$  

where $w_t^{ik}$ is additive measurement noise.
If Fixed Coefficients Are Used in Averaging: What Happens?

\[
\begin{align*}
    x_{t+1}^1 &= \frac{1}{2}(x_t^1 + y_{t}^{12}) \\
    x_{t+1}^2 &= \frac{1}{3}(x_t^2 + y_{t}^{21} + y_t^{23}) \\
    x_{t+1}^3 &= \frac{1}{2}(x_t^3 + y_{t}^{32})
\end{align*}
\]

This algorithm is essentially a noisy variant of equal-neighbor based algorithms (see related algorithms: Vicsek et. al. PRL’95 Jadbabaie, Lin, Morse’03, etc.)

Measurement noise causes divergence.
Stochastic Approximation

- We use the averaging rule (convex combination):

\[
x_{t+1}^i = (1 - a_t b_{ii}) x_t^i + a_t \sum_{k \in \mathcal{N}_i} b_{ik} y_{t}^{ik}, \quad t \geq 0
\]

\[
b_{ik} > 0 \text{ if and only if } k \in \mathcal{N}_i
\]

\[
b_{ii} = \sum_{k \in \mathcal{N}_i} b_{ik}
\]

- The state of a node remains the same if it has no neighbors. (This happens in leader following)
Stochastic Approximation

• The algorithm in vector form:
  \[ x_{t+1} = x_t + a_t B x_t + a_t \tilde{w}_t \]
  where \( B \) has zero row sum.

\[
B = \begin{bmatrix}
  -b_{11} & b_{12} & \cdots & b_{1n} \\
  b_{21} & -b_{22} & \cdots & b_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  b_{n1} & b_{n2} & \cdots & -b_{nn}
\end{bmatrix}
\]

where

\[ b_{ij} = 0 \text{ if } j \notin \mathcal{N}_i \cup \{i\}. \]

• \( B \) is unstable and may be viewed as the generator of a continuous time Markov chain.
Main Assumptions

• (A1) The measurement noises are independent random variables with zero mean, and independent of initial states. The noise and initial states have bounded second order moment.

• (A2) The digraph contains a spanning tree.

• (A3) The positive step size sequence \( \{a_t, t \geq 0\} \) satisfies:

\[
\sum_{i=0}^{\infty} a_i^2 < \infty, \quad \sum_{i=0}^{\infty} a_i = \infty
\]

Remark: The independence noise sequence assumption may be relaxed (for instance, a sequence of martingale differences for noise vectors)
Illustration with a Two-agent Model

• First, under (A1)-(A2) for noise and step size, it is relatively easy to show (a.s. and m.s.) convergence of the mid-point

\[ z_t = \frac{1}{2} (x_t^1 + x_t^2) \rightarrow z^* \]

• Next it suffices to show (a.s. and m.s.) convergence of the state gap

\[ \xi_t = x_t^1 - x_t^2 \]

• We have

\[ \xi_{t+1} = (1 - 2a_t)\xi_t + a_tv_t, \quad t \geq 0 \]

where

\[ v_t = w_t^{12} - w_t^{21} \]
The Diagram for State Gap

- Key idea: show benefits of reducing noise outweigh the disadvantage of reducing stability
State Gap as Noise Summation

• Denote $\bar{a}_t = 2a_t$ and
  \[
  \Pi_{l,k} = (1 - \bar{a}_l)(1 - \bar{a}_{l-1}) \cdots (1 - \bar{a}_{k+1})a_k
  \]
  for $l > k \geq T_1$. We set $\Pi_{k,k} = a_k$.

• The state gap satisfies
  \[
  \xi_{t+1} = (1 - \bar{a}_t)(1 - \bar{a}_{t-1}) \cdots (1 - \bar{a}_{T_1})\xi_{T_1}
  + \Pi_{t,T_1} v_{T_1}
  + \cdots + \Pi_{t,t-1} v_{t-1}
  + \Pi_{t,t} v_t
  \]

• To prove vanishing gap: Show $\Pi_{t,k}$ or related terms sufficiently small
Convergence Analysis

- Mean square convergence
- Sample path convergence
How to Prove M.S. Convergence?

• Use stochastic Lyapunov analysis to show all individual states attract to each other in mean square.

• Next, show the individual states actually go to the same limit.
The Lyapunov Function

• Let $S^{n \times n}$ be the set of symmetric matrices and denote

$$\mathcal{D} = \{ D \in S^{n \times n} : D \geq 0, \text{Null}(D) = \text{span}\{1_n\} \}$$

• Lemma. Under (A2) and given $D \in \mathcal{D}$, the Degenerate Lyapunov Eqn:

$$QB + B^TQ = -D$$

has a unique solution $Q \in \mathcal{D}$.

• The idea is to show the energy function $V(t) = Ex_t^TQx_t$ will decay to zero.
Energy Decay and Weak Consensus

• Theorem (weak consensus). Under (A1)-(A3),
  (i) There exist $c_1 > 0, c_2 > 0$, and a large $T_c > 0$ such that

  \[ V(t + 1) \leq (1 - a_t c_1 + a_t^2 c_2) V(t) + O(a_t^2) \]

  (ii) Consequently $\lim_{t \to \infty} V(t) = 0$, which implies

  \[ \lim_{t \to \infty} E|\hat{x}_t^i - \hat{x}_t^k|^2 = 0, \forall i, k. \]

  Stay in $\text{span}\{1_n\}$!

Remark: Here it is not clear yet whether they will converge to the same limit. (so, need an extra step!)
Mean Square Consensus

• Lemma. There is a unique probability measure \( \pi \) such that \( \pi^T B = 0 \). Further
  \[
  \pi^T x_{t+1} = \pi^T x_t + a_t \pi^T \tilde{w}_t
  \]
  and \( \pi^T x_t \) converges in m.s.

This Lemma combined with
  \[
  \lim_{t \to \infty} E|x_t^i - x_t^k|^2 = 0, \forall i, k.
  \]

\[\downarrow\]

Theorem. (A1)-(A3) ensures Mean Square consensus (Huang and Manton, ACC’07,08)
Simulations

- Averaging with fixed weights, noise var=0.01
- Stochastic Approx. with decreasing step size

5 individual trajectories

5 individual trajectories
Further Extension to Leader Following

- For leader following, the stochastic Lyapunov analysis is applicable to establish mean square convergence of all other agents’ states to that of the leader (i.e., 4 below).

- Left: use direct averaging    Right: use stochastic approx.

- Diagram showing iterates $x_t$ for two different methods.
Sample Path Behavior

• What is the group behavior along sample paths?

• In fact, this can be characterized by sample path convergence
Sample Path Convergence

- Theorem 1. Under (A1)-(A3), the Stochastic Approx. (SA) algorithm ensures strong consensus (i.e. sample path convergence).

- Remark: for strong consensus, the second order moment condition for the noise may be relaxed.
Sample Path Analysis via Change of Coordinates

- By choosing a suitable change of coordinates
  \[ z_t = [z_t^1, z_t^{(n-1)}]^T = \Phi^{-1}x_t, \] the consensus algorithm may be decomposed into the form (Huang & Manton, ACC’08)

\[
\begin{align*}
  z_{t+1}^1 &= z_t^1 + a_tv_t^1 \\
  z_{t+1}^{(n-1)} &= (I + a_t\tilde{B}_{n-1})z_t^{(n-1)} + a_tv_t^{(n-1)}
\end{align*}
\]

All eigenvalues of \( \tilde{B}_{n-1} \) have negative real parts
Thus, we only need to deal with a random walk and a stable linear SA model
Alternative Proving Tool: Double Array Analysis

• Theorem (Teicher, 1985). Let $\{w, w_t, t \geq\}$ be i.i.d. r.v.’s with zero mean and variance $Q$ and

$\{a_{ki}, 1 \leq i \leq l_k \uparrow \infty, k \geq 1\}$ a double array of constants. Assume

(i) $\max_{1 \leq i \leq l_k} |a_{ki}| h_i = O(1/\log k)$, where $0 < h_i \uparrow$

$\quad h_i = O(i^{1/\delta})$ for some $\delta \in [1, 2]$

(ii) $\sum_{i=1}^{\infty} P\{|w| > h_i\} < \infty$

(iii) $h_i/i \downarrow$, and $\sum_{i=1}^{l_k} |a_{ki}|^2 h_i^{2-\delta} = o(1/\log k)$,

$\quad \sum_{i=1}^{l_k} |a_{ki}|^2 h_i^{2-\delta} = O(1/\log l_k)$

Then

$\lim_{k \to \infty} \sum_{i=1}^{l_k} a_{ki} w_i = 0, \quad a.s.$
Performance?
Performance Analysis

• Due to consensus, denote the limit of the state vector by

\[ x_\infty = [x_\infty^1, \cdots, x_\infty^n]^T = x_\infty^1 1_n \]

• Convergence rate --- Roughly, how small is the error term \( x_t - x_\infty \) when \( t \) is large?
Performance (w/ spanning tree model)

• How fast to reach consensus? (characterized by asymptotic normal.)

• Take step size $a_t = a / t$. Denote $x_t = [x_t^1, \cdots, x_t^n]^T$

• Then under quite standard conditions for noise & coeffic. matrix, we show consensus and furthermore:

$$x_t = x_\infty 1_n + x_t^{e,a} + x_t^{e,b}$$

where $x_t^{e,a}$ depends on future noises & $x_t^{e,b}$ is linear in $x_t$

$$\sqrt{t}x_t^{e,a} \overset{d}{\to} N(0, Q_a), \quad \sqrt{t}x_t^{e,b} \overset{d}{\to} N(0, Q_b)$$

• (H.&M., ACC’08; H. CDC’08 sub) so error decays by rate $\frac{1}{\sqrt{t}}$
Illustration of Asymptotic Normality

- Left bottom \( x_t \)
- Right bottom \( t^{1/2} x_t^{e,b} \)
Additional Uncertainty Factors

- Random communication link failures
- Quantization effects
Random Link Failures

• The stochastic algorithm may still be applied for the randomly varying topology.

• In this case, the coefficient matrix in the consensus algorithm is given as a sequence of random matrices $B_t$ with mean $\overline{B}$.
Random Link Failures (ctn)

- The consensus algorithm

\[ x_{t+1} = x_t + a_t B_t x_t + a_t \text{“noise“} \]

\[ = x_t + a_t \bar{B} x_t + a_t (B_t - \bar{B}) x_t + a_t \text{“noise“} \]

- This algorithm may be viewed as the standard one (with fixed topology) subject to unbiased perturbations.

- In particular, for i.i.d. link failures with additive measurement noise, a perturbed Lyapunov analysis may be applied to establish convergence (Huang and Manton, ACC’08, and Preprint (submitted to IEEE, June’07))
Quantized Data---How to Achieve Convergence?
Probabilistic Quantization (PQ)

- Suppose the state $x^i_t$ is between two quantization levels $r_k < r_{k+1}$

- The idea of PQ is to produce a randomized output $Q_i(t)$ at the quantizer such that it takes the lower and upper level with probability

  \[ p_k = \frac{(r_{k+1} - x^i_t)}{(r_{k+1} - r_k)}, \quad p_{k+1} = 1 - p_k \]

  respectively
Probabilistic Quantization (PQ)

• This approach has been successfully applied for:

  • sensor network signal processing (Xiao, Cui, Luo, and Goldsmith, 2006), and
  • consensus models (Aysal, Coates and Rabbat, 2007)
PQ Combined with SA

- In PQ, we may view and quantization error as an additive uncorrelated noise.
- In the consensus algorithm, a decreasing step size may be further used to damp out the noise. Convergence results may be proved. (Huang, Dey, Nair, and Manton, CDC’08 submitted)
- Left: deterministic quantization; Right: PQ
Concluding Remarks

- Stochastic consensus and convergence

- The key is a decreasing step size for cautious learning

- Stochastic Lyapunov analysis is useful

- Many application opportunities in sensor network signal processing (see, e.g. S. Boyd, J. Hespanha) – networked estim. Prob., sensornet time synchronization, sensornet localization etc. etc. etc.

  Many practical modeling scenarios -- high order (inertia) models and asynchronous algorithms, approximate average consensus, etc. etc.