Resilience to Malicious Activity in Distributed Optimization for Cyberphysical Systems

Michal Yemini[1]

Joint work with Angelia Nedić[2], Stephanie Gil[3] and Andrea Goldsmith[4]

[2]Arizona State University, [3]Harvard University, [4]Princeton University

[1]Bar-Ilan University

CoE, February 2023

Outline

1. Distributed Optimization Systems.
2. Malicious Agents in Distributed Optimization Systems.
3. Agents' Trust Values in Cyberphysical Systems.
4. Characterizing Trust-Based Resilience in Distributed Optimization Systems.
5. Numerical Results.
6. Conclusions.

Distributed Optimization

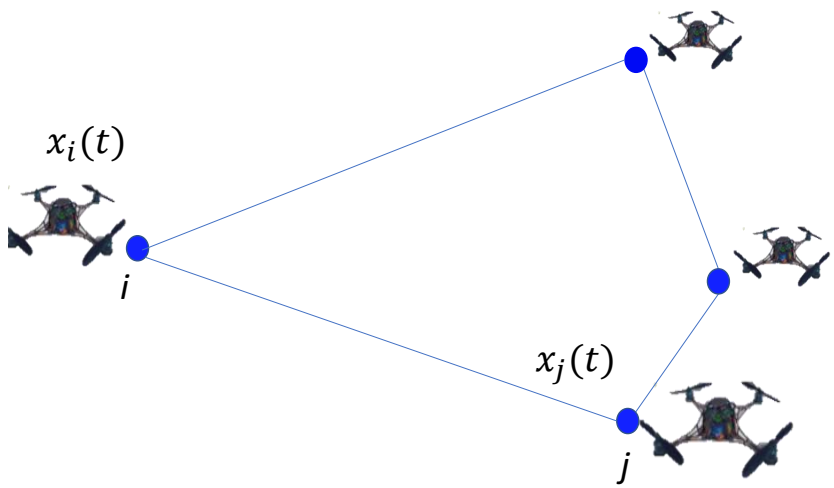
Leaderless optimization and control for multi-agent systems.

- Machine learning (training a shared model with local datasets).
- Robotic and drone networks (rendezvous problem).
- Sensor networks (data fusion).

Distributed optimization:

$$x^\star = \arg\min_{x \in \mathcal{X}} f(x), \text{ with}$$

$$f(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x).$$



Assumption

We assume that $\mathcal{X} \subset \mathbb{R}^d$ is compact and convex and that there exists a known value $\eta > 0$ such that

$$\|x\| \leq \eta, \quad \forall x \in \mathcal{X}. \tag{1}$$

Assumption

The functions f_i are μ-strongly convex and have$L$-Lipschitz continuous gradients, i.e.,
$\|\nabla f_i(x) - \nabla f_i(y)\| \leq L\|x - y\|$, *for all* $x \in \mathbb{R}^d$.

It follows that there is a scalar G such that:

$$\|\nabla f_i(x)\| \leq G, \; \forall x \in \mathcal{X}, i \in \mathcal{L}.$$

Solving Distributed Optimization Problems

Let us consider **connected** graph $G = (\mathbb{V}, \mathbb{E})$, a **stochastic weight** matrix W and **initial vector** values $x(0)$. Then, the distributed optimization problem can be solved using the following dynamic:

$$\begin{aligned} c_i(t) &= w_{ii}(t)x_i(t) + \sum_{j \in \mathcal{N}_i} w_{ij}(t)x_j(t), \\ y_i(t) &= c_i(t) - \gamma(t)\nabla f_i(c_i(t)), \\ x_i(t+1) &= \Pi_{\mathcal{X}}\left(y_i(t)\right), \end{aligned} \tag{2}$$

where $\mathcal{N}_i = \{j \in \mathbb{V} \mid \{i, j\} \in \mathbb{E}\}$, and $\gamma(t) \geq 0$ such that:

$$\sum_{t=0}^{\infty} \gamma(t) = \infty \text{ and } \sum_{t=0}^{\infty} \gamma^2(t) < \infty.$$

For the time being, let us assume that $W(t) = \overline{W}$ is a fixed doubly stochastic matrix and denote by $\rho_{\mathcal{L}} < 1$ its second largest eigenvalue modulus. Additionally, denote

$$\overline{h}(T) \triangleq \frac{G^2 T}{\mu} + \frac{2G^2 T}{\mu(1-\rho_{\mathcal{L}})} + \frac{8(\mu+L)G^2}{\mu^2(1-\rho_{\mathcal{L}})^2} \ln\left(\frac{T+2}{2}\right) + \frac{2\eta G}{1-\rho_{\mathcal{L}}} + \frac{2(\mu+L)(\mu\eta+2G)^2}{\mu^2(1-\rho_{\mathcal{L}})^2} + \frac{2G^2 + 4G\eta(\mu+L)}{\mu(1-\rho_{\mathcal{L}})^3} + \frac{G^2(\mu+L)}{\mu^2(1-\rho_{\mathcal{L}})^4}.$$

If all the agents are truthful, the dynamic (2) converges to the optimal point for every initial point $x_i(0) \in \mathcal{X}$, $i \in \mathbb{V}$:

$$\lim_{t\to\infty} \|x_i(t) - x^\star\| = 0, \; \forall\, i \in \mathcal{L}.$$

Moreover, if $\gamma(t) = \frac{2}{\mu(t+1)}$, then

$$\frac{1}{n}\sum_{i\in\mathcal{L}} \|x_i(T) - x^\star\|^2 \leq \min\left\{4\eta^2, \frac{4\overline{h}(T)}{\mu T(T+1)}\right\}, \tag{3}$$

for any initial points $x_i(0) \in \mathcal{X}$, $i \in \mathbb{V}$, and any $T \geq 1$.

Distributed Optimization with Malicious Agents

In practice, some agents can be malicious and input values to take the dynamic (2) away from its optimal value.
In this case, $\mathbb{V} = \mathcal{L} \cup \mathcal{M}$ where $\mathcal{L}$ is the set of legitimate agents and $\mathcal{M}$ is the set of malicious agents.

We are interested in solving the following problem without knowing the identities of the legitimate and malicious agents in advance:

$$x^\star_{\mathcal{L}} = \arg\min_{x \in \mathcal{X}} f(x), \text{ with } f(x) = \frac{1}{|\mathcal{L}|} \sum_{i \in \mathcal{L}} f_i(x). \tag{4}$$

Sundaram and Gharesifar, 2019 show that even a single malicious agent can prevent the naïve implementation of the dynamic (2) from converging to $x^\star_{\mathcal{L}}$.

Agents' Trust Values in Cyberphysical Systems I

Additional works have studied the number of malicious agents that can be mitigated. However, the guaranteed number is small and can be smaller than half of the network connectivity.

Prior works have used the **data values** to overcome/detect malicious behavior. The **physical** aspects of the problem have not been considered. A particular example is the **wireless communication channels**.

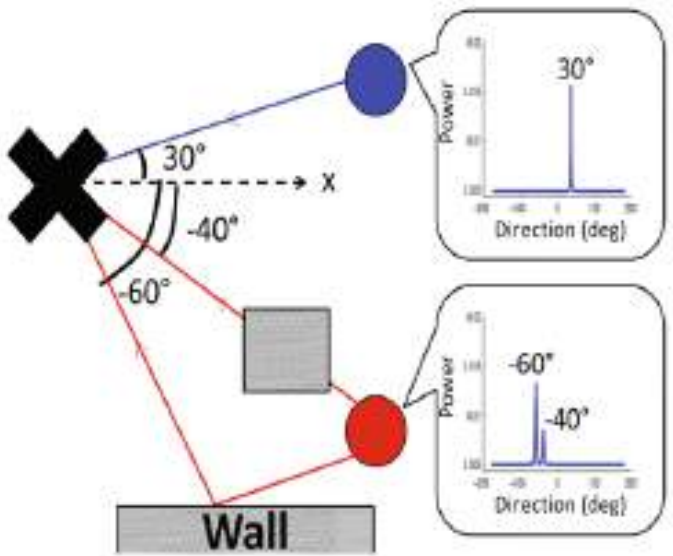
In cyberphysical systems:

- Malicious agents can lie about their location.
- A malicious agent can create many fictitious identities (Sybil attack).

Agents' Trust Values in Cyberphysical Systems II

Each transmitted signal leads to received signal characteristics:

- Number of paths, delays.
- Angles of arrival.
- Power order of the angles of arrival.
- Power of the received signals.



*Guaranteeing spoof-resilient multi-robot networks, S. Gil *et al* 2017.

Agents' Trust Values in Cyberphysical Systems III

We can generate trust values that capture the event that an agent

- lies about its location
  - Location Verification Systems for VANETs in Rician Fading Channels, S. Yan *et al* 2016.
- uses a Sybil attack and creates multiple fictitious agents
  - Detecting Colluding Sybil Attackers in Robotic Networks using Backscatters Y. Huang *et al* 2021.

    (Limited to single antenna malicious agents.)
  - Guaranteeing spoof-resilient multi-robot networks, S. Gil *et al* 2017.

    (Limited to single antenna malicious agents.)
  - The Mason Test: A Defense Against Sybil Attacks in Wireless Networks Without Trusted Authorities, Liu *et al* 2015.

    (Assumes limited mobility of malicious agents and no beamforming).

We denote by $\alpha_{ij}(t) \in [0, 1]$ the instantaneous single sample trust agent i gives agent j at time a t.

Research Objectives I

Recall the dynamic (2):

$$
\begin{aligned}
c_i(t) &= w_{ii}(t)x_i(t) + \sum_{j\in\mathcal{N}_i} w_{ij}(t)x_j(t), \\
y_i(t) &= c_i(t) - \gamma(t)\nabla f_i(c_i(t)), \\
x_i(t+1) &= \Pi_{\mathcal{X}}\left(y_i(t)\right).
\end{aligned}
$$

Objective 1

We wish to construct weight sequences $\{w_{ij}(t)\}$, $i \in \mathcal{L},\ j \in \mathcal{N}_i$ in the method (2) such that they converge over time to some *nominal weights* $\overline{w}_{ij}, i \in \mathcal{L},\ j \in \mathcal{N}_i$, almost surely (a.s.), where $\overline{w}_{ij} = 0$ for all malicious neighbors $j \in \mathcal{N}_i \cap \mathcal{M}$ of agent $i \in \mathcal{L}$.

Research Objectives II

Objective 2

Utilizing the proposed weights $\{w_{ij}(t)\}_{t=1,\ldots}$, we aim to show that the iterates given by (2) converge (in some sense) to the true optimal point $x^\star_{\mathcal{L}} \in \mathcal{X}$.

Objective 3

We aim to establish an upper bound on the expected value of $\|x_i(t) - x^\star_{\mathcal{L}}\|^2$, for all $i \in \mathcal{L}$, as a function of the time t, for the iterates $x_i(t)$ produced by the method.

Cumulative Trust Values

We assume that:

- $\alpha_{ij}(t)$ are statistically independent.
- There exist scalars $E_{\mathcal{L}} > 0$ and $E_{\mathcal{M}} < 0$ such that

$$E_{\mathcal{L}} \triangleq E(\alpha_{ij}(t)) - 1/2 \qquad \text{for all } i \in \mathcal{L},\ j \in \mathcal{N}_i \cap \mathcal{L},$$

$$E_{\mathcal{M}} \triangleq E(\alpha_{ij}(t)) - 1/2 \qquad \text{for all } i \in \mathcal{L},\ j \in \mathcal{N}_i \cap \mathcal{M}.$$

To capture the **history** of observations $\alpha_{ij}(t)$, we define:

$$\beta_{ij}(t) \triangleq \sum_{k=0}^{t-1} (\alpha_{ij}(k) - 1/2) \ \text{ for } t \geq 1, i \in \mathcal{L}, j \in \mathcal{N}_i,$$

with $\beta_{ij}(0) = 0$.
Agent i classifies agent j as legitimate if $\beta_{ij}(t) \geq 0$ and malicious otherwise.

Motivation: Finite Correct Classification Time

Lemma

For every $t \geq 0$ and $i \in \mathcal{L}$

$$\Pr\left(\beta_{ij}(t) < 0\right) \leq \exp(-2tE_{\mathcal{L}}^2), j \in \mathcal{N}_i \cap \mathcal{L},$$
$$\Pr\left(\beta_{ij}(t) \geq 0\right) \leq \exp(-2tE_{\mathcal{M}}^2), j \in \mathcal{N}_i \cap \mathcal{M}.$$

This is an immediate result of the Chernoff-Hoeffding Inequality.

Proposition

There exists a (random) finite time instant $T_f > 0$ such that every legitimate agent i correctly classifies its neighbors for all $t \geq T_f$ almost surely.

This proposition follows by the Borel-Cantelli Lemma

Trust-Based Weights I

Define the time dependent **trusted neighborhood** for agent $i \in \mathcal{L}$:

$$\mathcal{N}_i(t) \triangleq \{j \in \mathcal{N}_i : \beta_{ij}(t) \geq 0\}. \tag{5}$$

For all $t \geq 0$, let

$$d_i(t) \triangleq |\mathcal{N}_i(t)| + 1 \geq 1 \qquad \text{for all } i \in \mathcal{L}.$$

We define the weights $w_{ij}(t)$ as follows: for every $i \in \mathcal{L}$, $j \in \mathcal{N}_i$,

$$w_{ij}(t) = \begin{cases} \frac{\mathbb{1}_{\{t \geq T_0\}}}{2 \cdot \max\{d_i(t), d_j(t)\}} & \text{if } j \in \mathcal{N}_i(t), \\ 0 & \text{if } j \notin \mathcal{N}_i(t) \cup \{i\}, \\ 1 - \sum_{m \in \mathcal{N}_i} w_{im}(t) & \text{if } j = i. \end{cases} \tag{6}$$

where T_0 is the number of trust observations collected before the legitimate agents trust any of their neighbors.

Trust-Based Weights II

By the Borel-Cantelli lemma $w_{ij}(t)$ converges a.s. to the matrix $\overline{w}_{ij}$, where

$$\overline{w}_{ij} = \begin{cases} \frac{1}{2 \cdot \max\{d_i, d_j\}} & \text{if } j \in \mathcal{N}_i \cap, \\ 0 & \text{if } j \notin \mathcal{N}_i \cup \{i\}, \\ 1 - \sum_{m \in \mathcal{N}_i} w_{im} & \text{if } j = i, \end{cases} \tag{7}$$

$d_i = |\mathcal{N}_i| + 1$.
This is a special matrix that is doubly stochastic for the legitimate agents and ignores the malicious agent's inputs.

Asymptotic Convergence to the Optimal Nominal Point

Theorem (Convergence a.s. to the optimal point)

The sequence $\{x_i(t)\}$ *converges a.s. to* $x^\star_{\mathcal{L}}$ *for every* $i \in \mathcal{L}$ *and* $T_0 \geq 0$.

Theorem (Convergence in mean to the optimal point)

For every $T_0 \geq 0$, *the sequence* $\{x_i(t)\}$ *converges in the* r*-th mean to* $x^\star_{\mathcal{L}}$ *for every* $i \in \mathcal{L}$ *and* $r \geq 1$, *i.e.,*

$$\lim_{t\to\infty} \mathbf{E}\left[\|x_i(t) - x^\star_{\mathcal{L}}\|^r\right] = 0, \text{ for all } r \geq 1.$$

The Expected Convergence Rate via the Correct Classification Time I

Consequently we can upper bound the convergence rate as follows:

Theorem

For every $t \geq T_0$ we have that the expected error is bounded by a decaying function such that:

$$\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}\mathbf{E}[\|x_i(t)-x^\star\|^2]$$
$$\leq \min_{m\in[T_0:t-1]}\left\{\min\left\{4\eta^2, \frac{4\overline{h}\,(t-m)}{\mu(t-m)(t-m+1)}\right\}+4\eta^2 p_e(m)\right\}.$$

The Expected Convergence Rate via the Correct Classification Time II

$$\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}\mathbf{E}[\|x_i(t)-x^\star\|^2]\leq \min\left\{4\eta^2,\ \frac{4\overline{h}\,(t-T_0)}{\mu(t-T_0)(t-T_0+1)}\right\}+4\eta^2 p_e\,(T_0)\,, \tag{8}$$

$$\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}\mathbf{E}[\|x_i(t)-x^\star\|^2]\leq \min\left\{4\eta^2,\ \frac{16\overline{h}\left(\frac{t-T_0}{2}\right)}{\mu(t-T_0)(t-T_0+2)}\right\}+4\eta^2 p_e\left(\frac{t+T_0}{2}-1\right), \tag{9}$$

$$\begin{aligned}\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}}\mathbf{E}[\|x_i(t)-x^\star\|^2]\leq \min\left\{4\eta^2,\ \frac{4\overline{h}\left(t-\lceil\frac{\ln(t)}{2\min\{E_{\mathcal{L}}^2,E_{\mathcal{M}}^2\}}\rceil\right)}{\mu\left(t-\lceil\frac{\ln(t)}{2\min\{E_{\mathcal{L}}^2,E_{\mathcal{M}}^2\}}\rceil\right)\left(t-\lceil\frac{\ln(t)}{2\min\{E_{\mathcal{L}}^2,E_{\mathcal{M}}^2\}}\rceil+1\right)}\right\}\\ +4\eta^2\cdot\frac{D_{\mathcal{L}}+D_{\mathcal{M}}}{T}.\end{aligned} \tag{10}$$

Expected Deviation from Mean Value

Denote,

$$D_{\mathcal{L}} \triangleq \sum_{i \in \mathcal{L}} |\mathcal{N}_i \cap \mathcal{L}| \quad \text{and} \quad D_{\mathcal{M}} \triangleq \sum_{i \in \mathcal{L}} |\mathcal{N}_i \cap \mathcal{M}|.$$

Additionally, denote the following upper bound of a misclassification error:

$$p_c(k) \triangleq \mathbb{1}_{\{k \geq 0\}} \left[D_{\mathcal{L}} e^{-2kE_{\mathcal{L}}^2} + D_{\mathcal{M}} e^{-2kE_{\mathcal{M}}^2} \right].$$

Recall that $\rho_{\mathcal{L}} < 1$ is the second largest eigenvalue modulus of the doubly-stochastic nominal weight matrix $\overline{W}$.

Denote, $\overline{x}_{\mathcal{L}}(t) \triangleq \frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} x_i(t)$, and

$$\delta_{\mathcal{M}}(t,T_0) \triangleq 2\eta\rho_{\mathcal{L}}^{t-T_0} + \frac{(2\eta\sqrt{p_c(T_0)} + G\gamma(0))\rho_{\mathcal{L}}^{(t-T_0)/2}}{1-\rho_{\mathcal{L}}} + \frac{2(\eta\sqrt{p_c((t+T_0)/2)} + G\gamma((t-T_0)/2))}{1-\rho_{\mathcal{L}}}. \quad (11)$$

Lemma

For every $t \geq 0$

$$\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} \mathbf{E}\|x_i(t) - \overline{x}_{\mathcal{L}}(t)\| \leq \delta_{\mathcal{M}}(t,T_0), \text{ and}$$

$$\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} \mathbf{E}\|x_i(t) - \overline{x}_{\mathcal{L}}(t)\|^2 \leq \delta_{\mathcal{M}}^2(t,T_0).$$

Tightening the Convergence Results

Theorem

For every collection $x_i(0) \in \mathcal{X}$, $i \in \mathcal{L}$,*of initial points i.e.,*

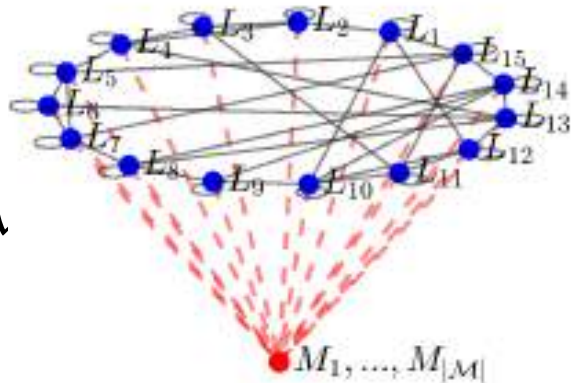
$$\lim_{t\to\infty} \mathbf{E}\left[\|x_i(t) - x^\star_{\mathcal{L}}\|^2\right] = 0, \ \forall\, i \in \mathcal{L}. \tag{12}$$

Moreover, let $\gamma(t) = \frac{2}{\mu(t+2)}$. *Then, for every* $T_0 \geq 0$ *and* $T \geq T_0$ *there exists a function* $C_{\mathcal{M}}(T_0)$ *that decreases exponentially with* T_0 *and is independent of* T *such that for any collection* $x_i(0) \in \mathcal{X}$, $i \in \mathcal{L}$, *and for all* $T \geq T_0$,

$$\frac{1}{|\mathcal{L}|}\sum_{i\in\mathcal{L}} \mathbf{E}\left[\|x_i(T) - x^\star_{\mathcal{L}}\|^2\right] \leq \min\left\{4\eta^2, \frac{4\overline{h}(T-T_0) + C_{\mathcal{M}}(T_0)}{\mu(T-T_0)(T-T_0+1)}\right\}. \tag{13}$$

Numerical Results I

- $|\mathcal{L}| = 15$ legitimate agents, $|\mathcal{M}| \in \{15, 30\}$;
- $d = 1$ and $\eta = 50$;
- $E(\alpha_{ij}) = E_{\mathcal{L}} = 0.55$ for $i \in \mathcal{L},\ j \in \mathcal{N}_i \cap \mathcal{L}$,
- $E(\alpha_{ij}) = E_{\mathcal{M}} = 0.45$ for $i \in \mathcal{L},\ j \in \mathcal{N}_i \cap \Lambda$
- $\alpha_{ij} \sim U\left[E(\alpha_{ij}) - \frac{\ell}{2}, E(\alpha_{ij}) + \frac{\ell}{2}\right]$, where $\ell = 0.6, 0.8$.



The legitimate agents aim to minimize the function

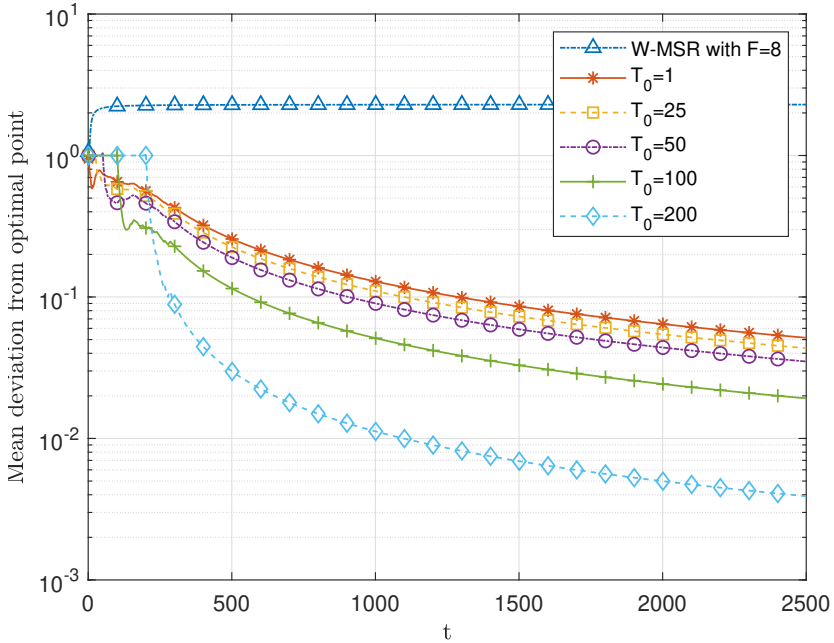
$$\arg \min_{x \in [-\eta, \eta]} \left\{ \frac{1}{|\mathcal{L}|} \sum_{i \in \mathcal{L}} (x - u_i)^2 \right\} \approx 31.4,$$

where u_i were chosen from the interval $[-200, 200]$.

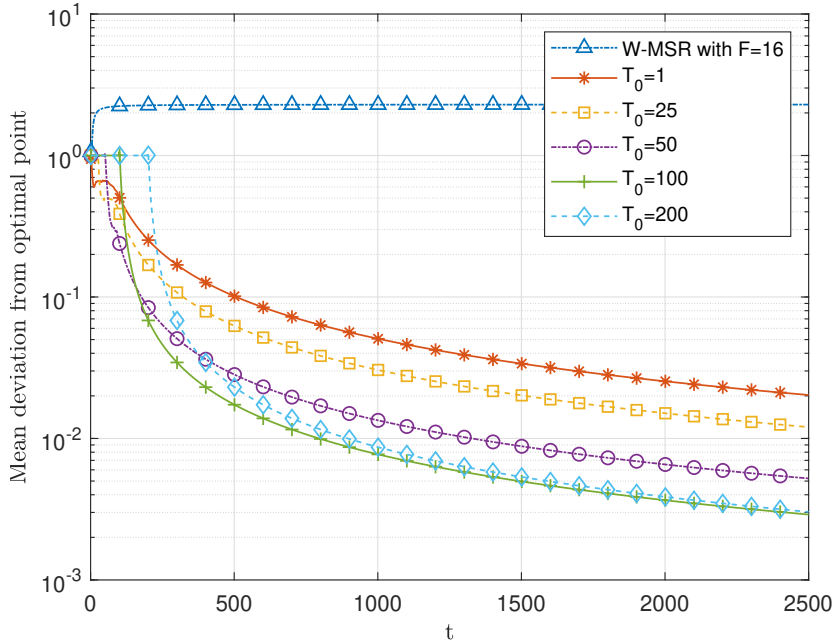
Classical bound must fulfill $|\mathcal{M}| < \frac{3+|\mathcal{M}|}{2} \Rightarrow |\mathcal{M}| < 3$.

Numerical Results II

Denote $\bar{e}(t) \triangleq \frac{1}{|\mathcal{L}|} \sum_{i \in \mathcal{L}} |x_i(t) - x^\star_{\mathcal{L}}|$.



(a) $\frac{\bar{e}(t)}{\bar{e}(0)}$ for $|\mathcal{M}| = 15,\ \ell = 0.8$.



(b) $\frac{\bar{e}(t)}{\bar{e}(0)}$ for $|\mathcal{M}| = 30,\ \ell = 0.6$.

Numerical Results III

Next, we extend these numerical results to a higher-dimensional setup where $d = 5$.

$$\min_{x \in [-\eta, \eta]^d} \left\{ \frac{1}{|\mathcal{L}|} \sum_{i \in \mathcal{L}} \frac{1}{2} (a_i^T x - b_i)^2 + \frac{1}{2} \|x\|^2 \right\}. \tag{14}$$

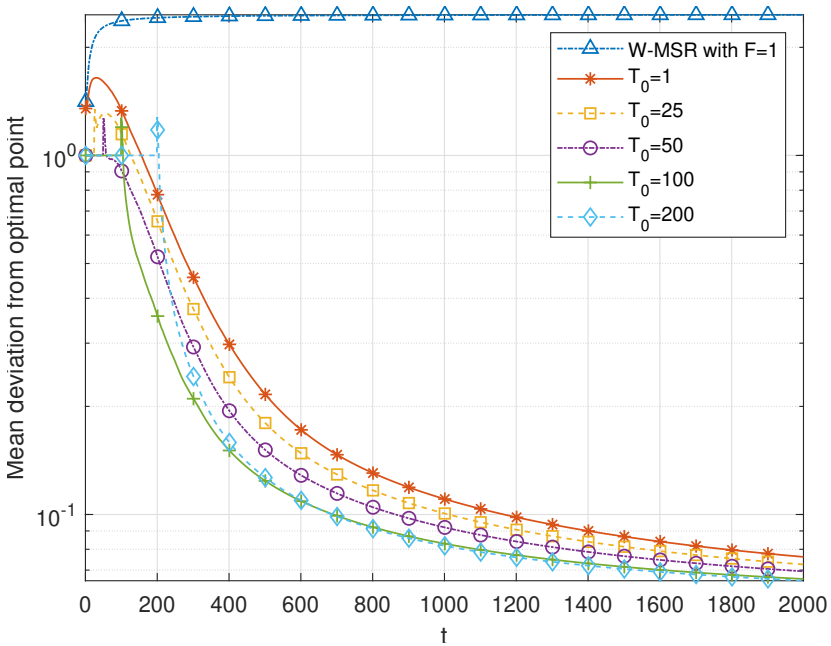
In this setup

$$\nabla f_i(x) = a_i (a_i^T x - b_i) + x.$$

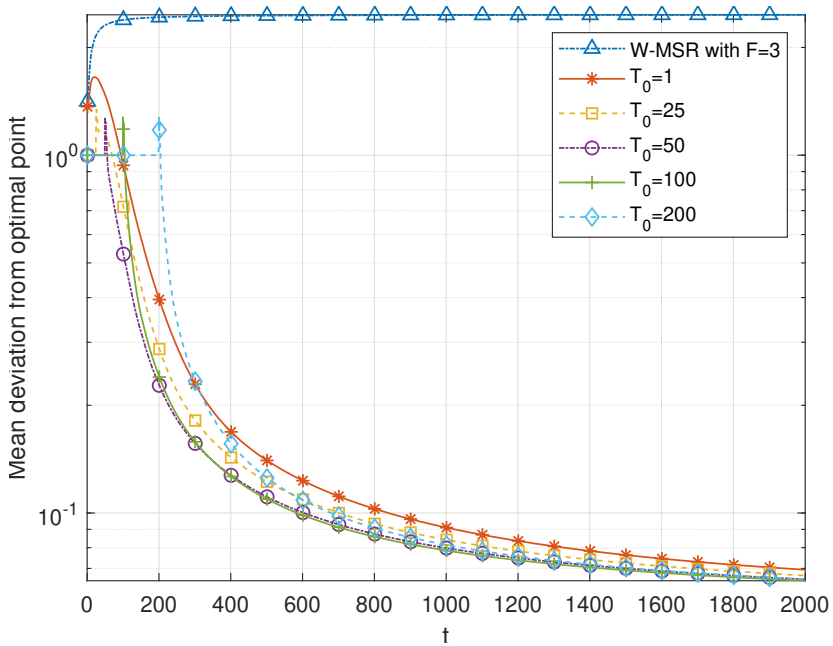
The projection is with respect to the 5-dimensional box $[-50, 50]^5$.

Numerical Results IV

Denote $\bar{e}(t) \triangleq \frac{1}{|\mathcal{L}|} \sum_{i \in \mathcal{L}} \|x_i(t) - x^\star_{\mathcal{L}}\|$.



(a) $\frac{\bar{e}(t)}{\bar{e}(0)}$ for $|\mathcal{M}| = 15,\ \ell = 0.8$.



(b) $\frac{\bar{e}(t)}{\bar{e}(0)}$ for $|\mathcal{M}| = 30,\ \ell = 0.6$.

Conclusions

- Physical-based trust values to increase resiliency to malicious inputs.
- Trust-based weight matrix.
- Finite detection time a.s.
- Convergence to the optimal nominal value.
- Expected convergence rate.
- Numerical results that validate our analytical results.

Thank you!

Reach out for further discussions: michal.yemini@biu.ac.il

CDC paper: *Resilience to Malicious Activity in Distributed Optimization for Cyberphysical Systems*, December 2022.

Journal version (with proofs) on arXiv:
Resilient Distributed Optimization for Multi-Agent Cyberphysical Systems: https://arxiv.org/pdf/2212.02459.pdf