Dynamic Median Consensus over Random Networks

Shuhua Yu, Yuan Chen, Soummya Kar

Electrical and Computer Engineering, Carnegie Mellon University

Dec 16th, 2021

Background

Distributed computing

Consider a distributed system where data are distributed across networked agents.

- Majority vote.
- Computing average, median, quantiles.
- Distributed data-driven optimization.

Distributed setup

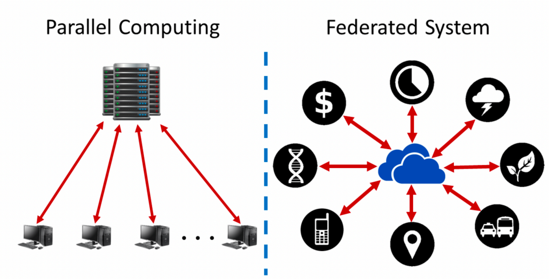

Figure: Distributed setups with central processor.[1]

- Requires a computation coordinator.
- The coordinator may be a communication bottleneck.

[1] Zhixiong Yang, Arpita Gang, and Waheed U Bajwa. "Adversary-resilient distributed and decentralized statistical inference and machine learning: An overview of recent advances under the Byzantine threat model". In: *IEEE Signal Processing Magazine* 37.3 (2020), pp. 146–159.

Decentralized setup

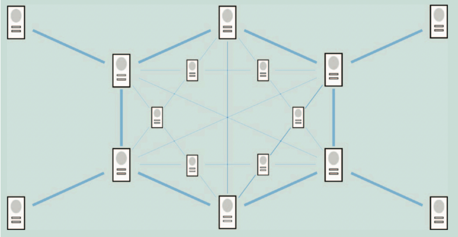

Figure: Decentralized setup.[2]

- Local computations and communications.
- No central compute node.
- Less communication per node.
- Applications in sensor networks, unmanned aerial vehicles.

[2] Yuan Chen, Soummya Kar, and José MF Moura. "The internet of things: Secure distributed inference". In: *IEEE Signal Processing Magazine* 35.5 (2018), pp. 64–75.

Consensus

Consensus in decentralized setup:

Consensus

Suppose each node n holds a decision variable x_n^t at time t. All nodes reach consensus if for any $m \neq n$, $\lim_{t \to \infty} |x_n^t - x_m^t| = 0$.

Average consensus

Suppose each node n holds some initial state θ_n, and holds a local estimate x_n^t for $\bar{\theta} := N^{-1} \sum_{n=1}^{N} \theta_n$ at time t. All nodes reach average consensus if for any $n \in [N]$, $\lim_{t \to \infty} x_n^t = \bar{\theta}$.

Average vs median

Average is vulnerable to outliers.

Figure: One outlier can dominate the average

Median is more robust to outliers[3].

	m1	m2	m3	m4	m5	average	median
record 1	6.27	6.34	6.25	6.31	6.28	6.29	6.28
record 1	6.27	6.34	6.25	63.1	6.28	17.65	6.28

[3] Peter J Rousseeuw and Mia Hubert. "Robust statistics for outlier detection". In: *Wiley interdisciplinary reviews: Data mining and knowledge discovery* 1.1 (2011), pp. 73–79.

Median consensus

L_1 vs L_2 minimization

Given a set of scalars $\theta_1, \ldots, \theta_N$,

$$\text{mean:} \quad \underset{x \in \mathbb{R}}{\operatorname{argmin}} \frac{1}{N} \sum_{n=1}^{N} (x - \theta_n)^2,$$

$$\text{median:} \quad \underset{x \in \mathbb{R}}{\operatorname{argmin}} \frac{1}{N} \sum_{n=1}^{N} |x - \theta_n| \,.$$

Median consensus

Suppose each node n holds some initial state θ_n, and holds a local estimate x_n^t for median$(\theta_1, \ldots, \theta_N)$. Median consensus is achieved if all nodes reach consensus and $\lim_{t \to \infty} x_n^t \in \text{median}(\theta_1, \ldots, \theta_N)$.

Application: a subCULTron project

- A decentralized multi-robot systems.

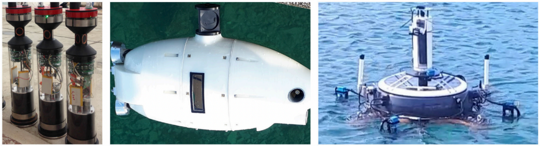

Figure: aMussels (left), aFish (middle) and aPad (right).[4]

- Underwater swarm: 100 aMussels, 10 aFish and 5 aPads.
- Goal: measuring environment parameters such as oxygen or turbidity.
- Challenges: sensors prone to faults/errors/outliers.

[4] Goran Vasiljević et al. "Dynamic Median Consensus for Marine Multi-Robot Systems Using Acoustic Communication". In: *IEEE Robotics and Automation Letters* 5.4 (2020), pp. 5299–5306.

Problem setup

Dynamic median consensus

Goal: dynamic median consensus

Let Θ be the medians of a set of distinct numbers $\{\theta_n\}_{n\in[N]}$. Suppose each node n maintains a local estimate x_n^t, the goal is for all nodes to reach median consensus, i.e., reach consensus and $\lim_{t\to\infty} x_n^t \in \Theta$, with

- local dynamic observations on θ_n,
- local communications in random networks.

Dynamic observations

Node n observes that

$$\theta_n^t = \theta_n + v_n^t + w_n^t,$$

where

- v_n^t is some decaying bias,
- w_n^t is some white noise.

Assumptions

- For every $n \in [N]$, $|v_n^t| \leq v_0(t+1)^{-\delta}$ a.s. for some positive constants δ, v_0.
- Each w_n^t satisfies that $\mathbb{E}\left(w_n^t\right) = 0, \mathrm{Var}(w_n^t) < \infty$.
- $\{w_n^t\}_{n \in [N], t \geq 0}$ is i.i.d. distributed over time and across agents.
- $\{v_n^t\}, \{w_n^t\}$ are mutually independent.

Random networks

Time-varying, undirected, simple random graph $G^t = ([N], E^t)$.

$$
\begin{aligned}
\Omega_n^t &= \{m : (m,n) \in E^t\}; \\
D^t[n,n] = \left|\Omega_n^t\right|, D^t[m,n] &= 0 \text{ if } m \neq n; \\
A^t[m,n] = 1 \text{ if } (m,n) \in E^t, &\text{ otherwise } 0; \\
L^t &= D^t - A^t.
\end{aligned}
$$

We assume $\{G^t\}$ is connected on average.[5]

Assumption

We assume $\{L^t\}$ is an i.i.d. sequence with $\lambda_2(\mathbb{E}(L^t)) > 0$.

[5] Soummya Kar, José MF Moura, and Kavita Ramanan. "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication". In: *IEEE Transactions on Information Theory* 58.6 (2012), pp. 3575–3605.

Random networks: example

A simple connected and undirected network $G = ([N], E)$ where each link in E has dropout probability in $[0, 1)$.

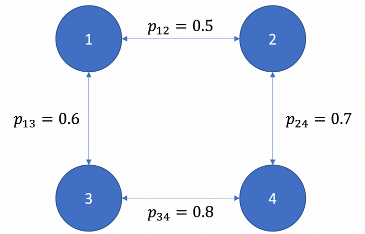

Figure: A simple network with random dropout

Contributions

- Prior works consider determinstically bounded observation noises[6].
- Prior works require the network to be connected all the time[7].
- We relax these assumptions, provide a *consensus+innovations* type algorithm with variance reduction and clipped innovations, and show almost sure convergence in sublinear rate.

[6] Zohreh Al Zahra Sanai Dashti, Carla Seatzu, and Mauro Franceschelli. "Dynamic consensus on the median value in open multi-agent systems". In: *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE. 2019, pp. 3691–3697.

[7] Alessandro Pilloni et al. "Robust distributed consensus on the median value for networks of heterogeneously perturbed agents". In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE. 2016, pp. 6952–6957.

DMED Algorithm

Consensus+innovations

A distributed inference framework. For instance,

$$x_n^{t+1} = \underbrace{x_n^t - \beta_t \sum_{l \in \Omega_n^t} (x_n^t - x_l^t)}_{\textit{consensus}} + \underbrace{\alpha_t K_n^t [H_n^\top R_n^{-1} (y_n^t - H_n x_n^t)]}_{\textit{innovations}}$$

is a *consensus+innovations*[8] type distributed linear estimator, where K_n^t, H_n, R_n are local variables only known to agent n.

[8] Soummya Kar and José MF Moura. "Consensus+ innovations distributed inference over networks: cooperation and sensing in networked systems". In: *IEEE Signal Processing Magazine* 30.3 (2013), pp. 99–109.

Variance reduction

Recall that $\theta_n^t = \theta_n + v_n^t + w_n^t$, where $|v_n^t| \leq v_0(t+1)^{-\delta}$ a.s. and w_n^t is a white noise. Let $\bar{\theta}_n^0 = \theta_n^0$ and choose η_t,

$$\bar{\theta}_n^{t+1} \leftarrow (1-\eta_t)\bar{\theta}_n^t + \eta_t \theta_n^t.$$

Why:

- if $v_n^t = 0$, take $\eta_t = 1/(t+1)$, use LLN,
- if $w_n^t = 0$, take $\bar{\theta}_n^t = \theta_n^t$.

Local convergence

Let $\eta_t = \eta_0(t+1)^{-\tau_4}$. If $\delta \geq 1$, take any $0 < \tau_4 < 1$, otherwise take $\delta \leq \tau_4 < 1$. Then, for every $0 < \epsilon \leq \tau_4$, we have

$$\lim_{t\to\infty} (t+1)^{\tau_4-\epsilon} \left|\bar{\theta}_n^t - \theta_n\right|^2 = 0, \text{ a.s.}$$

Clipped innovations

Consensus+Clipped Innovations,

$$x_n^{t+1} \leftarrow x_n^t - \beta_t \sum_{m \in \Omega_n^t} (x_n^t - x_m^t) - \alpha_t \mathrm{clip}(x_n^t - \bar{\theta}_n^t, \gamma_t),$$

where

$$\mathrm{clip}(x, \gamma_t) = \begin{cases} (x/\left|x\right|)\gamma_t, & |x| \geq \gamma_t, \\ x & \text{otherwise.} \end{cases}$$

Motivation:

- Global objective is $\min_\theta \sum_{n \in [N]} |\theta - \theta_n|$. Local cost function $|x - \theta_n|$ with subgradient $\mathrm{sign}(x - \theta_n)$.
- $\alpha_t \mathrm{clip}(x_n^t - \bar{\theta}_n^t, \gamma_t) = \alpha_t \gamma_t \mathrm{sign}(x_n^t - \bar{\theta}_n^t)$ or $\alpha_t (x_n^t - \bar{\theta}_n^t)$.
- Clipping operation "smooth" the algorithm behavior when x_n^t is close to Θ.

Algorithm 1: Distributed Median Estimator for Dynamic observations

Input: $\{\alpha_t\}_{t\geq 0}, \{\beta_t\}_{t\geq 0}, \{\gamma_t\}_{t\geq 0}, \{\eta_t\}_{t\geq 0}$;
Initialization: Set arbitrary x_n^0 and $\bar{\theta}_n^0 = \theta_n^0$ for all $n \in [N]$;
for $t = 0, \ldots, T$ **do**
  for $n = 1, \ldots, N$ *in parallel* **do**
    VR: $\bar{\theta}_n^{t+1} \leftarrow (1 - \eta_t)\bar{\theta}_n^t + \eta_t \theta_n^t$;
    C+Cl: $x_n^{t+1} \leftarrow x_n^t - \beta_t \sum_{m \in \Omega_n^t}(x_n^t - x_n^t) - \alpha_t \text{clip}(x_n^t - \bar{\theta}_n^t, \gamma_t)$
  end
end
Output: $\{x_n^T\}_{n\in[N]}$

Convergence

Define $\text{dist}(x_n^t, \Theta) = \min_{\theta \in \Theta} |x_n^t - \theta|$.

Asymptotic convergence

Let $\alpha_t = \alpha_0(t+1)^{-\tau_1}, \beta_t = \beta_0(t+1)^{-\tau_2}, \gamma_t = \gamma_0(t+1)^{-\tau_3}$. Choose $0 < \tau_2 < \tau_1 < 1, \tau_3 < \min\{1 - \tau_1, \tau_4/2\}$. Then, we have for all $n \in [N]$, $\lim_{t \to \infty} (t+1)^{\tau_3} \text{dist}(x_n^t, \Theta) = 0$ a.s.

- Take $\tau_1 = 0.5, \tau_2 = 0.3, \tau_3 = 0.4, \tau_4 = 0.9$, we obtain $O((t+1)^{-0.4})$ convergence rate.
- Consensus error decays faster at $O\left((t+1)^{-(\tau_1 - \tau_2 + \tau_3)}\right)$ a.s.
- The theorem holds for every $\tau_3 < \min\{1 - \tau_1, \tau_4/2\}$.
- Since $0 < \tau_4 < 1$, best guaranteed rate is near $O(\sqrt{t})$.

Experiments: networks

Two geometric random graphs with 40 nodes.

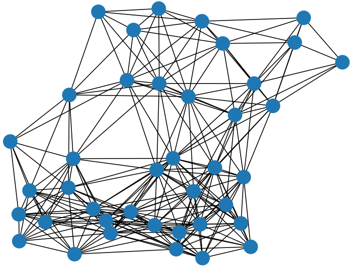

Figure: Graph 1

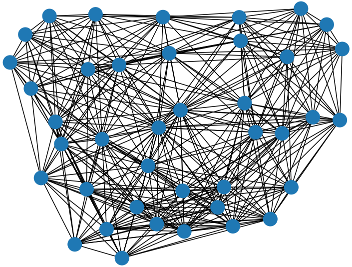

Figure: Graph 2

Experiments: τ_1, τ_4

Data setup: Graph 2 with dropout 0.1. $\theta_n = 5n, v_n = 50/(t+1), w_n^t \sim \mathcal{N}(0,4), \alpha_t = (t+1)^{-\tau_1}, \beta_t = 0.1(t+1)^{-\tau_2}, \gamma_t = 20(t+1)^{-\tau_3}, \eta_t = 10(t+1)^{-\tau_4}$.

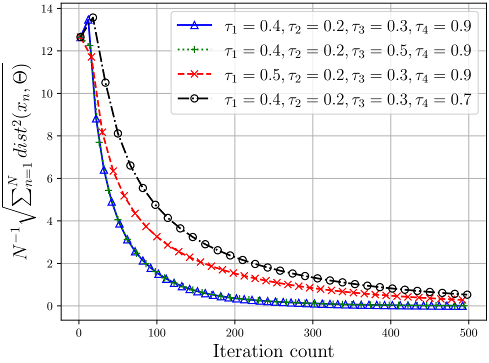

Figure: Tuning τ_1 and τ_4.

Experiments: τ_2, τ_3

Fix (variance reduction rate) $\tau_4 = 0.9$.

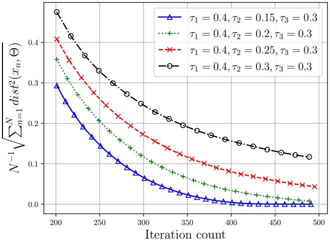

Figure: Tuning τ_2.

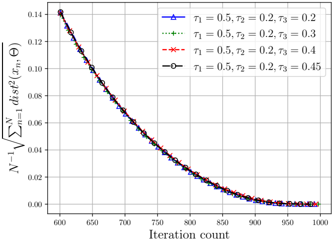

Figure: Tuning τ_3.

Experiments: clipping vs sign

Fix $\tau_4 = 0.9$, and replace $\alpha_t \text{clip}(x_n^t - \bar{\theta}_n^t, \gamma_t)$ with $\alpha_t \gamma_t \text{sign}(x_n^t - \bar{\theta}_n^t)$.

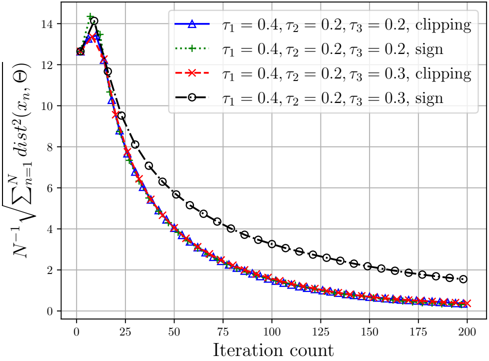

Figure: *Clipped innovations* vs *signed innovations*

Experiments: connectivity

Experiments on networks with different connectivity, $\lambda_2(L_1) \approx 1.8, \lambda_2(L_2) \approx 7.2$.

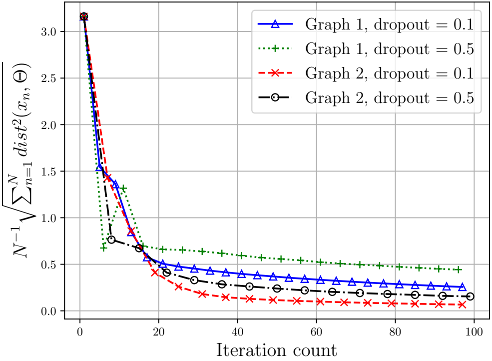

Figure: Experiments in different connectivities

Experiments: summary

- As suggested by the theorem: τ_1, τ_4 determines the convergence rates.
- Empirical findings: convergence rates are affected by τ_2, and better connectivity accelerates convergence.

Discussions

Discussions

Conclusions:

- Proved median consensus for dynamic observations in random networks.
- Demonstrated effectiveness and illustrated parameters on synthetic data.

Directions:

- Understand the difference between *clipped* and *signed innovations*.
- Understand how the connectivity affects the convergence rate.
- Can we use this algorithm to robustify other decentralized algorithms?

Questions, comments, suggestions?