Consensus over Stochastically Switching Directed Topologies

S. Vanka, V. Gupta and M. Haenggi

Department of Electrical Engineering
University of Notre Dame

1 Introduction

2 Problem Formulation and Main Results

3 Key Ingredients of the Proof
- Expected deviation from average consensus point is zero
- Constructing a martingale
- Bounding martingale differences
- Using the Azuma-Hoeffding Inequality

4 Applications
- Consensus over Fading Channels

5 Conclusions

6 Future Work

Reaching Consensus

- Decentralized algorithm for a system of n agents $1, 2, \ldots, n$ to achieve agreement over value of their state
  - Initial state x_o^i for i^{th} node
- Iterative exchange of information specified by a time-varying communication digraph $G = (V, E)$, where
  - Vertex set V: set of all participating agents $1, 2, \ldots, n$
  - An edge $e_{ji} \triangleq \overrightarrow{(j, i)} \in E$ iff the message from node j is used by i
- Average Consensus Algorithm: $x_t = W_{t-1} x_{t-1}$, where $x_t \triangleq [x_t^{(1)} \; x_t^{(2)} \ldots x_t^{(n)}]^T$ is the state vector
- Average Consensus Point: $x_{\text{av}}^* = n^{-1} \mathbf{1}^* x_o$
- $W_t = I - hL_t$, where L_t is the Laplacian of G_t
- For all t, G_t is balanced $\Rightarrow$ W_t is doubly stochastic $\Rightarrow$ x_t can reach average consensus

Prior Work

- Deterministically varying communication topologies
  - Roots going back to Tsitsiklis (1984)
  - Problems studied include criteria for convergence, optimizing convergence rate etc. (Jadbabaie, Lin and Morse (2003), Xiao and Boyd (2004), Olfati-Saber and Murray (2004), Ren and Beard (2006). . .)
- Randomly varying communication topologies
  - Problems studied include randomized gossip, convergence on random graphs, consensus in networks with noise and packet losses (Boyd *et al.* (2006), Hatano and Mesbahi (2005), Porfiri and Stilwell (2007), Salehi and Jadbabaie (2008), Fagnani and Zampieri (2008), Huang (2007a&b), Hovareshti *et al.* (2008). . .)

Average Consensus over Random Networks

- State after $t+1$ iterations:

$$x_{t+1} = W_t W_{t-1} \cdots W_0 x_o.$$

$\Rightarrow$ Time evolution determined by matrix product $W_t W_{t-1} \ldots W_0$ of random stochastic matrices

- Suppose that
  - $\{W_t : t = 0, 1, \ldots\}$ is matrix-valued i.i.d. sequence
  - All update matrices W_t have positive diagonal elements
  - The linear system $x_{t+1} = \mathbb{E}[W]x_t$ asymptotically reaches consensus
- It is known that
  - The state x_t almost surely reaches consensus (Salehi-Jadbabaie 2008)
  - The consensus point is random variable

Average Consensus over Random Networks

- State after $t+1$ iterations:

$$x_{t+1} = W_t W_{t-1} \cdots W_0 x_o.$$

$\Rightarrow$ Time evolution determined by matrix product $W_t W_{t-1} \ldots W_0$ of random stochastic matrices

- Suppose that
  - $\{W_t : t = 0, 1, \ldots\}$ is matrix-valued i.i.d. sequence
  - All update matrices W_t have positive diagonal elements
  - The linear system $x_{t+1} = \mathbb{E}[W] x_t$ asymptotically reaches consensus
- It is known that
  - The state x_t almost surely reaches consensus (Salehi-Jadbabaie 2008)
  - The consensus point is random variable

Average Consensus over Random Networks

- State after $t+1$ iterations:

$$x_{t+1} = W_t W_{t-1} \cdots W_0 x_o.$$

$\Rightarrow$ Time evolution determined by matrix product $W_t W_{t-1} \ldots W_0$ of random stochastic matrices

- Suppose that
  - $\{W_t : t = 0, 1, \ldots\}$ is matrix-valued i.i.d. sequence
  - All update matrices W_t have positive diagonal elements
  - The linear system $x_{t+1} = \mathbb{E}[W]x_t$ asymptotically reaches consensus
- It is known that
  - The state x_t almost surely reaches consensus (Salehi-Jadbabaie 2008)
  - The consensus point is random variable

Problem Formulation

- Linear dynamics $x_{t+1} = W_t x_t$, fixed initial state x_0
- Ideally: Reach average consensus point $\mathbf{1}(n^{-1}\mathbf{1}^* x_0)$

  But what if not all W_t are balanced?

- Matrix sequence W_t :
  - An i.i.d. sequence of stochastic matrices
  - All W_t have positive diagonal entries
  - The system $x(t+1) = \mathbb{E}[W]x(t)$ reaches average consensus

  $\Rightarrow x(t)$ almost surely reaches consensus

- Can we quantify this deviation from the average consensus point?

Problem Formulation

- Linear dynamics $x_{t+1} = W_t x_t$, fixed initial state x_0
- Ideally: Reach average consensus point $\mathbf{1}(n^{-1}\mathbf{1}^* x_0)$

  But what if not all W_t are balanced?
- Matrix sequence W_t :
  - An i.i.d. sequence of stochastic matrices
  - All W_t have positive diagonal entries
  - The system $x(t+1) = \mathbb{E}[W]x(t)$ reaches average consensus

  $\Rightarrow x(t)$ almost surely reaches consensus
- Can we quantify this deviation from the average consensus point?

Problem Formulation

- Linear dynamics $x_{t+1} = W_t x_t$, fixed initial state x_0
- Ideally: Reach average consensus point $\mathbf{1}(n^{-1}\mathbf{1}^* x_0)$

  But what if not all W_t are balanced?
- Matrix sequence W_t :
  - An i.i.d. sequence of stochastic matrices
  - All W_t have positive diagonal entries
  - The system $x(t+1) = \mathbb{E}[W]x(t)$ reaches average consensus

$\Rightarrow$ $x(t)$ almost surely reaches consensus

- Can we quantify this deviation from the average consensus point?

Problem Formulation

- Linear dynamics $x_{t+1} = W_t x_t$, fixed initial state x_0
- Ideally: Reach average consensus point $\mathbf{1}(n^{-1}\mathbf{1}^* x_0)$

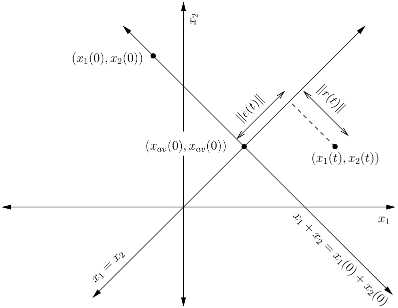
  But what if not all W_t are balanced?
- Matrix sequence W_t :
  - An i.i.d. sequence of stochastic matrices
  - All W_t have positive diagonal entries
  - The system $x(t+1) = \mathbb{E}[W]x(t)$ reaches average consensus

  $\Rightarrow x(t)$ almost surely reaches consensus
- Can we quantify this deviation from the average consensus point?

Geometric Interpretation

- Instantaneous average $x_{av}(t) \triangleq \frac{\mathbf{1}^* x_t}{n}$
- For balanced communication graphs $x_{av}(t) = x_{av}(0) =$ average consensus point
- Interested in characterizing the deviation $x_t - \mathbf{1}x_{av}(0)$ of *instantaneous state from average consensus point*

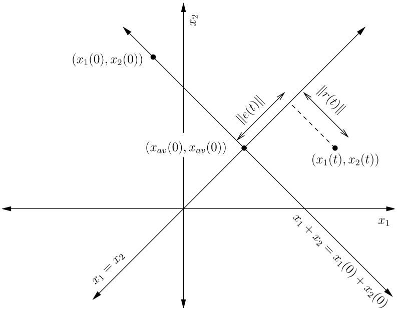
$n = 2$



Geometric Interpretation

- Now x_t can be written as a sum of e_t and r_t where
  - $e_t \triangleq \mathbf{1}(x_{av}(0) - x_{av}(0))$: Deviation of Instantaneous Average
  - $r_t \triangleq x_t - \mathbf{1}x_{av}(t)$: Disagreement
- Define $\delta_t \triangleq x_{av}(t) - x_{av}(0)$, so that $\|e_t\| = \delta_t$

$n = 2$



Our Results

For a fixed initial state x_0, $\bar{W} \triangleq \mathbb{E}[W]$, $P \triangleq I - n^{-1}\mathbf{1}\mathbf{1}^*$. Let $1 = \lambda_1 \geq \lambda_2 \geq \cdots \lambda_n > -1$ be the eigenvalues of $\bar{W}$, and $\mu \triangleq \max(|\lambda_2|, |\lambda_n|)$. Then for all $\epsilon > 0$ and t:

- Deviation δ_t of instantaneous average from the average consensus point

$$\boxed{\mathbb{P}(|\delta_t| \geq \epsilon) \leq \min(1, 2\exp(-\epsilon^2\beta_t))}$$

- Distance r_t from the consensus subspace

$$\boxed{\mathbb{P}(\|r_t - P\bar{W}^t x_o\|_\infty \geq \epsilon) \leq \min(1, 2\exp(-\epsilon^2\beta_t))}$$

where $\beta(t) \triangleq \frac{1-\mu^2}{2C^2\|x_o\|_\infty^2(1-\mu^{2t})}$.

- Will highlight the key results used to derive the concentration bound on the deviation δ_t from the average consensus point
- Proof has four main ingredients
  - Expected deviation from average consensus point is zero
  - Constructing a martingale
  - Bounding the differences between successive terms of this sequence by leveraging the connectivity of $\bar{W}$
  - Using the Azuma-Hoeffding inequality to bound the deviation δ_t from its mean

Expected Deviation in the Consensus Subspace

Can show that $\mathbb{E}[\delta_t] = 0$, exploiting:

(1) The matrix sequence $\{W_t\}$ is i.i.d.

(2) $\bar{W}$ is doubly stochastic

$$\mathbb{E}[\delta_t] \quad = \quad n^{-1}\mathbf{1}^* \left(\mathbb{E}\left[\prod_{k=0}^{t-1} W_k x_o \right] - x_o \right)$$

Expected Deviation in the Consensus Subspace

Can show that $\mathbb{E}[\delta_t] = 0$, exploiting:

(1) The matrix sequence $\{W_t\}$ is i.i.d.

(2) $\bar{W}$ is doubly stochastic

$$\begin{aligned} \mathbb{E}[\delta_t] &= n^{-1}\mathbf{1}^* \left(\mathbb{E}\left[\prod_{k=0}^{t-1} W_k x_o \right] - x_o \right) \\ &\overset{(1)}{=} n^{-1}\mathbf{1}^* (\bar{W}^t - I) x_o \end{aligned}$$

Expected Deviation in the Consensus Subspace

Can show that $\mathbb{E}[\delta_t] = 0$, exploiting:

(1) The matrix sequence $\{W_t\}$ is i.i.d.

(2) $\bar{W}$ is doubly stochastic

$$\begin{aligned}\mathbb{E}[\delta_t] &= n^{-1}\mathbf{1}^*\left(\mathbb{E}\left[\prod_{k=0}^{t-1} W_k x_o\right] - x_o\right)\\ &\overset{(2)}{=} n^{-1}(\mathbf{1}^*\bar{W}^t - \mathbf{1}^*)x_o = 0.\end{aligned}$$

Definition of a martingale

Definition (Chap. 12, Mitzenmacher and Upfal, 2005)

A sequence $Y_1, Y_2, \ldots, Y_t$ is a martingale with respect to a sequence $X_1, X_2, \ldots, X_t$ if:

1. $\mathbb{E}[|Y_k|] < \infty$ for $k = 1, 2, \ldots, t$.
2. $\mathbb{E}(Y_{k+1} \mid X_1, \ldots, X_k) = Y_k$, for $k = 1, \ldots, t-1$.

The sequence $\{X_k\}$ is called a filtration.

For the current problem, define a sequence $\{Y_k(t)\}$ for $k = 1, 2, \ldots t$:

$$\begin{aligned} X_k &= W_{t-k} \\ Y_k(t) &\triangleq \mathbb{E}[\delta_t | W_{t-1}, W_{t-2}, \ldots W_{t-k}] \end{aligned}$$

Definition of a martingale

Definition (Chap. 12, Mitzenmacher and Upfal, 2005)

A sequence $Y_1, Y_2, \ldots, Y_t$ is a martingale with respect to a sequence $X_1, X_2, \ldots, X_t$ if:

1. $\mathbb{E}[|Y_k|] < \infty$ for $k = 1, 2, \ldots, t$.
2. $\mathbb{E}(Y_{k+1} \mid X_1, \ldots, X_k) = Y_k$, for $k = 1, \ldots, t-1$.

The sequence $\{X_k\}$ is called a filtration.

For the current problem, define a sequence $\{Y_k(t)\}$ for $k = 1, 2, \ldots t$:

$$\begin{aligned} X_k &= W_{t-k} \\ Y_k(t) &\triangleq \mathbb{E}[\delta_t | W_{t-1}, W_{t-2}, \ldots W_{t-k}] \end{aligned}$$

Verifying that $\{Y_k(t)\}$ is a martingale w.r.t. $\{W_{t-k}\}$

Property 1: $\mathbb{E}[|Y_k(t)|] < \infty$. Follows from the fact that $|Y_k(t)|$ is upper-bounded by $2\|x_o\|_\infty$

Property 2: $\mathbb{E}[Y_{k+1}(t) \mid W_{t-1}, W_{t-2}, \cdots, W_{t-k}] = Y_k$. Follows from definition:

$$
\begin{aligned}
&\mathbb{E}[Y_{k+1}(t) \mid W_{t-1}, W_{t-2}, \cdots, W_{t-k}] \\
=&\mathbb{E}[\mathbb{E}[\delta_t \mid W_{t-1}, \cdots, W_{t-k+1}] \mid W_{t-1}, \cdots, W_{t-k}] \\
=&\mathbb{E}[\delta_t \mid W_{t-1}, \cdots, W_{t-k}] = Y_k.
\end{aligned}
$$

Bounding martingale differences

$$\|Y_{k+1}(t) - Y_k(t)\|_\infty$$

Bounding martingale differences

$$
\begin{aligned}
&\|Y_{k+1}(t) - Y_k(t)\|_\infty \\
&= \|n^{-1}\mathbf{1}^* W_{t-1}\cdots W_{t-k+1}(\bar{W} - W_{t-k})\bar{W}^{t-k}x_o\|_\infty \\
&\overset{(1)}{\leq} \underbrace{\|n^{-1}\mathbf{1}^*\|_\infty}_{1} \prod_{l=1}^{l=k-1} \underbrace{\|W_{t-l}\|_\infty}_{1} \|(\bar{W} - W_{t-k})\bar{W}^{t-k}\|_\infty \|x_o\|_\infty \\
&\leq \|(\bar{W} - W_{t-k})\bar{W}^{t-k}\|_\infty \|x_o\|_\infty
\end{aligned}
$$

Key Ideas

1. ℓ_∞-norm is sub-multiplicative

Bounding martingale differences

$$
\begin{aligned}
&\|Y_{k+1}(t) - Y_k(t)\|_\infty \\
&\leq \|(\bar{W} - W_{t-k})\bar{W}^{t-k}\|_\infty \|x_o\|_\infty \\
&\stackrel{(2)}{\leq} \|\cancelto{0}{(\bar{W} - W_{t-k})(n^{-1}\mathbf{1}\mathbf{1}^*)} + (\bar{W} - W_{t-k})C_1\mu^{t-k})\|x_o\|_\infty
\end{aligned}
$$

Key Ideas

1. ℓ_∞-norm is sub-multiplicative
2. $\bar{W}$ and W_{t-k} are stochastic

Bounding martingale differences

$$\|Y_{k+1}(t) - Y_k(t)\|_\infty \stackrel{(1,3)}{\leq} C\mu^{t-k}\|x_o\|_\infty$$

Key Ideas

1. ℓ_∞-norm is sub-multiplicative
2. $\bar{W}$ and W_{t-k} are stochastic
3. Spectral gap of of $\bar{W}$ is positive

Using the Azuma-Hoeffding Inequality

Theorem (Chap. 12, Mitzenmacher and Upfal '05)

Suppose $\{Y_k : k = 0, 1, \ldots\}$ *is a martingale and* $|Y_{k+1} - Y_k| < c_k$ *almost surely for all* k. *Then, for all positive integers* t *and* $\epsilon > 0$,

$$\mathbb{P}(|Y_t - Y_0| \geq \epsilon) \leq 2\exp\left(-\frac{\epsilon^2}{2\sum_{k=1}^{t} c_k^2}\right).$$

For our martingale:

- $Y_t(t) = \delta_t$, $Y_0(t) = \mathbb{E}[\delta_t] = 0$.
- $\|Y_{k+1}(t) - Y_k(t)\|_\infty \leq C\mu^{t-k}\|x_o\|_\infty \triangleq c_k(t)$

Observing that $2\sum_{k=0}^{k=t} c_k^2(t) = \beta_t^{-1}$ and using the result:

$$\boxed{\mathbb{P}(|\delta_t| \geq \epsilon) \leq 2\exp(-\epsilon^2\beta_t)}$$

Using Concentration Inequalities to Bound Higher Moments of δ_t

Theorem (Chap. 13, Folland 1999)

If X is a positive random variable, its r^{th} moment is given by

$$\mathbb{E}[X^r] = \int |X|^r d\mu \equiv r \int_0^\infty u^{r-1} \gamma(u) du,$$

where $\gamma(u) \triangleq \mathbb{P}(X \geq u)$. If $\gamma(u) \leq \lambda(u)$ for all u,

$$\mathbb{E}[X^r] \leq r \int_0^\infty u^{r-1} \lambda(u) du.$$

$\Rightarrow$ Bounds every moment of a random variable.

Asymptotic Distribution of δ_t

- Results valid for all $t \geq 0$.
- Since $\lim_{t\to\infty} x(t) = \mathbf{1}\alpha$ almost surely:

$$\lim_{t\to\infty} \mathbb{P}(|\delta_t| \geq \epsilon) \leq \lim_{t\to\infty} 2\exp(-\epsilon^2\beta_t)$$

holds almost surely. But

- $\lim_{t\to\infty} \delta_t = \alpha - x_{\text{av}}(0)$ almost surely,
- $\beta_\infty \triangleq \lim_{t\to\infty} \beta_t = (1-\mu^2)(2C^2\|x_o\|_\infty^2)^{-1}$.

Therefore

$$\boxed{\mathbb{P}(|\alpha - x_{\text{av}}(0)| \geq \epsilon) \leq 2\exp(-\epsilon^2\beta_\infty)}$$

System Model

- n consensus-seeking nodes, labelled $V = \{1, 2, \ldots, n\}$ located at $\{r_1, r_2, \ldots, r_n\}$.
- Node i initially holds a value x_o^i, $i = 1, 2, \ldots, n$.
- Edges in interconnection topology established by wireless links, with node i transmitting with power P_i
- Synchronous state update after all nodes transmit once

Communication Model

- Block fading Rayleigh channels, independent across iterations
- Medium access using Time Division Multiple Access (TDMA) protocol
- Link Failure Model:
  - A node i can successfully receive a message from node j iff the Signal-to-Noise-Ratio (SNR_{ji}) exceeds a known threshold Θ
  - A failed link is said to be in *outage*

Communication Model

- Block fading Rayleigh channels, independent across iterations
- Medium access using Time Division Multiple Access (TDMA) protocol
- Link Failure Model:
  - A node i can successfully receive a message from node j iff the Signal-to-Noise-Ratio (SNR_{ji}) exceeds a known threshold Θ
  - A failed link is said to be in *outage*

Communication Model

- Block fading Rayleigh channels, independent across iterations
- Medium access using Time Division Multiple Access (TDMA) protocol
- Link Failure Model:
  - A node i can successfully receive a message from node j iff the Signal-to-Noise-Ratio (SNR_{ji}) exceeds a known threshold Θ
  - A failed link is said to be in *outage*

Communication Model

- Block fading Rayleigh channels, independent across iterations
- Medium access using Time Division Multiple Access (TDMA) protocol
- Link Failure Model:
  - A node i can successfully receive a message from node j iff the Signal-to-Noise-Ratio (SNR_{ji}) exceeds a known threshold Θ
  - A failed link is said to be in *outage*

Randomness in information exchange

- Fading causes random link outages
- Information exchange specified by a *random* digraph over n nodes, with each edge $\overrightarrow{(j,i)}$ present with a probability

$$p_{ji} = \exp\left(\frac{-\Theta N_0}{P_j \|r_j - r_i\|^{-\alpha}}\right),$$

where N_0 is the noise variance and $\alpha \geq 2$ is the path loss exponent. L_{ij} is a Bernoulli random variable with parameter p_{ji}

- Therefore $\bar{W} = I - h\bar{L}$, where

$$\bar{L}_{ij} = \begin{cases} -p_{ji} & i \neq j \\ \sum_{j \neq i} p_{ji} & i = j \end{cases}$$

Randomness in information exchange

- Fading causes random link outages
- Information exchange specified by a *random* digraph over n nodes, with each edge $\overrightarrow{(j,i)}$ present with a probability

$$p_{ji} = \exp\left(\frac{-\Theta N_0}{P_j \|r_j - r_i\|^{-\alpha}}\right),$$

where N_0 is the noise variance and $\alpha \geq 2$ is the path loss exponent.
L_{ij} is a Bernoulli random variable with parameter p_{ji}

- Therefore $\bar{W} = I - h\bar{L}$, where

$$\bar{L}_{ij} = \begin{cases} -p_{ji} & i \neq j \\ \sum_{j \neq i} p_{ji} & i = j \end{cases}$$

Randomness in information exchange

- Fading causes random link outages
- Information exchange specified by a *random* digraph over n nodes, with each edge $\overrightarrow{(j,i)}$ present with a probability

$$p_{ji} = \exp\left(\frac{-\Theta N_0}{P_j \|r_j - r_i\|^{-\alpha}}\right),$$

where N_0 is the noise variance and $\alpha \geq 2$ is the path loss exponent. L_{ij} is a Bernoulli random variable with parameter p_{ji}

- Therefore $\bar{W} = I - h\bar{L}$, where

$$\bar{L}_{ij} = \begin{cases} -p_{ji} & i \neq j \\ \sum_{j \neq i} p_{ji} & i = j \end{cases}$$

Randomness in information exchange

- Fading causes random link outages
- Information exchange specified by a *random* digraph over n nodes, with each edge $\overrightarrow{(j,i)}$ present with a probability

$$p_{ji} = \exp\left(\frac{-\Theta N_0}{P_j \|r_j - r_i\|^{-\alpha}}\right),$$

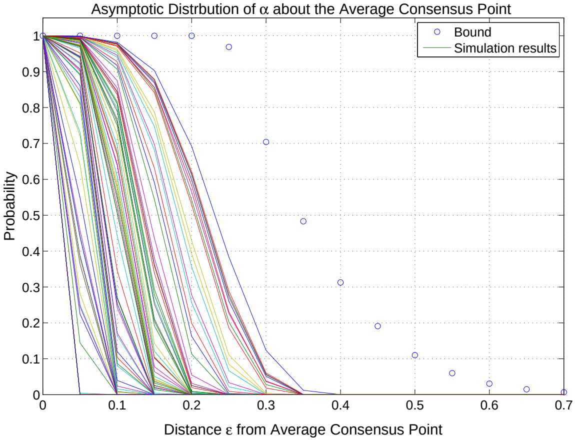
where N_0 is the noise variance and $\alpha \geq 2$ is the path loss exponent. L_{ij} is a Bernoulli random variable with parameter p_{ji}

- Therefore $\bar{W} = I - h\bar{L}$, where

$$\bar{L}_{ij} = \begin{cases} -p_{ji} & i \neq j \\ \sum_{j \neq i} p_{ji} & i = j \end{cases}$$

Example

$n = 2$ nodes, Link failures $p_{01} = 0.2, p_{10} = 0.8, ||x_o|| = 0.4$



Conclusions

- Framework developed to study probabilistic state evolution in average consensus algorithm
- Some Applications
  - Concentration Bounds for State in Randomized Average Consensus Algorithms
  - Distribution of products of random stochastic matrices

Future Work

- Obtaining sharper concentration bounds for instantaneous state average for consensus in wireless networks
- Extending convergence results with non-iid update matrices
- Applying techniques developed to the performance of distributed randomized algorithms