Resilient decentralized optimization in multi-agent networks with data injection attack

Shuhua Yu, Yuan Chen, Soummya Kar

2021 55th Asilomar Conference on Signals, Systems, and Computers | 978-1-6654-5828-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/IEEECONF53345.2021.9723327

***Abstract*—This paper studies resilient decentralized optimization over multi-agent networks. In particular, we consider the scenario where all networked agents are supposed to parallelly minimize the same objective function with a unique minimizer, but some agents are under data injection attack that perturb their minimizers away from the true minimizer. The goal is to ensure that all agents resiliently recover the true minimizer. We propose a *consensus+innovations* type of algorithm with *signed innovations*, to track the coordinate-wise median of all local (possibly perturbed) minimizers. We assume that all local iterates are *sublinearly convergent*, and the inter-agent undirected communication network is *connected on average*. We show that, as long as at least one agent is unattacked, and the median of all local minimizers is the true minimizer, the proposed decentralized algorithm asymptotically converges to the true minimizer almost surely at an sublinear rate. Numerical experiments are provided to demonstrate the effectiveness of the algorithm.**

***Index Terms*—resilience, multi-agent systems, median consensus, decentralized optimization**

I. Introduction

Decentralized learning and inference over multi-agent networks have attracted increasing attention in many applications like *Internet of Things* [1] and learning paradigms such as *federated learning* [2]. Meanwhile, the distributed nature of multi-agent networks casts security concerns on data integrity and agent integrity. As a result, there is a rich literature of prior art providing an abundance of approaches to robustify decentralized learning and inference procedures, such as [1], [3] and references therein.

This paper studies optimization in the decentralized setup where the adversary only attack the distributed data, but is not allowed to manipulate the networked agents [4], in contrast to the Byzantine decentralized optimization setup in [5], [6]. Unlike [4], we assume all networked agents process the same copy of data or homogeneous data in the streaming data case. However, some members in the multi-agent networks are under data injection attack. This scenario arises where, for example, the adversary injects malicious data into the training datasets or hijacks the sensors of some agents to mislead local data processing procedures.

In this paper, a network of agents cooperate to solve local copies of the *same* optimization problem with unique optimizer. The agents solve their local problems iteratively (e.g., via gradient descent). A fraction of the agents are subject to data injection attack, which effectively alter the local optimization problems. Due to these attack, some of the agents' local minimizer sequences (i.e., the iterate sequence from the local optimization algorithm) may converge to a different (incorrect) value. We present an approach, based on median consensus, to ensure that all agents, even those under attack, effectively solve the optimization problem. On top of the differences in attack model, to address the adversarial effects, most existing works design screening operators to trim away abnormal data [4]–[6], or solve TV-norm penalized approximation [7], [8], except that [9] integrates a median solver for resilient state estimation, which we believe can be improved with the results in this work and [10].

Our contributions are summarized as follows. We study distributed optimization in undirected inter-agent communication networks with random topology. Each agent is assigned a local optimization problem and computes an iterative sequence that converges to the minimizer of its local problem almost surely (a.s.). To address the considered data injection attack, we propose a *consensus+innovations* type of algorithm with *signed innovations*, that provably converges to the coordinate-wise median of all local minimizers.

The current work can be contrasted with similar works on median consensus [10]–[12]. In particular, this works uses *signed innovations* instead of *clipped innovations* as in [10], which enables new analysis to cope with the non-distinct cases that are not addressed by [10]. In our setup, non-distinctness is critical since all those sequences on unattacked agents still converge to the same true minimizer.

The rest of the paper goes as follows. In section II, we present the problem formulation. In section III, we develop our algorithm and present the main theorem with proofs given in section IV. In section V, we show numerical results on a classification problem.

Notations. Agents exchange messages over a time-varying network denoted as $G_t = (V, E_t)$ at discrete time t. The set of agents $V = [N]$ is indexed from 1 to N. The set of communication links E_t is time-varying. For each agent n, Ω_n^t denotes the set of neighbors at time t. We denote the Laplacian of network G_t as $L_t = D_t - A_t$ where D_t is a diagonal matrix whose i-th diagonal entry is the number of neighbors of agent i and $A_t[i,j] = 1$ if there exists a link between agent i and agent j at time t, otherwise $A_t[i,j] = 0$. A network is connected if and only if the second smallest eigenvalue of its Laplacian is positive [13]. We use $\mathbf{1}_n$ to denote column vector of n ones, and I_n for identity matrix in $\mathbb{R}^{n\times n}$. We use $\|\cdot\|$ for Euclidean norm on vectors and spectral norm on matrices.

II. Problem statement

N networked agents attempt to minimize the same objective function f with a unique minimizer $\mathbf{w}^* \in \mathbb{R}^d$. In the absence

Shuhua Yu and Soummya Kar are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213 USA. Emails: shuhuay@andrew.cmu.edu, soummyak@andrew.cmu.edu.

of attack, agents are able to iteratively compute $\mathbf{w}^*$ which we take as granted. Some of the agents, however, fall under data injection attack, and, as a result, may compute local minimizers that are different than $\mathbf{w}^*$. The agents' collective goal is to simultaneously 1) each compute local minimizer, and 2) track the coordinate-wise median of *all* of the agents' local minimizers.

Assumption 1. *If agent n is under attack, it has a perturbed minimizer $\mathbf{w}_n^* = \mathbf{w}^* + \Delta_n$ with some unknown Δ_n. $\{\mathbf{w}_n^*\}_{n\in[N]}$ contains $\mathbf{w}^*$ and its coordinate-wise median is $\mathbf{w}^*$.*

Remark 1. *Assumption 1 holds true if at least one agent is free of attack and $\{\mathbf{w}_n^*\}_{n\in[N]}$ has coordinate-wise median $\mathbf{w}^*$. A sufficient condition for Assumption 1 is that strictly less than half agents are under attack. An example of the considered data injection attack is as follows: consider a network of agents that each try to learn a binary classifier based on their local (labeled) training data sets. Initially, all of the agents have identical local training data. Then, an attacker flips the labels on some of the examples for less than half of the agents.*

Assumption 2. *Each sequence of local iterates $\{\mathbf{w}_n^t\}_{t\geq 0}, n \in [N]$, converges to local optimizer $\mathbf{w}_n^*$ a.s. and sublinearly, i.e., $\mathbb{P}\big((t+1)^\delta \|\mathbf{w}_n^t - \mathbf{w}_n^*\| = 0\big) = 1$, for positive constant δ.*

Remark 2. *The almost sure convergence of SGD and other gradient methods are well studied in the literature [14], [15].*

Assumption 3. *The Laplacian sequence $\{L_t\}_{t\geq 0}$ associated with $\{G_t\}_{t\geq 0}$ is an i.i.d sequence whose expectation, denoted as $\bar{L}$, exists and satisfies $\lambda_2(\bar{L}) > 0$.*

Remark 3. *This assumption models many practical communication networks. First, static networks (V, E) that are undirected and connected are clearly subsumed. On top of that, it models phenomena such as random link failures, i.e., the links in E have failure probabilities in $[0, 1)$, which captures random failures in practical networks such as wireless sensor networks [16].*

Let $(\Omega, \mathcal{F})$ be the probability space where random variables L_t and $\mathbf{w}_n^t$ are defined, and let $\mathcal{F}_t$ denote the corresponding natural filtration, i.e., $\mathcal{F}_t$ is the sigma algebra $\sigma(\{L_{t'}\}_{t'\leq t}, \{\mathbf{w}_n^{t'}\}_{n\in[N],t'\leq t})$. In this paper, unless otherwise stated, all inequalities involving random variables hold a.s. We use ω to denote a random variable on the sample path $\omega \in \Omega$.

III. Algorithm and Main result

We propose a *consensus+innovations* algorithm with *signed innovations* to address the median tracking problem. Assume that all $\mathbf{w}_n^t$ are given in the same timescale. Each agent n simultaneously maintains a median estimator that updates as

$$\mathbf{x}_n^{t+1} = \mathbf{x}_n^t - \beta_t \sum_{m\in\Omega_n^t} (\mathbf{x}_n^t - \mathbf{x}_m^t) - \alpha_t sign(\mathbf{x}_n^t - \mathbf{w}_n^t), \quad (1)$$

where the *sign* operator is applied on each component of the argument vector, and

$$\alpha_t = \frac{\alpha_0}{(t+1)^{\tau_1}}, \ \beta_t = \frac{\beta_0}{(t+1)^{\tau_2}}$$

for some constants α_0, β_0, and $0 < \tau_2 < \tau_1 < 1$. Denote

$$\mathbf{x}^t = [(\mathbf{x}_1^t)^\top, \ldots, (\mathbf{x}_n^t)^\top]^\top, \ \mathbf{w}^t = [(\mathbf{w}_1^t)^\top, \ldots, (\mathbf{w}_n^t)^\top]^\top.$$

Then, all local updates can be written as

$$\mathbf{x}^{t+1} = (I_{Nd} - \beta_t L_t \otimes I_d)\mathbf{x}^t - \alpha_t sign(\mathbf{x}^t - \mathbf{w}^t). \quad (2)$$

Theorem 1. *Under Assumptions 1-3, local median estimates $\{\mathbf{x}_n^t\}_{t\geq 0}$ of every agent $n \in [N]$ generated by (1) converges to $\mathbf{w}^*$ a.s. in that $\mathbb{P}\big(\lim_{t\to\infty}(t+1)^{\tau_3}\|\mathbf{x}_n^t - \mathbf{w}^*\|_2 = 0\big) = 1$ for every $0 < \tau_3 < \min\{\delta, \tau_1 - \tau_2\}$, and for all n simultaneously.*

Remark 4. *Theorem 1 addresses general median consensus problem that satisfy Assumptions 1-3. We obtain sample wise convergence rate $O(1/t^{\tau_3})$. For example, if local iterates $\{\mathbf{w}_n^t\}$ converges at $O(1/\sqrt{t})$, we can take $\tau_1 = 0.9, \tau_2 = 0.4$. Then, τ_3 can be arbitrarily close to $1/2$.*

IV. Proof of Theorem 1

For the simplicity of presentation, we prove Theorem 1 for the case $d = 1$. Indeed, if each component of the $\mathbf{x}_n^t - \mathbf{w}^*$ converges to 0, then for any finite d, $\|\mathbf{x}_n^t - \mathbf{w}^*\|$ converges to 0 at the same rate. We first show all local median estimates reach consensus. Define $P_N = N^{-1}\mathbf{1}_N\mathbf{1}_N^\top$.

Lemma 1. *Under Assumption 3, for every $0 < \epsilon \leq \tau_1 - \tau_2$, the iterates $\{\mathbf{x}^t\}$ generated by (2) satisfies that $\mathbb{P}\big(\lim_{t\to\infty}(t+1)^{\tau_1-\tau_2-\epsilon}\|\mathbf{x}^t - P_N\mathbf{x}^t\|_2 = 0\big) = 1$.*

Proof. For $d = 1$, (2) reduces to

$$\mathbf{x}^{t+1} = (I_N - \beta_t L_t)\mathbf{x}^t - \alpha_t sign(\mathbf{x}^t - \mathbf{w}^t).$$

Denote $\widehat{\mathbf{x}}^t = \mathbf{x}^t - P_N\mathbf{x}^t$. Multiplying $I_N - P_N$ on both sides of the equation above gives

$$\widehat{\mathbf{x}}^{t+1} = (I_N - P_N - \beta_t L_t)\widehat{\mathbf{x}}^t - \alpha_t(I_N - P_N)sign(\mathbf{x}^t - \mathbf{w}^t).$$

Taking Euclidean norm of $\widehat{\mathbf{x}}^{t+1}$ leads to

$$\|\widehat{\mathbf{x}}^{t+1}\| \leq \|(I_N - P_N - \beta_t L_t)\widehat{\mathbf{x}}^t\| + \alpha_t\sqrt{N}, \quad (3)$$

where we have used

$$\|I_N - P_N\| = 1,$$
$$\|sign(\mathbf{x}^t - \mathbf{w}^t)\| \leq \sqrt{N}.$$

By Lemma 4.4 in [17], there exist a measurable $\mathcal{F}_{t+1}$ adapted and $\mathbb{R}_+$ valued process $\{r_t\}$ and a constant c_r, such that for sufficiently large t,

$$\|(I_N - P_N - \beta_t L_t)\widehat{\mathbf{x}}^t\| \leq (1 - r_t)\|\widehat{\mathbf{x}}^{t+1}\| \quad (4)$$

with

$$\mathbb{E}(r_t \mid \mathcal{F}_t) \leq \frac{c_r}{(t+1)^{\tau_2}}, a.s.$$

By (3)(4) we have

$$\|\widehat{\mathbf{x}}^{t+1}\| \leq (1 - r_t)\|\widehat{\mathbf{x}}^t\| + \alpha_t\sqrt{N}.$$

The relation above falls into the pursuit of Lemma 4.2 in [17], and thus $(t+1)^{\tau_1-\tau_2-\epsilon}\|\widehat{\mathbf{x}}^t\|$ converges to 0 a.s. for every $0 < \epsilon \leq \tau_1 - \tau_2$. □

With Lemma 1 and Assumption 2, we next develop an error bound to estimate $sign(\mathbf{x}^t - \mathbf{w}^t)$. Let x_n^t and w_n^t denote the nth component of $\mathbf{x}^t$ and $\mathbf{w}^t$, respectively. Then, x_n^t, w_n^t are the median estimate and the local optimization iterate of agent n at time t in 1-dimensional case, and the local minimizer $\mathbf{w}_n^*$ reduces to scalar w_n^* and the true minimizer reduces to w^*. Let $\bar{x}^t = N^{-1}\sum_{n=1}^N x_n^t$, and define $e_n^t = x_n^t - \bar{x}^t + w_n^t - w_n^*$.

Lemma 2. *Under Assumption 2 and 3, let*

$$\eta = \tau_1 - \tau_2 - \epsilon_1$$

for arbitrary small $0 < \epsilon_1 < \tau_1 - \tau_2$. Then, there exist positive constants c_η, c_δ such that

$$|e_n^t| \le \frac{c_\eta}{(t+1)^\eta} + \frac{c_\delta}{(t+1)^\delta}, a.s.$$

Proof. We consider the sample path $\omega \in \Omega$ such that there exist $c_{\eta,\omega}, c_{\delta,\omega}$, for all $n \in [N]$, we have

$$|w_{n,\omega}^t - w_n^*| \le \frac{c_{\eta,\omega}}{(t+1)^\eta}, \tag{5}$$

$$|x_{n,\omega}^t - \bar{x}_\omega^t| \le \frac{c_{\delta,\omega}}{(t+1)^\delta}. \tag{6}$$

By triangular inequality,

$$|e_{n,\omega}^t| \le \frac{c_{\eta,\omega}}{(t+1)^\eta} + \frac{c_{\delta,\omega}}{(t+1)^\delta}. \tag{7}$$

By Assumption 2 and Lemma 1, the set of all such sample paths ω has probability 1, and thus the lemma follows. □

We next show that for sufficiently large t, there exists a local contraction region for $\bar{x}^t$. The following lemma characterizes a threshold sequence for the local contraction.

Lemma 3. *Define the threshold sequence*

$$\gamma_t = \frac{\gamma_0}{(t+1)^{\tau_3}},$$

for some constant γ_0 and choose τ_3 such that

$$0 < \tau_3 < \min\{\delta, \tau_1 - \tau_2\}.$$

Then, for any constants c_δ and c_η, it holds that for sufficiently large t,

$$\gamma_t - \frac{\alpha_t}{N} \le \gamma_{t+1}, \tag{8}$$

$$\frac{c_\eta}{(t+1)^\eta} + \frac{c_\delta}{(t+1)^\delta} + \alpha_t \le \gamma_{t+1}. \tag{9}$$

Proof. By mean value theorem, there exists $t' \in (t, t+1)$ such that

$$\gamma_t - \gamma_{t+1} = \frac{\gamma_0\tau_3}{(t'+1)^{1+\tau_3}} < \frac{\gamma_0\tau_3}{(t+1)^{1+\tau_3}}.$$

Since $1 + \tau_3 > 1 > \tau_1$, for sufficiently large t we have

$$\frac{\gamma_0\tau_3}{(t+1)^{1+\tau_3}} \le \frac{\alpha_0}{N(t+1)^{\tau_1}},$$

and thus (8) follows. Next, for sufficiently large t, we have

$$\gamma_{t+1} = \frac{\gamma_0}{(t+2)^{\tau_3}} = \frac{\gamma_0}{(t+1)^{\tau_3}} \frac{1}{\left(1 + \frac{1}{t+1}\right)^{\tau_3}} \ge \frac{\gamma_t}{2},$$

which follows from the fact that for sufficiently large t,

$$\left(1 + \frac{1}{t+1}\right)^{\tau_3} \le 2.$$

Thus, to ensure (9), it suffices to ensure that

$$\frac{c_\eta}{(t+1)^\eta} + \frac{c_\delta}{(t+1)^\delta} + \frac{\alpha_0}{(t+1)^{\tau_1}} \le \frac{\gamma_0}{2(t+1)^{\tau_3}},$$

which holds true for suffciently large t due to $\tau_3 < \min\{\tau_1 - \tau_2, \delta\}$. Therefore, the lemma is proved. □

We next show the existence of the local contraction region for sufficiently large t.

Lemma 4. *Define γ_t as in Lemma 3. Then, almost surely, there exists a finite T_0 such that if for some $T_1 \ge T_0, |\bar{x}^{T_1} - w^*| \le \gamma_{T_1}$, then $|\bar{x}^t - w^*| \le \gamma_t$ for all $t \ge T_1$.*

Proof. We consider the sample path $\omega \in \Omega$ on which Lemma 2 holds, i.e., (5)(6) are satisfied. Multiplying $N^{-1}\mathbf{1}^\top$ on both sides of (2) leads to

$$\bar{x}_\omega^{t+1} = \bar{x}_\omega^t - \frac{\alpha_t}{N}\sum_{n=1}^N sign(x_{n,\omega}^t - w_{n,\omega}^t). \tag{10}$$

Define

$$d_{\min} = \min_{w_n^* \ne w^*} |w_n^* - w^*|,$$

i.e., the least perturbation of local minimizers. If $d_{\min} > 0$, define t_1 as the least t such that $\gamma_t < d_{\min}$. Then take t_2 as the least $t \ge t_1$ such that

$$\frac{\alpha_0}{(t+1)^{\tau_1}} < \frac{c_{\eta,\omega}}{(t+1)^\eta} + \frac{c_{\delta,\omega}}{(t+1)^\delta} < \frac{\gamma_0}{(t+1)^{\tau_3}}, \tag{11}$$

where by $\tau_3 < \min\{\delta, \tau_1 - \tau_2\}$ and $\eta < \tau_1 - \tau_2$, such t_2 must exist. Without loss of generality, we consider

$$\bar{x}_\omega^t \ge w^*. \tag{12}$$

We next consider two excluding cases for $\bar{x}^t$ that satisfies the hypothesis of the lemma.

Case 1. If

$$\frac{c_{\eta,\omega}}{(t+1)^\eta} + \frac{c_{\delta,\omega}}{(t+1)^\delta} < \bar{x}_\omega^t - w^* \le \gamma_t. \tag{13}$$

Then,

$$x_{n,\omega}^t - w_{n,\omega}^t = x_{n,\omega}^t - \bar{x}_\omega^t + \bar{x}_\omega^t - w_n^* + w_n^* - w_{n,\omega}^t. \tag{14}$$

For $n \in [N]$ such that $w_n^* = w^*$, by (7), (13)(14) we have

$$sign(x_{n,\omega}^t - w_{n,\omega}^t) = sign(\bar{x}_\omega^t - w^*). \tag{15}$$

For $m \in [N]$ such that $w_m^* \ne w^*$, we have

$$\begin{aligned} &x_{m,\omega}^t - w_{m,\omega}^t \\ &= x_{m,\omega}^t - \bar{x}_\omega^t + \bar{x}_\omega^t - w^* + w^* - w_m^* + w_m^* - w_{m,\omega}^t. \end{aligned}$$

If such m exists, take t_3 as the least $t \ge t_2$ such that

$$\frac{c_{\eta,\omega}}{(t+1)^\eta} + \frac{c_{\delta,\omega}}{(t+1)^\delta} + \gamma_t < d_{\min} \le |w^* - w_m^*|. \tag{16}$$

Then, by (13)(16), we have

$$sign(x^t_{m,\omega} - w^t_{m,\omega}) = sign(w^* - w^*_m). \tag{17}$$

By Assumption 1, w^* is the unique median of $\{w^*_n\}$, combined with (12)(13)(15)(17), we have

$$1 \le \sum_{n=1}^{N} sign(x^t_{n,\omega} - w^t_{n,\omega}) \le N.$$

Putting the display above into (10) gives that

$$\bar{x}^t_\omega - w^* - \alpha_t \le \bar{x}^{t+1}_\omega - w^* \le \bar{x}^t_\omega - w^* - \frac{\alpha_t}{N}. \tag{18}$$

Take t_4 as the least $t \ge t_3$ such that (8)(9) in Lemma 3 hold. By the first part of (11) and (12)(13), we have $\bar{x}^{t+1}_\omega - w^* > 0$. Then, with (8) we have $|\bar{x}^{t+1}_\omega - w^*| \le \gamma_{t+1}$.

Case 2. If

$$0 \le \bar{x}^t - w^* \le \frac{c_{\eta,\omega}}{(t+1)^\eta} + \frac{c_{\delta,\omega}}{(t+1)^\delta}.$$

In this case, upper bounds (5)(6) cannot determine the relationships between $sign(x^t_{n,\omega} - w^t_{n,\omega})$ and $sign(\bar{x}^t_\omega - w^*)$ for $n \in [N]$ with $w^*_n = w^*$ as in (15), but with the same argument, (17) still holds true. Then, (10) leads to

$$-\alpha_t \le \bar{x}^{t+1}_\omega - w^* \le \frac{c_{\eta,\omega}}{(t+1)^\eta} + \frac{c_{\delta,\omega}}{(t+1)^\delta} + \alpha_t.$$

Then, from (9) we obtain $|\bar{x}^{t+1}_\omega - w^*| \le \gamma_{t+1}$. Combing those two cases and taking $T_{0,\omega} = t_4$, since the set of all such ω has probability 1, the existence of T_0 has probability 1. □

We next show that $\bar{x}^t$ eventually falls into the local contraction region identified in Lemma 4.

Lemma 5. *For T_0, γ_t defined in Lemma 4, there must exist a finite $T_1 \ge T_0$ such that $|\bar{x}^{T_1} - w^*| \le \gamma_{T_1}$.*

Proof. We prove this lemma by contradiction. Consider the sample path $\omega \in \Omega$ such that Lemma 2 holds. Suppose for all $t \ge T_{0,\omega}$, we have $|\bar{x}^t_\omega - w^*| > \gamma_t$. Without loss of generality, we assume $\bar{x}^t_\omega > w^*$. By the same analysis that follows from (13), relation (15) still holds true for all n with $w^*_n = w^*$, and (17) holds true for all m such that $w^*_m < w^*$. Since w^* is the unique median for $\{w^*_n\}_{n\in[N]}$ and $w^* \in \{w^*_n\}_{n\in[N]}$, we have

$$\sum_{w^*_n \le w^*} sign(x^t_{n,\omega} - w^t_{n,\omega}) \ge \begin{cases} \frac{N+1}{2}, & N \text{ is odd}, \\ \frac{N}{2} + 2, & N \text{ is even}. \end{cases}$$

It follows that

$$1 \le \sum_{n=1}^{N} sign(x^t_{n,\omega} - w^t_{n,\omega}) \le N.$$

Putting the display above into (10), by (11) and the contradiction hypothesis, we have

$$|\bar{x}^{t+1}_\omega - w^*| \le |\bar{x}^t_\omega - w^*| - \frac{\alpha_t}{N}.$$

Then, summing over t from $T_{0,\omega}$ to infinity we have the contradiction

$$\lim_{t\to\infty} |\bar{x}^t_\omega - w^*| \le |\bar{x}^{T_0}_\omega - w^*| - \frac{1}{N}\sum_{t=T_{0,\omega}}^{\infty} \alpha_t = -\infty,$$

where the last equality follows from $0 < \tau_1 < 1$. Therefore, the hypothesis fails to hold and the lemma follows. □

Proof of Theorem 1. By Lemma 4-5, almost surely, there exists some finite T_1 such that $\forall t \ge T_1, |\bar{x}^t - w^*| \le \gamma_t$. Thus,

$$\mathbb{P}\big(\lim_{t\to\infty} (t+1)^{\tau_3} |\bar{x}^t - w^*| = 0\big) = 1. \tag{19}$$

By triangular inequality, we have

$$|x^t_n - w^*| \le |x^t_n - \bar{x}^t| + |\bar{x}^t - w^*|. \tag{20}$$

Combing (19), (20), Lemma 1, and $\tau_3 < \tau_1 - \tau_2$, we have for all n simultaneously,

$$\mathbb{P}\big(\lim_{t\to\infty} (t+1)^{\tau_3} |x^t_n - w^*| = 0\big) = 1.$$

□

V. Numerical Experiments

We consider a binary classification task on Fashion-mnist dataset [18]. Consider the scenario where a network (see Fig. 1 for the simulated network) of 15 agents train logistic regression models to classify two classes "pullover" versus "coat" where each class initially has the same 6k training and 1k test data points.. The trainning data on 5 red agents in Fig.

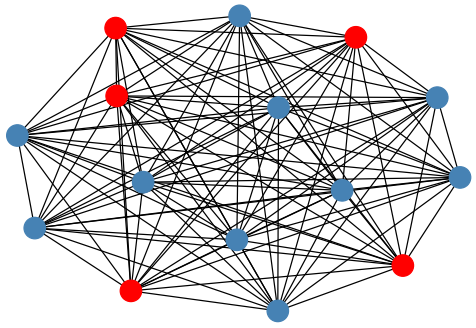

Fig. 1. Network of 15 agents, 5 red agents have injected data

1 are corrupted as follows: the labels of ρ-fraction randomly sampled data points from each class are flipped, i.e. changed from 0 to 1 or from 1 to 0. Agents may communicate over the time-varying network depicted in Fig. 1, where, in each time step, every communication link (graph edge) may fail with probability 0.2.

Each agent n trains a logistic regression model in parallel on local dataset and updates its weights $\mathbf{w}^t_n$ by vanilla SGD with mini-batch 200 and stepsize 0.1. Since data on red agents are corrupted, their weights converge to some points that performs poorly on test data. In each step t, each agent maintains an estimate $\mathbf{x}^t_n$ for the median of all the local minimizers as specified in (1). We compare the average test accuracies of

$\mathbf{w}_n^t$ and $\mathbf{x}_n^t$ on attacked agents and unattacked agents in Fig. 2, and that on attacked agents with different ρ.

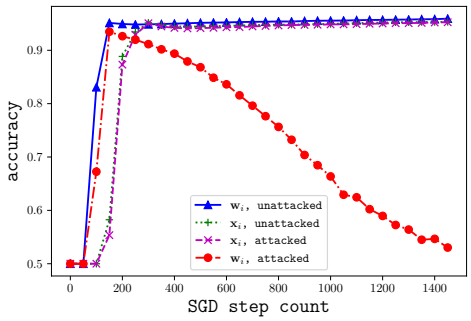

Fig. 2. Average test accuracy on attacked/unattacked agents with $\rho = 0.5$.

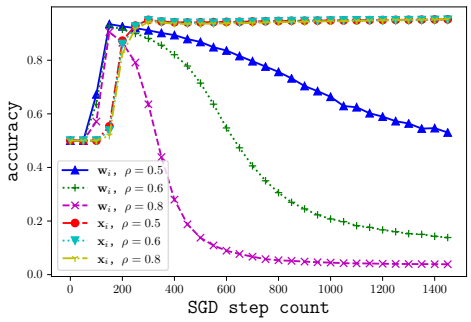

Fig. 3. Average test accuracy on attacked agents with different attack.

The performance of local classifiers $\mathbf{w}_n^t$ on attacked agents deteriorate with increasing iterations. On the other hand, the median estimators $\mathbf{x}_n^t$ for all local optimizers $\mathbf{w}_n^*$ remains a robust estimator for model minimizer, and achieves the same test accuracies on unattacked agents. Without the median estimator $\mathbf{x}_n^t$, agent can only rely on local optimizer $\mathbf{w}_n^t$ that is arbitrarily bad under data injection attack with big portion ρ, as shown in Fig. 3. The code can be found here[1].

VI. Conclusion

In this paper, we have studied multi-agent optimization under data attack over random networks. The agents cooperate to *the same* optimization problem, and the agents each have local "copies" of the problem that are initially *identical* . The agents follow an iterative procedure to find the minimizer to their local problem. A fraction of the agents falls under attack, and, as a result, in the absence of cooperation, these agents may fail to find the correct minimizer. We presented method based on distributed median consensus for all of the agents, even those under attack, to resiliently identify the correct minimizer to the original optimization problem. Finally, we presented a numerical example that demonstrated our approach on a classification task where a fraction of the agents train on compromised data. Our method ensures that, even when agents train with incorrect labels, through cooperation with other agents, they are able to learn the correct classifier. Future directions may include integrating the median consensus process into decentralized optimization where agents optimize heterogeneous functions.

[1]https://colab.research.google.com/drive/1-qV3keKhIGdiyxMC3uSEotTTXfVSqrH-?usp=sharing

VII. Acknowledgement

This research was partially supported by NSF under grant CNS-1837607.

References

[1] Y. Chen, S. Kar, and J. M. F. Moura, "The internet of things: Secure distributed inference," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 64–75, 2018.

[2] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[3] Z. Yang, A. Gang, and W. U. Bajwa, "Adversary-resilient distributed and decentralized statistical inference and machine learning: An overview of recent advances under the byzantine threat model," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 146–159, 2020.

[4] Y. Chen, S. Kar, and J. M. F. Moura, "Resilient distributed parameter estimation with heterogeneous data," *IEEE Transactions on Signal Processing*, vol. 67, no. 19, pp. 4918–4933, 2019.

[5] L. Su and N. H. Vaidya, "Byzantine-resilient multiagent optimization," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2227–2233, 2020.

[6] Z. Yang and W. U. Bajwa, "Byrdie: Byzantine-resilient distributed coordinate descent for decentralized learning," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 4, pp. 611–627, 2019.

[7] W. Ben-Ameur, P. Bianchi, and J. Jakubowicz, "Robust distributed consensus using total variation," *IEEE Transactions on Automatic Control*, vol. 61, no. 6, pp. 1550–1564, 2015.

[8] J. Peng, W. Li, and Q. Ling, "Byzantine-robust decentralized stochastic optimization over static and time-varying networks," *Signal Processing*, vol. 183, p. 108020, 2021.

[9] J. G. Lee, J. Kim, and H. Shim, "Fully distributed resilient state estimation based on distributed median solver," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3935–3942, 2020.

[10] S. Yu, Y. Chen, and S. Kar, "Dynamic median consensus over random networks," *arXiv preprint arXiv:2110.05317*, 2021.

[11] A. Pilloni, A. Pisano, M. Franceschelli, and E. Usai, "Robust distributed consensus on the median value for networks of heterogeneously perturbed agents," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 6952–6957, IEEE, 2016.

[12] Z. A. Z. Sanai Dashti, C. Seatzu, and M. Franceschelli, "Dynamic consensus on the median value in open multi-agent systems," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 3691–3697, IEEE, 2019.

[13] B. Mohar, Y. Alavi, G. Chartrand, and O. Oellermann, "The laplacian spectrum of graphs," *Graph theory, combinatorics, and applications*, vol. 2, no. 871-898, p. 12, 1991.

[14] D. P. Bertsekas and J. N. Tsitsiklis, "Gradient convergence in gradient methods with errors," *SIAM Journal on Optimization*, vol. 10, no. 3, pp. 627–642, 2000.

[15] O. Sebbouh, R. M. Gower, and A. Defazio, "Almost sure convergence rates for stochastic gradient descent and stochastic heavy ball," in *Conference on Learning Theory*, pp. 3935–3971, PMLR, 2021.

[16] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.

[17] S. Kar, J. M. F. Moura, and H. V. Poor, "Distributed linear parameter estimation: Asymptotically efficient adaptive strategies," *SIAM Journal on Control and Optimization*, vol. 51, no. 3, pp. 2200–2229, 2013.

[18] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.