Accelerating Distributed Optimization via Fixed-Time Convergent Flows

Kunal Garg * **Mayank Baranwal** **

* *Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: kgarg@mit.edu).*
** *Tata Consultancy Services Research, Mumbai, Maharashtra 400708 India (e-mail: baranwal.mayank@tcs.com)*

Abstract: Distributed optimization has gained significant attention in recent years, primarily fueled by the availability of a large amount of data and privacy-preserving requirements. This paper presents a fixed-time convergent optimization algorithm for solving a potentially non-convex optimization problem using a first-order multi-agent system. Each agent in the network can access only its private objective function, while local information exchange is permitted between the neighbors. The proposed optimization algorithm combines a fixed-time convergent distributed parameter estimation scheme with a fixed-time distributed consensus scheme as its solution methodology. The results are presented under the assumption that the team objective function is strongly convex, as opposed to the common assumptions in the literature requiring each of the local objective functions to be strongly convex. The results extend to the class of possibly non-convex team objective functions satisfying only the Polyak-Łojasiewicz (PL) inequality. It is also shown that the proposed continuous-time scheme, when discretized using Euler's method, leads to consistent discretization.

Keywords: Distributed optimization; Fixed-time stability; Consistent discretization

1. INTRODUCTION

Over the past decade, distributed optimization problems over a peer-to-peer network have received considerable attention due to the size and complexity of the dataset, privacy concerns, and communication constraints among multiple agents (Lin et al., 2017; Pan et al., 2018). These distributed convex optimization problems take the following form:

$$\min_{\mathbf{x}\in\mathbb{R}^d} \quad F(\mathbf{x}) = \sum_{i=1}^{N} f_i(\mathbf{x}), \tag{1}$$

where $F(\cdot)$ is the team objective function, and the convex function $f_i : \mathbb{R}^d \to \mathbb{R}$ represents the local objective function of the i^{th} agent, where $i \in \{1, 2, \dots, N\}$ for some positive integer N. Distributed optimization problems find applications in several domains including, but not limited to, sensor networks, satellite tracking (Hu and Shao, 2016), and large-scale machine learning (Nathan and Klabjan, 2017). Distributed optimization problems facilitate distributed coordination among the agents, as well as minimization of the team objective function. Consequently, these problems are inherently more complex than other multi-agent control problems, such as, distributed consensus.

In recent years, the use of continuous-time dynamical systems for distributed optimization has emerged as a viable alternative (Lin et al., 2017; Pan et al., 2018; Feng and Hu, 2017; Hu and Yang, 2018). This viewpoint enables the use of tools from Lyapunov theory and differential equations for the analysis and design of optimization procedures. It is worth mentioning that most of the existing continuous-time schemes for distributed optimization are only asymptotically (or exponentially at best) convergent. On the other hand, most practical multi-agent optimization tasks, such as distributed economic dispatch, often undergo frequent changes in operating conditions, thereby requiring the optima to be achieved in a finite time.

The notion of finite-time convergence in optimization is closely related to finite-time stability (Bhat and Bernstein, 2000) in control theory. In contrast to asymptotic stability (AS), finite-time stability is a concept that guarantees the convergence of solutions in a finite amount of time. In (Lu and Tang, 2012), a continuous-time zero-gradient-sum (ZGS) with an exponential convergence rate was proposed, which, when combined with a finite-time consensus protocol, was shown to achieve finite-time convergence in (Feng and Hu, 2017). A drawback of ZGS-type algorithms is the requirement of strong convexity of the local objective functions and the choice of specific initial conditions $x_i(0)$ for each agent i such that $\sum_{i=1}^{N} \nabla f_i(x_i(0)) = 0$. In (Lin et al., 2017), a novel continuous-time distributed optimization algorithm, based on private (nonuniform) gradient gains, was proposed for convex functions with quadratic growth and achieved convergence in a finite time. A finite-time tracking and consensus-based algorithm were recently proposed in (Hu and Yang, 2018), which again achieves convergence in a finite time under a time-invariant communication topology.

Fixed-time stability (FxTS) (Polyakov, 2012) is a stronger notion than finite-time stability (FTS), where the time of convergence does not depend upon the initial condition.

To the best of our knowledge, distributed optimization procedures with fixed-time convergence have not been addressed in the literature for a general class of non-linear, potentially non-convex, objective functions. The use of FxTS theory for distributed optimization was first investigated in (Garg et al., 2020) where centralized optimization problems were studied. The authors in (Wang et al., 2020) further specialized it to the case of strongly convex functions, however, at the expense of using a Hessian-based (second-order) schemes that do not scale well with the dimension d of the underlying state-space. Moreover, the distributed protocol in (Wang et al., 2020) requires each of the individual private objective functions to be strongly convex. In the particular case of quadratic objective functions, the scheme proposed in (Garg et al., 2020) can be suitably modified to incorporate both inequality and equality constraints (Baranwal et al., 2020).

Despite growing interests in the use of continuous-time dynamical systems towards distributed optimization with fixed-time convergence guarantees, the existing literature makes various simplifying assumptions, including but not limited to, requiring agents to satisfy ZGS condition, use of second-order (Hessian-based) optimization schemes, necessitating all private objective functions to be strongly convex or with bounded growth, and existence of a time-invariant communication topology. In addition, prior work does not discuss how efficient their proposed methods are during implementation using iterative, discrete methods. It is worth noting that while continuous-time dynamical systems are studied for ease of understanding the behavior of an optimization algorithm, in practice, it is inevitable to use a discrete-time, iterative method to solve optimization problems. In view of the limitations stated above, this paper presents a fixed-time convergent, distributed optimization scheme that extend to a broad class of local objective functions under relaxed assumptions on convexity and information to be exchanged with the neighbors.

- We consider the problem of distributed optimization of the sum of local objective functions, assuming that only the global objective function is strongly convex.
- The results extend to a class of possibly non-convex functions satisfying only the Polyak-Łojasiewicz (PL) inequality. PL inequality is a relaxation of strong-convexity and is popularly used to design exponentially stable gradient-flows in the centralized optimization problems (Garg and Panagou, 2020). To the best of the our knowledge, this is the first work that utilizes this condition in distributed optimization.
- We show that trajectories of dynamics obtained by discretizing the proposed continuous-time dynamics using Euler discretization converge to an arbitrarily small neighborhood of the optimal point within a fixed number of iterations, leading to a *consistent* discretization. This is a rather significant result as it bridges the gap between the continuous-time analysis and discrete-time implementation.

A note on mathematical notations: We use $\mathbb{R}$ to denote the set of real numbers and $\mathbb{R}_+$ to denote non-negative real numbers. Given a function $f : \mathbb{R}^d \to \mathbb{R}$, the gradient at some point $x \in \mathbb{R}^d$ is denoted by $\nabla f(x)$. $\|x\|$ denotes the 2-norm of x. $\mathcal{G} = (A, \mathcal{V})$ represents an undirected graph with the adjacency matrix $A = [a_{ij}] \in \mathbb{R}^{N\times N}$, $a_{ij} \in \{0, 1\}$ and the set of nodes $\mathcal{V} = \{1, 2, \cdots, N\}$. The set of 1-hop neighbors of node $i \in \mathcal{V}$ is represented by $\mathcal{N}_i$, i.e., $\mathcal{N}_i(t) = \{j \in \mathcal{V} \mid a_{ij} = 1\}$. The second smallest eigenvalue of a matrix is denoted by $\lambda_2(\cdot)$. We define the function $\text{sign}^\mu : \mathbb{R}^d \to \mathbb{R}^d$ as

$$\text{sign}^\mu(x) = x\|x\|^{\mu-1}, \quad \mu \geq 0, \tag{2}$$

with $\text{sign}^\mu(0) = 0$. We use $1_N, 0_N \in \mathbb{R}^N$ to denote vectors consisting of ones and zeros, respectively, of dimension N.

2. PROBLEM FORMULATION

2.1 Problem statement

Consider the system consisting of N nodes with graph structure $\mathcal{G} = (A, \mathcal{V})$ specifying the communication links. The objective is to find $x^* \in \mathbf{R}^d$ that solves

$$\begin{aligned} &\min_{x_1, x_2 \cdots, x_N} \sum_{i=1}^N f_i(x_i), \\ &\text{s.t. } x_1 = x_2 = \cdots = x_N. \end{aligned} \tag{3}$$

In this work, we assume that the minimizer $x^* = x_1^* = x_2^* = \cdots = x_N^*$ for (3) exists and is unique. [1] We make the following assumption on the inter-node communications.

Assumption 1. *The communication topology between the agents is connected and undirected.*

To motivate the dynamical system approach considered, let us first revisit the gradient decent (GD) method to minimize an unconstrained function $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$:

$$x_{k+1} = x_k - \eta \, \nabla\mathcal{F}(x_k),$$

where $\eta > 0$ is the step-size. We can re-write the above as $\frac{x_{k+1}-x_k}{\eta} = -\nabla\mathcal{F}(x_k)$ and in the limit $\eta \to 0$, we obtain the continuous-time equivalent of GD, termed as gradient-flow, given as $\dot{x} = -\nabla\mathcal{F}(x)$. More generally, we can write this dynamical system as $\dot{x} = u$ where u can be designed to solve a given problem (e.g., for unconstrained minimization of $\mathcal{F}$, $u = -\nabla\mathcal{F}$ and for constrained minimization of $\mathcal{F}$ over a convex set $\mathcal{C}$, one can define $u = -k(x - \text{P}_\mathcal{C}(x - \nabla\mathcal{F}(x)))$ using the projection operator $\text{P}_\mathcal{C}$. We use this dynamical systems viewpoint to solve the constrained optimization problem (3) in a distributed fashion. Let $x_i \in \mathbb{R}^d$ represent the state of agent i modeled using a first-order integrator system:

$$\dot{x}_i = u_i, \tag{4}$$

where $u_i \in \mathbb{R}^d$ can be regarded as a *control input*, that depends upon the states of the agent i, and the states of the neighboring agents $j_1, j_2, \cdots, j_l \in \mathcal{N}_i$. The problem statement is formally given as follows.

Problem 1. *Design u_i for each agent $i \in \mathcal{V}$, such that $x_1 = x_2 = \cdots = x_N = x^*$ is achieved under (4) within a fixed time, for any initial condition $\{x_1(0), x_2(0), \cdots, x_N(0)\}$, where x^* solves (3).*

2.2 Preliminaries

In this subsection, we present relevant definitions and results on FxTS. Consider the system:

$$\dot{x} = \phi(x), \tag{5}$$

[1] Existence and uniqueness of global minimizer is trivially satisfied for a strongly convex team objective function. While the PL inequality (see Assumption 4) does not imply convexity, it implies invexity, i.e., the stationary points are global minimizers.

where $x \in \mathbb{R}^d$, $\phi : \mathbb{R}^d \to \mathbb{R}^d$ and $\phi(0) = 0$. The authors in (Polyakov, 2012) presented the following result for fixed-time stability, where the time of convergence is finite and is uniformly bounded for any initial condition $x(0)$.

Lemma 1 ((Polyakov, 2012)). *Suppose there exists a positive definite, radially unbounded, continuously differentiable function $V : \mathbb{R}^d \to \mathbb{R}$, i.e., $V \in \mathcal{C}^1$ such that $V(0) = 0$ and $V(x) > 0$ for $x \neq 0$, and:*

$$\dot{V}(x) \leq -aV(x)^p - bV(x)^q, \quad \forall x \neq 0, \tag{6}$$

with $a, b > 0$, $0 < p < 1$ and $q > 1$. Then the origin of (5) is FxTS, i.e., $x(t) = 0$ for all $t \geq T$, where the settling time T satisfies $T \leq \frac{1}{a(1-p)} + \frac{1}{b(q-1)}$.

Next, we present some well-known results that will be useful in proving our claims on fixed-time parameter estimation and consensus protocols.

Lemma 2 ((Zuo and Tie, 2016)). *Let $z_i \in \mathbb{R}_+$ for $i \in \{1, 2, \cdots, N\}$, $N \in \mathbb{Z}_+$. Then the following hold:*

$$\sum_{i=1}^{N} z_i^p \geq \left(\sum_{i=1}^{N} z_i\right)^p, \ 0 < p \leq 1, \tag{7a}$$

$$\sum_{i=1}^{N} z_i^p \geq N^{1-p}\left(\sum_{i=1}^{N} z_i\right)^p, \ p > 1. \tag{7b}$$

Lemma 3. *Let $\mathcal{G} = (A, \mathcal{V})$ be an undirected graph consisting of N nodes located at $x_i \in \mathbb{R}^d$ for $i \in \{1, 2, \cdots, N\}$ and $\mathcal{N}_i$ denotes the in-neighbors of node i. Then,*

$$\sum_{i=1}^{N} \sum_{j \in \mathcal{N}_i} \mathrm{sign}(x_i - x_j) = 0. \tag{8}$$

Lemma 4. *Let $w : \mathbb{R}^d \to \mathbb{R}^d$ be an odd mapping, i.e., $w(x) = -w(-x)$ for all $x \in \mathbb{R}^d$ and let the graph $\mathcal{G} = (A, \mathcal{V})$ be undirected. Let $\{x_i\}$ and $\{e_i\}$ be the sets of arbitrary vectors with $i \in \mathcal{V}$ and $x_{ij} := x_i - x_j$ and $e_{ij} := e_i - e_j$. Then, the following holds*

$$\sum_{i,j=1}^{N} a_{ij} e_i^{\mathsf{T}} w(x_{ij}) = \frac{1}{2}\sum_{i,j=1}^{N} a_{ij} e_{ij}^{\mathsf{T}} w(x_{ij}). \tag{9}$$

Lemma 5 ((Mesbahi and Egerstedt, 2010)). *Let $\mathcal{G} = (A, \mathcal{V})$ be an undirected, connected graph. Let $L_A := [l_{ij}] \in \mathbb{R}^{N \times N}$ be graph Laplacian matrix defined as $l_{ij} = \begin{cases} \sum_{k=1, k\neq i}^{N} a_{ik}, & i = j \\ -a_{ij}, & i \neq j \end{cases}$. Then the Laplacian L_A has following properties:*

1) L_A is positive semi-definite, $L_A 1_N = 0_N$, $\lambda_2(L_A) > 0$.

2) $x^{\mathsf{T}} L_A x = \frac{1}{2}\sum_{i,j=1}^{N} a_{ij}(x_j - x_i)^2$, and if $1^{\mathsf{T}} x = 0$, then $x^{\mathsf{T}} L_A x \geq \lambda_2(L_A) x^{\mathsf{T}} x$.

3. MAIN RESULTS

Our approach to fixed-time multi-agent distributed optimization is based on first designing a centralized fixed-time protocol that relies upon global information. Then, the quantities in the centralized protocol are estimated in a distributed manner. In summary, the algorithm proceeds by first estimating global quantities (g^* as defined in (11)) required for the centralized protocol, then driving the agents to reach consensus ($x_i(t) = x(t)$ for all $i \in \mathcal{V}$), and finally driving the *common* trajectory $x(t)$ to the optimal point x^*, all within a fixed time T. Recall that agents are said to have reached consensus on states x_i if $x_i = x_j$ for all $i, j \in \mathcal{V}$. To this end, we define first a centralized fixed-time protocol. Note that agents' states are driven by the same input under centralized settings and are initialized to the same starting point. In a distributed setting, this behavior translates to agents having already reached consensus and subsequently being driven by a common input (see Remark 3). We make the following assumptions.

Assumption 2. *Functions f_i are convex, twice differentiable and the Hessian $\nabla^2 F(x) = \sum_{i=1}^{N} \nabla^2 f_i(x) \succeq kI$, where $k > 0$, for all $x \in \mathbb{R}^d$, i.e., function F is strongly convex with modulus k.*

Remark 1. *Assumption 2 can be satisfied even if just one of the objective functions is strongly convex.*

Assumption 3. *Each node i receives $x_j, \nabla f_j(x_j)$ from each of its neighboring nodes $j \in \mathcal{N}_i$.*

Note that under Assumption 2, the agents only need to exchange their state values x_i and the gradients $\nabla f_i(x_i)$ with their neighbors. We first present a centralized protocol that guarantees solution of (3) in a fixed time. All the results in the following section assume that Assumptions 1, 2, 3 hold, unless specified otherwise.

3.1 Centralized protocol

Lemma 6 (**Centralized fixed-time protocol**). *Suppose the dynamics of each agent $i \in \mathcal{V}$ is given by*

$$u_i = g^*, \quad x_i(0) = x_j(0) \quad \forall\ i, j \in \mathcal{V}, \tag{10}$$

where g^* is defined as:

$$g^*(x) = -\left(\sum_{i=1}^{N} \nabla f_i(x) + \mathrm{sign}^{l_1}\left(\sum_{i=1}^{N} \nabla f_i(x)\right) + \mathrm{sign}^{l_2}\left(\sum_{i=1}^{N} \nabla f_i(x)\right)\right) \tag{11}$$

where $l_1 > 1$ and $0 < l_2 < 1$, and $x_i = x$ for each $i \in \mathcal{V}$, for all $t \geq 0$. Then the trajectories of all agents converge to the optimal point x^*, i.e., the minimizer of the team objective function (3) in a fixed time $T_{sc} > 0$.

Proof. Consider a candidate Lyapunov function:

$$V(x) = \frac{1}{2}\left(\sum_{i=1}^{N} \nabla f_i(x)\right)^{\mathsf{T}}\left(\sum_{i=1}^{N} \nabla f_i(x)\right).$$

By taking its time-derivative along (10), we obtain:

$$\begin{aligned}
\dot{V}(x) &= \left(\sum_{i=1}^{N} \nabla f_i(x)\right)^{\mathsf{T}}\left(\sum_{i=1}^{N} \nabla^2 f_i(x)\dot{x}\right) \\
&= -\left(\sum_{i=1}^{N} \nabla f_i\right)^{\mathsf{T}}\left(\sum_{i=1}^{N} \nabla^2 f_i\right)\left(\sum_{i=1}^{N} \nabla f_i \right. \\
&\quad \left. + \mathrm{sign}^{l_1}\left(\sum_{i=1}^{N} \nabla f_i\right) + \mathrm{sign}^{l_2}\left(\sum_{i=1}^{N} \nabla f_i\right)\right), \\
&\leq -2kV - k\left\|\sum_{i=1}^{N} \nabla f_i\right\|^{l_1+1} - k\left\|\sum_{i=1}^{N} \nabla f_i\right\|^{l_2+1} \\
&\leq -k2^{\frac{1+l_1}{2}} V^{\frac{1+l_1}{2}} - k2^{\frac{1+l_2}{2}} V^{\frac{1+l_2}{2}},
\end{aligned}$$

where the first inequality follows from the fact that $\sum_{i=1}^{N} \nabla^2 f_i \succeq kI$. Thus, using Lemma 1, we have that there exists $T_{sc} < \infty$ such that for all $t \geq T_{sc}$, $x(t) = x^*$ starting from any initial condition. □

The centralized fixed-time protocol inherently assumes that the agents can directly access the global quantity $\sum_{i=1}^{N} \nabla f_i$. In a distributed setting, this quantity needs to be estimated and is not directly accessible. Before presenting the algorithm to compute this global quantity in a distributed manner, we first present an extension of Lemma 6 under further relaxation of Assumption 2. The notion of *gradient-dominance* or Polyak-Łojasiewicz (PL) inequality has been explored extensively in optimization literature to show exponential convergence. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to satisfy PL inequality, or is gradient dominated, with $\mu_f > 0$ if

$$\frac{1}{2}\|\nabla f(x)\|^2 \geq \mu_f(f(x) - f^*) \quad \forall x \in \mathbb{R}^n, \tag{12}$$

where $f^* = f(x^*)$ is the value of the function at its minimizer x^*. We make the following assumption on the team objective function.

Assumption 4. ***(Gradient dominated)*** *The function F is radially unbounded, has a unique minimizer $x = x^*$, and satisfies the PL inequality, or is gradient dominated, i.e., there exists $\mu > 0$ such that*

$$\frac{1}{2}\left\|\sum_{i=1}^{N}\nabla f_i(x)\right\|^2 \geq \mu(F(x) - F^*) = \mu\sum_{i=1}^{N}(f_i(x) - f_i^*), \tag{13}$$

where $F^* = F(x^*)$ and $f_i^* = f_i(x^*)$.

Remark 2. *As noted in (Karimi et al., 2016), PL inequality is the weakest condition among other similar conditions popularly used in the literature to show linear convergence in discrete-time (exponential, in continuous-time). Notably, a strongly convex function F satisfies PL inequality. Furthermore, note that under Assumption 4, it is not required that the function F is convex, as long as its minimizer exists and is unique.*

It is easy to show that if a function $F : \mathbb{R}^m \to \mathbb{R}$ is strongly convex, then the function $G : \mathbb{R}^n \to \mathbb{R}$, defined as $G(x) = F(Ax)$, where $A \in \mathbb{R}^{n\times m}$ is not full row-rank, may not be strongly convex. On the other hand, as shown in (Karimi et al., 2016, Appendix 2.3), G still satisfies PL inequality for any matrix A. Below, an example of an important class of problems is given for which the objective function satisfies PL inequality.

Example 1. ***Least squares****: Consider the problem*

$$\min_x \|Ax - b\|^2 = \sum_{i=1}^{n}\|A_i x - b_i\|^2, \tag{14}$$

where $x \in \mathbb{R}^n, A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$. Here, the function $F(x) = \|x - b\|^2$ is strongly-convex, and hence, $G(x) = \|Ax - b\|^2$ satisfies PL inequality for any matrix A.

The objective function (14) satisfies PL inequality, but need not be strongly convex for any A, thus, one can use (10) to find the optimal solution for (14) in a fixed time.

Lemma 7. *Let Assumption 4 hold. Suppose the dynamics of each agent $i \in \mathcal{V}$ in the network is given by* (10) *where g^* given as* (11) *with $x_i(t) = x(t)$ for each $i \in \mathcal{V}$, for all $t \geq 0$. Then the trajectories of all agents converge to the optimal point x^*, i.e., the minimizer of the team objective function* (3) *in a fixed time $T_{PL} > 0$.*

Proof. Consider the candidate Lyapunov function as $V(x) = \sum_{i=1}^{N}(f_i(x) - f_i(x^*)) = (F(x) - F^*)$. Note that V is positive definite and per Assumption 4, radially unbounded. Taking its time derivative along the trajectories of (10), we obtain

$$\begin{aligned}
\dot{V}(x) &= -\sum_{i=1}^{N}\nabla f_i^{\intercal}\left(\sum_{i=1}^{N}\nabla f_i + \text{sign}^{l_1}\left(\sum_{i=1}^{N}\nabla f_i\right)\right.\\
&\qquad \left. + \text{sign}^{l_2}\left(\sum_{i=1}^{N}\nabla f_i\right)\right)\\
&= -\|\nabla F(x)\|^2 - \|\nabla F(x)\|^{l_1+1} - \|\nabla F(x)\|^{l_2+1}\\
&\overset{(13)}{\leq} -2\mu(F(x) - F^*) - (2\mu)^{\frac{1+l_1}{2}}(F(x) - F^*)^{\frac{1+l_1}{2}}\\
&\qquad - (2\mu)^{\frac{1+l_2}{2}}(F(x) - F^*)^{\frac{1+l_2}{2}}\\
&\leq -4\mu V(x) - (4\mu)^{\frac{1+l_1}{2}}V(x)^{\frac{1+l_1}{2}} - (4\mu)^{\frac{1+l_2}{2}}V(x)^{\frac{1+l_2}{2}}\\
&\leq -k_1 V(x)^{\frac{1+l_1}{2}} - k_2 V(x)^{\frac{1+l_2}{2}}.
\end{aligned}$$

Thus, using Lemma 1, we obtain that there exits $T_{PL} < \infty$ such that for all $t \geq T_{PL}$, we have that $V(x(t)) = 0$, or equivalently, $F(x(t)) = F^*$. Under Assumption 4, we have that F has a unique minimizer, and thus, $F(x(t)) = F^*$ implies that $x(t) = x^*$, which completes the proof. ☐

Remark 3. *Lemmas 6 and 7 represent centralized protocols for convex optimization of team objective functions. Here, the agents are already in consensus and have access to the global information $\sum_{i=1}^{N}\nabla f_i(x)$. In the distributed setting, agents can only access their local information, as well as x_j, $\nabla f_j(x_j)$ for all $j \in \mathcal{N}_i$, and will not be in consensus in the beginning.*

3.2 Distributed estimation of global parameters

We now present results for distributed estimation of global quantity that achieves consensus in a fixed time so that the problem can be solved in a distributed setting. For each agent $i \in \mathcal{V}$, define g_i as:

$$g_i = -\left(N\theta_i + \text{sign}^{l_1}(N\theta_i) + \text{sign}^{l_2}(N\theta_i)\right), \tag{15}$$

where g_i denotes agent i's estimate of g^* and $\theta_i : \mathbb{R}_+ \to \mathbb{R}^d$ is the estimate of the global (centralized) quantities, whose dynamics is defined as

$$\dot{\theta}_i = \omega_i + h_i, \tag{16}$$

where h_i is defined as $h_i = \frac{d}{dt}\nabla f_i(x_i)$. The signal $\omega : \mathbb{R}_+ \to \mathbb{R}^d$, defined as

$$\begin{aligned}
\omega_i = p\sum_{j\in\mathcal{N}_i}&\Big(\text{sign}(\theta_j - \theta_i) + \gamma\,\text{sign}^{\nu_1}(\theta_j - \theta_i)\\
&+ \delta\,\text{sign}(\theta_j - \theta_i)^{\nu_2}\Big),
\end{aligned} \tag{17}$$

where $p, \gamma, \delta > 0$, and $0 < \nu_2 < 1 < \nu_1$, are suitably chosen in order to achieve consensus over the quantities θ_i, as shown later. The functions $\{h_i\}$ are needed to drive the consensus values to the global quantities to be estimated. Observe that $\{\theta_i\}$ are updated in (16) in a *distributed* manner. We make the following assumption on h_i.

Assumption 5. *The functions h_i, h_j satisfy $\|h_i(t) - h_j(t)\| \leq \rho$ for all $t \geq 0$, $i, j \in \mathcal{V}, i \neq j$, for some $\rho > 0$.*

This assumption can be easily satisfied if the graph is connected for all time t and the gradients and their derivatives are bounded (Hu and Yang, 2018).

Lemma 8. *Let Assumption 5 hold, and the gain p in* (17) *satisfies $p > \left(\frac{N-1}{2}\right)\rho$; then for each agent $i \in \mathcal{V}$, $\theta_i(t) = \theta_c(t) := \frac{1}{N}\sum_{j=1}^{N}\theta_j(t) = \frac{1}{N}\sum_{i=1}^{N}\nabla f_i(x_i(t))$, for all $t \geq T_p$ where*

$$T_p := \frac{2}{p\gamma N^{2(1-\kappa_1)}c^{\kappa_1}(\kappa_1 - 1)} + \frac{2}{p\delta c^{\kappa_2}(1 - \kappa_2)},$$

$\kappa_1 = \frac{1+\nu_1}{2}, \kappa_2 = \frac{1+\nu_2}{2}$ and $c = 4\lambda_2(L_A)$.

The proof is provided in Appendix A.

Lemma 9 (**Fixed-time parameter estimation**). *Let $\omega_i(0) = \mathbf{0}_d$ for each $i \in \mathcal{V}$ and the gain p in* (17) *satisfy $p > \left(\frac{N-1}{2}\right)\rho$. Then there exists a fixed-time $0 < T_p < \infty$ such that $g_i(t) = g_j(t)$ for all $i, j \in \mathcal{V}$ and $t \geq T_p$.*

Proof. The proof follows directly from Lemma 8, i.e., it holds that $\theta_i(t) = \theta_j(t)$ for all $t \geq T_{\text{p}}$, $i, j \in \mathcal{V}$. From the definition of g_i in (15), it follows that $g_i(t) = g_j(t)$ for all $t \geq T_{\text{p}}$ and for each $i, j \in \mathcal{V}$. $\square$

The centralized fixed-time protocol in Lemma 6 is based on two key assumptions: (a) Agents are being driven by the same input g^*, and (b) agents start at the same initial state,, i.e., $x_i(0) = x_j(0)$ for all $i, j \in \mathcal{V}$. To this end, Lemma 9 only ensures that the first of the two conditions is met. All agents must be driven to the same state in order to ensure the applicability of Lemma 6 in the distributed setting. Consequently, we propose the following update rule for each agent $i \in \mathcal{V}$ in the network:

$$u_i = \tilde{u}_i + g_i, \tag{18}$$

where g_i is as described in (15), and $\tilde{u}_i$ is defined as locally averaged signed differences:

$$\begin{aligned}\tilde{u}_i = q\sum\nolimits_{j\in\mathcal{N}_i}\Big(&\text{sign}(x_j - x_i) + \alpha\text{sign}^{\mu_1}(x_j - x_i)\\ &+ \beta\text{sign}^{\mu_2}(x_j - x_i)\Big),\end{aligned} \tag{19}$$

where $q, \alpha, \beta > 0$, $\mu_1 > 1$ and $0 < \mu_2 < 1$. The following results establish that the state update rule for each agent proposed in (18) ensures that the agents reach global consensus and optimality in fixed time.

Lemma 10 (**Fixed-time consensus**). *Under the effect of control law u_i in* (18) *with $\tilde{u}_i$ defined as in* (19)*, and $g_i(t) = g_j(t)$ for all $t \geq T_p$ and $i, j \in \mathcal{V}$, the closed-loop trajectories of* (4) *converge to a common point $\bar{x}$ for all $i \in \mathcal{V}$ in a fixed time T_{con}, i.e., $x_i(t) = \bar{x}(t)$ for all $t \geq T_p + T_{con}$.*

Proof. The proof follows from Lemma 8 and the fact that $g_i(t) = g_j(t)$ for all $t \geq T_{\text{p}}$, $i, j \in \mathcal{V}$. Thus, for $t \geq T_{\text{p}}$, the dynamics of agent i in the network is described by $\dot{x}_i(t) = \tilde{u}_i(t) + g_i(t)$ with $\|g_i(t) - g_j(t)\| = 0$ for all $i, j \in \mathcal{V}$. Moreover, $\tilde{u}_i$ has a form similar to ω_i. Thus, from Lemma 8, it follows that there exists a $T_{\text{con}} > 0$ such that $x_i(t) = \dfrac{1}{N}\sum_{j=1}^{N} x_j(t)$ for $t \geq T_{\text{p}} + T_{\text{con}}$, where T_{con} satisfies $T_{\text{con}} \leq \frac{2}{q\alpha N^{2(1-\tau_1)}\tilde{c}^{\tau_1}(\tau_1-1)} + \frac{2}{q\beta\tilde{c}^{\tau_2}(1-\tau_2)}$, where $\tau_1 := \frac{1+\mu_1}{2}$, $\tau_2 := \frac{1+\mu_2}{2}$, $\tilde{c} > 0$ is an appropriate constant. $\square$

Finally, the following result establishes that the agents track the optimal point in a fixed time.

Theorem 1 (**Fixed-time distributed optimization**). *Let each agent $i \in \mathcal{V}$ in the network be driven by the control input u_i* (18)*. If the functions satisfy Assumption 2 (respectively, Assumption 4), then the agents track the minimizer of the team objective function within fixed time $T = T_{\text{p}} + T_{\text{con}} + T_{sc}$ (respectively, $T = T_{\text{p}} + T_{\text{con}} + T_{PL}$).*

Proof. The proof follows directly from the previous results presented in this section. From Lemmas 8 and 10, it follows that $g_i(t) = g_j(t)$ for all $t \geq T_{\text{p}}$, and $x_i(t) = x_j(t)$ for all $t \geq T_{\text{p}} + T_{\text{con}}$. Since g_i is a function of θ_i, and from Lemma 9, we have that $\theta_i(t) = \sum_j \nabla f_j(x_j(t))$ for all $t \geq T_{\text{p}}$, with $\nabla f_i(x_i(t)) = \nabla f_j(x_j(t))$ for all $t \geq T_{\text{p}} + T_{\text{con}}$, we obtain that $g_i(t) = g^*(t)$ and $\tilde{u}_i(t) = 0$ for all $i \in \mathcal{V}$, $t \geq T_{\text{p}} + T_{\text{con}}$. Thus, if the objective functions satisfy Assumption 2 (respectively, Assumption 4), the conditions of the centralized fixed-time protocol in Lemma 6 are satisfied, and therefore, $x_i(t) = x^*$ for all $i \in \mathcal{V}$, for $t \geq T_{\text{p}} + T_{\text{con}} + T_{sc}$ (respectively, $t \geq T_{\text{p}} + T_{\text{con}} + T_{PL}$). $\square$

Note that the total time of convergence $T = T_{\text{p}} + T_{\text{con}} + T_{sc}$ ((respectively, $t \geq T_{\text{p}} + T_{\text{con}} + T_{PL}$)) depends upon the design parameters $p, q, \alpha, \beta, \gamma, \delta, \mu_1, \mu_2, \nu_1, \nu_2, l_1, l_2$. Hence, for a given user-defined time budget T_b, one can choose large values of these parameters so that $T \leq T_b$, and hence, convergence can be achieved within user-defined time T_b. The overall Fixed-time stable Distributed Optimization Algorithm (FxTS-DOA) with discrete-time iterative implementation is described in Algorithm 1.

Algorithm 1 Discretized FxTS-DOA

procedure FxTS Dist Opt($(A, \mathcal{V})$, $\{f_i(\cdot)\}$)
  Inputs: $p, q, l_1, l_2, \nu_1, \nu_2, \mu_1, \mu_2$; Step-size η
  Initialize local estimates $\{x_i\}$ for each $i \in \mathcal{V}$
  $\omega_i \leftarrow 0_{d\times 1}$ for each $i \in \mathcal{V}$
  $\theta_i \leftarrow 0_{d\times 1}$ for each $i \in \mathcal{V}$
  for $k = 1$, $k \leq$ `max-epochs` **do**
    Each agent computes its own gradient $\nabla_i f_i(x_i)$
    $\bar{u}_i \leftarrow q\big((x_j - x_i) + \text{sign}^{\mu_1}(x_j - x_i) + \text{sign}^{\mu_2}(x_j - x_i)\big)$
    $\omega_i \leftarrow \eta\big((\theta_j - \theta_i) + \text{sign}^{\nu_1}(\theta_j - \theta_i) + \text{sign}^{\nu_2}(\theta_j - \theta_i)\big)$
      ▷ Information sharing with neighbors
    $\theta_i \leftarrow \omega_i + \eta\nabla_i f_i(x_i)$
    $g_i \leftarrow -\big((N\theta_i) + \text{sign}^{l_1}(N\theta_i) + \text{sign}^{l_2}(N\theta_i)\big)$
    $x_i \leftarrow x_i + \eta(g_i + \bar{u}_i)$
      ▷ Agents update their estimates locally
  end for
  return $x_1 = x_2 = \cdots = x^*$
end procedure

4. DISCRETIZATION OF THE FXTS-DOA

Continuous-time dynamical systems, such as the one given by (4) with u_i given by (18), offer effective insights into designing accelerated schemes for solving a distributed optimization problem. However, in practice, a discrete-time implementation is used for solving optimization problems. (Polyakov et al., 2019) defines a discretization to be consistent with a fixed-time convergent dynamical system if the trajectories of the discretized system converge to an arbitrarily small neighborhood of the equilibrium point of the continuous-time system within a fixed number of steps, independent of the initial conditions. In order to prove that an Euler discretization scheme of the proposed method in Section 3 leads to a consistent discretization, it is sufficient to show that the closed-loop dynamics (4) under u_i given by (18) satisfies the conditions of (Garg et al., 2022, Theorem 3). Consider the proposed algorithm in Section 3. For $0 \leq t \leq T_{\text{p}} + T_{\text{con}}$, the dynamics for all θ_i, x_i can be written in a compact form as:

$$\dot{\theta} = F_1(\theta) + F_2(x), \quad \dot{x} = F_3(x) + F_4(\theta), \tag{20}$$

where

$$F_1(\theta) = \begin{bmatrix}\omega_1\\ \vdots\\ \omega_N\end{bmatrix}, F_2(x) = \begin{bmatrix}h_1\\ \vdots\\ h_N\end{bmatrix}, F_3(x) = \begin{bmatrix}\tilde{u}_1\\ \vdots\\ \tilde{u}_N\end{bmatrix}, F_4(\theta) = \begin{bmatrix}g_1\\ \vdots\\ g_N\end{bmatrix}.$$

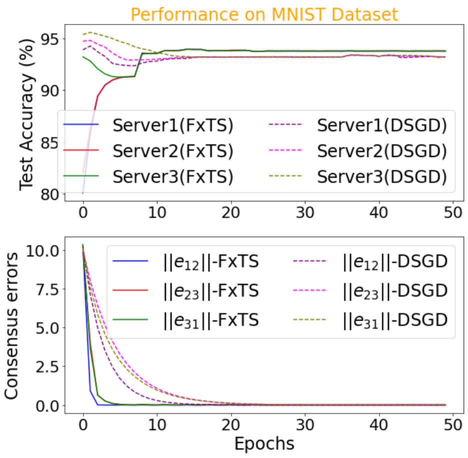

Fig. 1. Evaluation of FxTS-DOA for training of DNNs.

More compactly, define $z = [\theta^\intercal,\ x^\intercal]^\intercal \in \mathbb{R}^{2Nd}$ and $\mathcal{F}(z) := [(F_1(\theta) + F_2(x))^\intercal,\ F_3(x)^\intercal + F_4(\theta)^\intercal]^\intercal$ so that

$$\dot{z} \in \mathcal{F}(z). \tag{21}$$

We use the notion of differential inclusion in (21) since the right-hand side of (21) is not single-valued. The interested reader is referred to (Clarke et al., 2008) for more details. First, we show that the set-valued map $\mathcal{F}$ in (21) satisfies the conditions in (Garg et al., 2022, Theorem 3).

Lemma 11. *If the functions f_i satisfy either Assumption 2 (or Assumption 4) for all $i \in \mathcal{V}$, then $\mathcal{F}$ in (21) is upper semi-continuous set-valued map, taking non-empty, convex and compact values.*

Proof. Define $S = \{z \mid \mathcal{F}(z) = 0\}$ is the set of equilibrium points for the dynamics of variable z. Note that the equilibrium points of (21) are the points $x_i = x_j$ and $\theta_i = \theta_j$ for all $i \neq j$, which is a $2d-$dimensional subspace in $\mathbb{R}^{2Nd}$, and thus, S is a Lebesgue measure zero set in $\mathbb{R}^{2Nd}$. Note that the map $\mathcal{F}$ is continuous for all $z \in \mathbb{R}^{2Nd} \setminus \bigcup_{i \neq j} S_{ij}$ where

$$S_{ij} = \{z = [\theta^\intercal,\ x^\intercal]^\intercal \mid x_i = x_j, \theta_i = \theta_j\} \subset \mathbb{R}^{2Nd-2d}, \tag{22}$$

and is also locally essentially bounded. From (Danca, 2010, Remark 2), we obtain that the map $\mathcal{F}$ is upper semi-continuous with non-empty, compact and convex values for all $0 \leq t \leq T_\mathrm{p} + T_\mathrm{con}$.

Now, it holds that $\omega_i(t) = 0$ and $\tilde{u}_i(t) = 0$ for $i = \{1, 2, \ldots, N\}$, i.e., $F_1(\theta(t)) = F_3(x(t)) = 0$, for all $t \geq T_\mathrm{p} + T_\mathrm{con}$. Furthermore, for $t \geq T_\mathrm{p} + T_\mathrm{con}$, it holds that $g_i(\theta(t)) = g^*(x(t))$, and thus, $F_4(\theta(t)) = F_4(x(t))$. Hence, the augmented dynamics for $t \geq T_\mathrm{p} + T_\mathrm{con}$ reads:

$$\dot{\theta}(t) = F_2(x(t)), \quad \dot{x}(t) = F_4(x(t)). \tag{23}$$

Note that F_2 and F_4 are continuous functions in their arguments, and thus, the map $\mathcal{F}(z)$ the required conditions for all $t \geq T_\mathrm{p} + T_\mathrm{con}$, which completes the proof. □

Now, we are ready to present the main result of this section, which shows that when the closed-loop dynamics of (4) under $u = \tilde{u}_i + g_i$, written compactly as (21), is discretized using Euler discretization, the trajectories of the resulting discrete-time system reach an arbitrarily small neighborhood of the optimal point x^* within a fixed number of steps. To this end, define z^* as $z^* := \begin{bmatrix} I_N \otimes x^* \\ 0_{Nd} \end{bmatrix}$ where $\otimes$ denote the Kronecker product, $I_N \in \mathbb{R}^{N \times N}$ an identity matrix, and $0_{Nd} \in \mathbb{R}^{Nd}$ a vector consisting of 0s.

Theorem 2. *Assume that the functions f_i satisfy Assumption 2 (or Assumption 4) for all $i \in \mathcal{V}$ and let $p = q$, $\alpha = \gamma$, $\beta = \delta$, $l_1 = \mu_1 = \nu_1 = 1 + \frac{1}{\mu}$ and $l_2 = \mu_2 = \nu_2 = 1 - \frac{1}{\mu}$ for some $\mu > 1$. Consider the Euler discretization of (21) given by*

$$z_{k+1} \in z_k + \eta \mathcal{F}(z_k), \tag{24}$$

where $\eta > 0$. Then, for each $\epsilon > 0$, there exists $\eta^* > 0$ such that for all $\eta \in (0, \eta^*]$, the trajectories of (24) satisfy

$$\|z_k - z^*\| \leq \begin{cases} \dfrac{1}{\sqrt{c_1}} \left(\sqrt{\dfrac{a}{b}} \tan\left(\dfrac{\pi}{2} - \dfrac{\eta k \sqrt{ab}}{2\mu} \right) \right)^{\bar{\mu}} + \epsilon\,; & k \leq \dfrac{\mu \pi}{\sqrt{ab}\eta} \\ \epsilon\,; & \textit{otherwise}, \end{cases} \tag{25}$$

where $a, b, c_1, \mu > 0$.

The proof is provided in Appendix B. Thus, it is shown that the trajectories of the closed-loop dynamics (4) of each node i under the input (18), when discretized using Euler discretization, converge uniformly to an arbitrarily small neighborhood (dictated by ϵ) of the optimal point x^* within a fixed number of steps $\frac{\bar{\mu}}{2\sqrt{ab}\eta}$.

5. NUMERICAL VALIDATION

We validate the performance of the proposed DOA for distributed training of deep neural networks on the MNIST dataset. We assume a network of three servers connected in a line graph where each server has access to only one-third (20k) of the total (60k) training images. We consider a network with a single convolutional layer with ReLU activation (consisting of 32 filters of size 3×3), followed by a dense layer (with ReLU activation) of output size 128. The final linear layer transforms 128-dimensional input to a 10-dimensional output (corresponding to 10 classes) with SoftMax activation. The network comprises a total of 2.8×10^6 learnable parameters. The individual servers have their own local estimates of the neural network parameters. Figure 1 shows that the servers initialized with different parameters and having different test accuracies quickly converge to around 94% accuracy in less than ten epochs. Moreover, the norms of the consensus errors between servers i and j, denoted by $e_{ij} := x_i - x_j$, too, converge to zero, indicating that all the servers arrive at a similar estimate for all the neural network parameters. We also compare the performance of the proposed FxTS-DOA with the decentralized SGD (DSGD) (Koloskova et al., 2020) algorithm. As can be seen in Figure 1 for the DSGD method, even though the servers have better initial test accuracies to start with, the non-agreement between initial parameter estimates and large consensus errors eventually drives the cumulative test accuracy to $\sim$ 93.27%. Moreover, the servers achieve consensus on parameter estimates only after 20 epochs. On the other hand, the proposed FxTS-DOA trades off initial dip in test accuracies for super fast consensus on network parameters, eventually resulting in improved cumulative performance. Recall that the exchange of local estimates of parameters and gradients between any two neighbors occurs only once per epoch, i.e., the iteration complexity is only linear in the number of parameters (see lines 8-9 in Algorithm 1), resulting in significantly lower computational overhead.

Appendix A. PROOF OF LEMMA 8

Proof. The time derivative of θ_i is given by:

$$\dot{\theta}_i = p\sum_{j\in\mathcal{N}_i}\Big(\mathrm{sign}(\theta_j-\theta_i)+\gamma\mathrm{sign}^{\nu_1}(\theta_j-\theta_i) + \delta\mathrm{sign}(\theta_j-\theta_i)^{\nu_2}\Big)+h_i.$$

Define $\theta_{ji} \coloneqq \theta_j-\theta_i$ and $\theta_c \coloneqq \frac{1}{N}\sum_{j=1}^N\theta_j$, $i,j\in\mathcal{V}$. The difference between an agent i's state θ_i and the mean value θ_c of all agents' states is denote by $\tilde{\theta}_i \coloneqq \theta_i-\theta_c$. Similarly, $\tilde{\theta}_{ji} \coloneqq \tilde{\theta}_j-\tilde{\theta}_i$. The time-derivative of $\tilde{\theta}_i$ is given by:

$$\begin{aligned}\dot{\tilde{\theta}}_i &= \omega_i+h_i-\frac{1}{N}\sum_{j=1}^N\omega_j-\frac{1}{N}\sum_{j=1}^N h_j\\ &= \frac{1}{N}\sum_{j=1}^N(\omega_i-\omega_j)+\frac{1}{N}\sum_{j=1}^N(h_i-h_j) \qquad \text{(A.1)}\end{aligned}$$

Define the error vector $\tilde{\theta}=\begin{bmatrix}\tilde{\theta}_1 & \tilde{\theta}_2 & \cdots\tilde{\theta}_N\end{bmatrix}^\intercal$. Consider the candidate Lyapunov function defined as $V(\tilde{\theta})=\frac{1}{2}\|\tilde{\theta}\|^2=\frac{1}{2}\sum_{i=1}^N\tilde{\theta}_i^\intercal\tilde{\theta}_i$. Taking its time-derivative along the trajectories of (A.1) yields:

$$\dot{V}(\tilde{\theta}) = \underbrace{\frac{1}{N}\sum_{i,j=1}^N\tilde{\theta}_i^\intercal(\omega_i-\omega_j)}_{\dot{V}_1}+\underbrace{\frac{1}{N}\sum_{i,j=1}^N\tilde{\theta}_i^\intercal(h_i-h_j)}_{\dot{V}_2}. \qquad \text{(A.2)}$$

From (17), the first term $\dot{V}_1$ is rewritten as:

$$\begin{aligned}\dot{V}_1 &= \frac{1}{N}\sum_{i=1}^N\tilde{\theta}_i^\intercal\sum_{j=1}^N\Big(\omega_i-p\sum_{k\in\mathcal{N}_j}\Big(\mathrm{sign}(\theta_k-\theta_j) + \gamma\mathrm{sign}^{\nu_1}(\theta_k-\theta_j)+\delta\mathrm{sign}(\theta_k-\theta_j)^{\nu_2}\Big)\Big)\\ &\overset{(8)}{=} \frac{1}{N}\sum_{i=1}^N\tilde{\theta}_i^\intercal\sum_{j=1}^N\omega_i=\sum_{i=1}^N\tilde{\theta}_i^\intercal\omega_i=p\sum_{i=1}^N\tilde{\theta}_i^\intercal\sum_{j\in\mathcal{N}_i}(\mathrm{sign}(\theta_{ji})+ \gamma\mathrm{sign}^{\nu_1}(\theta_{ji})+\delta\mathrm{sign}(\theta_{ji})^{\nu_2})\\ &\overset{(9)}{=} \frac{p}{2}\sum_{i=1}^N\sum_{j\in\mathcal{N}_i}\tilde{\theta}_{ij}^\intercal(\mathrm{sign}(\theta_{ji})+\gamma\mathrm{sign}^{\nu_1}(\theta_{ji})+\delta\mathrm{sign}(\theta_{ji})^{\nu_2}),\end{aligned}$$

where the last equality follows with $w(x)=x$ in (9). Using this, and the fact that $\mathrm{sign}(\theta_{ij})^l=-\mathrm{sign}(\theta_{ji})^l$ for any odd $l\geq 0$, we obtain

$$\begin{aligned}\dot{V}_1 &= -\frac{p}{2}\sum_{i=1}^N\sum_{j\in\mathcal{N}_i}\tilde{\theta}_{ij}^\intercal(\mathrm{sign}(\theta_{ij})+\gamma\mathrm{sign}^{\nu_1}(\theta_{ij})+\delta\mathrm{sign}(\theta_{ij})^{\nu_2})\\ &= -\frac{p}{2}\sum_{i=1}^N\sum_{j\in\mathcal{N}_i}\left(\|\tilde{\theta}_{ij}\|+\gamma\|\tilde{\theta}_{ij}\|^{\nu_1+1}+\delta\|\tilde{\theta}_{ij}\|^{\nu_2+1}\right), \qquad \text{(A.3)}\end{aligned}$$

where the last equality follows from $\tilde{\theta}_{ij}=(\theta_i-\theta_c)-(\theta_j-\theta_c)=\theta_{ij}$. The second term in (A.2) can be bounded as:

$$\begin{aligned}\dot{V}_2 &= \frac{1}{2N}\sum_{i,j=1}^N\tilde{\theta}_{ij}^\intercal(h_i-h_j)\leq\frac{1}{2N}\sum_{i,j=1}^N\|\tilde{\theta}_{ij}\|\|h_i-h_j\|\\ &\leq \frac{\rho}{2N}\sum_{i,j=1}^N\|\tilde{\theta}_{ij}\|\leq\frac{\rho}{2N}\left(N\max_i\sum_{j=1,j\neq i}^N\|\tilde{\theta}_{ij}\|\right)\\ &\leq \frac{\rho}{2}\frac{(N-1)}{2}\sum_{i=1}^N\sum_{j\in\mathcal{N}_i}\|\tilde{\theta}_{ij}\|, \qquad \text{(A.4)}\end{aligned}$$

where the last inequality follows from connectivity of $\mathcal{G}$. Thus, from (A.3) and (A.4), it follows that

$$\begin{aligned}\dot{V}(\tilde{\theta}) &\leq -\frac{1}{2}\left(p-\rho\frac{(N-1)}{2}\right)\sum_{i=1}^N\sum_{j\in\mathcal{N}_i}\|\tilde{\theta}_{ij}\|\\ &\quad -\frac{1}{2}p\gamma\sum_{i=1}^N\sum_{j\in\mathcal{N}_i}\|\tilde{\theta}_{ij}\|^{\nu_1+1}-\frac{1}{2}p\delta\sum_{i=1}^N\sum_{j\in\mathcal{N}_i}\|\tilde{\theta}_{ij}\|^{\nu_2+1}\\ &\leq -\frac{1}{2}p\gamma\sum_{i=1}^N\sum_{j\in\mathcal{N}_i}\|\tilde{\theta}_{ij}\|^{\nu_1+1}-\frac{1}{2}p\delta\sum_{i=1}^N\sum_{j\in\mathcal{N}_i}\|\tilde{\theta}_{ij}\|^{\nu_2+1}\\ &\leq -\frac{p\gamma}{2}\sum_{i=1}^N\sum_{j=1}^N\left(a_{ij}\|\tilde{\theta}_{ij}\|^2\right)^{\kappa_1}-\frac{p\delta}{2}\sum_{i=1}^N\sum_{j=1}^N\left(a_{ij}\|\tilde{\theta}_{ij}\|^2\right)^{\kappa_2},\end{aligned}$$

where $\kappa_1=\frac{1+\nu_1}{2}$, $\kappa_2=\frac{1+\nu_2}{2}$. Define $\eta_{ij}=a_{ij}\|\tilde{\theta}_{ij}\|^2$. With this, and using the fact that $\nu_1>1$ and $\nu_2<1$, we obtain:

$$\begin{aligned}\dot{V}(\tilde{\theta}) &\leq -\frac{p\gamma}{2}\sum_{i,j=1}^N\eta_{ij}^{\kappa_1}-\frac{p\delta}{2}\sum_{i,=1}^N\eta_{ij}^{\kappa_2}\\ &\overset{(7)}{\leq} -\frac{p\gamma}{2}N^{2(1-\kappa_1)}\left(\sum_{i,j=1}^N\eta_{ij}\right)^{\kappa_1}-\frac{p\delta}{2}\left(\sum_{i,j=1}^N\eta_{ij}\right)^{\kappa_2}.\end{aligned}$$

We have
$$\begin{aligned}\sum_{i,j=1}^N\eta_{ij} &= \sum_{i,j=1}^N a_{ij}\|\tilde{\theta}_{ij}\|^2\\ &= 2\tilde{\theta}^\intercal L_A\otimes I_N\tilde{\theta}\geq 2\lambda_2(L_A\otimes I_N)\tilde{\theta}^\intercal\tilde{\theta}=cV,\end{aligned}$$

where $c=4\lambda_2(L_A)$. With this, we obtain that

$$\dot{V}(\tilde{\theta})\leq-\frac{p\gamma}{2}N^{2(1-\kappa_1)}c^{\kappa_1}V(\tilde{\theta})^{\kappa_1}-\frac{p\delta}{2}c^{\kappa_2}V(\tilde{\theta})^{\kappa_2}.$$

With $\nu_1>1$, we have $\kappa_1>1$, and with $\nu_2<1$, we have $\kappa_2<1$. Hence, using Lemma 1, we obtain that $V(\tilde{\theta}(t))=0$, i.e., $\theta_i(t)=\theta_c(t)$, for all $t\geq T_\mathrm{p}$, where $T_\mathrm{p}=\frac{2}{p\gamma N^{2(1-\kappa_1)}c^{\kappa_1}(\kappa_1-1)}+\frac{2}{p\delta c^{\kappa_2}(1-\kappa_2)}$. Using the fact that $\sum_{i=1}^N\omega_i(t)=0$ for all $t\geq 0$, we obtain that

$$\begin{aligned}&\sum_{i=1}^N\dot{\theta}_i(t)=\sum_{i=1}^N\omega_i(t)+\sum_{i=1}^N h_i(t)=\sum_{i=1}^N h_i(t)=\sum_{i=1}^N\frac{d}{dt}\zeta_i(t),\\ &\implies\sum_{i=1}^N\theta_i(t)=\sum_{i=1}^N\zeta_i(t)+c.\end{aligned}$$

With $\theta_i(0)=\zeta_i(0)\implies c=0$, completing the proof. □

Appendix B. PROOF OF THEOREM 2

Proof. First, consider the closed-loop dynamics (21) for $t\leq T_\mathrm{p}+T_\mathrm{con}$. From Lemma 9, it holds that the function $V_1(\theta)=\frac{1}{2}\|\tilde{\theta}\|^2$, where $\tilde{\theta}$ is as defined in Lemma 8 satisfies $\dot{V}_1(\theta(t))\leq-a_1V_1(\theta)^{\kappa_1}-a_2V_1(\theta)^{\kappa_2}$, where $a_1=\frac{p\gamma}{2}N^{2(1-\kappa_1)}c^{\kappa_1}, a_2=\frac{p\delta}{2}c^{\kappa_2}$, $\kappa_1=\frac{1+\nu_1}{2}>1$, $\kappa_2=\frac{1+\nu_2}{2}<1$ and $c=4\lambda_2(L_A)$. Similarly, since $p=q$, $\alpha=\gamma$, $\beta=\delta$, $\mu_1=\nu_1$ and $\mu_2=\nu_2$, the function $V_2(x)=\frac{1}{2}\|\tilde{x}\|^2$ where $\tilde{x}_i(t)=x_i(t)-\frac{1}{N}\sum_{j=1}^N x_j(t)$, satisfies $\dot{V}_2(x(t))\leq-a_1V_2(x)^{\kappa_1}-a_2V_2(x)^{\kappa_2}$. Now, define $z=[x^\top\ \theta^\top]^\top$ and $V(z(t))=V_1(\theta(t))+V_2(x(t))$, so that

$$\begin{aligned}\dot{V}(z(t)) &\leq -a_1V_1(\theta)^{\kappa_1}-a_2V_1(\theta)^{\kappa_2}-a_1V_2(x)^{\kappa_1}-a_2V_2(x)^{\kappa_2}\\ &= -a_1(V_1(\theta)^{\kappa_1}+V_2(x)^{\kappa_1})-a_2(V_1(\theta)^{\kappa_2}+V_2(x)^{\kappa_2}),\end{aligned}$$

for all $t\leq T_\mathrm{p}+T_\mathrm{con}$. Now, using Lemma 2, it holds that $V_1(\theta)^{\kappa_1}+V_2(x)^{\kappa_1}\geq 2^{1-\kappa_1}(V_1(\theta)+V_2(x))^{\kappa_1}=2^{1-\kappa_1}V(z)^{\kappa_1}$ and $V_1(\theta)^{\kappa_2}+V_2(x)^{\kappa_2}\geq(V_1(\theta)+V_2(x))^{\kappa_2}=V(z)^{\kappa_2}$. Thus, it holds that $\dot{V}(z(t))\leq-a_12^{1-\kappa_1}V(z(t))^{\kappa_1}-a_2V(z(t))^{\kappa_2}$ for all $t\leq T_\mathrm{p}+T_\mathrm{con}$. Hence, $z=\bar{z}$ is an FxTS equilibrium point of (21) where $\bar{z} = \begin{bmatrix} I_N \otimes \bar{\theta} \\ I_N \otimes \bar{x} \end{bmatrix}$ with $\bar{\theta} = \frac{1}{N}\sum_{j=1}^N \theta_j$ and $\bar{x} = \frac{1}{N}\sum_{j=1}^N x_j$. Note also that V_1 and V_2 are quadratic in the error $\tilde{\theta}$ and $\tilde{x}$, respectively. Hence, it holds that $V(z(t)) = \frac{1}{2}\|\tilde{\theta}\|^2 + \frac{1}{2}\|\tilde{x}\|^2 \leq \frac{1}{2}\|\tilde{z}\|^2$ where $\tilde{z} = z - z = \begin{bmatrix} \tilde{\theta} \\ \tilde{x} \end{bmatrix}$. Furthermore, it holds that $V(z(t)) = \frac{1}{2}\|\tilde{\theta}(t)\|^2 + \frac{1}{2}\|\tilde{x}(t)\|^2 \geq \frac{1}{4}\|\tilde{z}(t)\|^2$. Now, consider the time interval $t \geq T_{\rm p} + T_{\rm con}$. The dynamics for z reads $\dot{z}(t) = F(z(t)) = \begin{bmatrix} F_2(x(t)) \\ F_4(x(t)) \end{bmatrix}$, for $t \geq T_{\rm p} + T_{\rm con}$. Consider the Lyapunov candidate $V_3(z(t)) = \frac{1}{2}\sum_{i=1}^N \|\nabla f_i(x(t))\|^2 + \frac{1}{2}\|\tilde{\theta}(t)\|^2$. Note that from Lemma 9, it follows that $\tilde{\theta}(t) = 0$ for $t \geq T_{\rm p}$. If f_i satisfies Assumption 2, then from Lemma 6, it follows that the time derivative of V along the trajectories of z for $t \geq \max\{T_1, T_2\}$ reads

$$\begin{aligned}\dot{V}_3(z(t)) &\leq -k2^{\frac{1+l_1}{2}} V_3(z(t))^{\frac{1+l_1}{2}} - k2^{\frac{1+l_2}{2}} V_3(z(t))^{\frac{1+l_2}{2}} \\ &= -k2^{\kappa_1} V_3(z(t))^{\kappa_1} - k2^{\kappa_2} V_3(z(t))^{\kappa_2},\end{aligned}$$

since $l_1 = \mu_1$ and $l_2 = \mu_2$. Note that from (Karimi et al., 2016, Theorem 2), it follows that under strong convexity implies quadratic growth, and thus, we obtain that the function V satisfies the quadratic growth requirement in (Garg et al., 2022, Theorem 3). If, on the other hand, f_i satisfies Assumption 4, then from Lemma 7, it follows that the time derivative of V along the trajectories of z for $t \geq \max\{T_1, T_2\}$ reads

$$\begin{aligned}\dot{V}_3(z(t)) &\leq -k_1^{\frac{1+l_1}{2}} V_3(z(t))^{\frac{1+l_1}{2}} - k_2^{\frac{1+l_2}{2}} V_3(z(t))^{\frac{1+l_2}{2}} \\ &\leq -k_1^{\kappa_2} V_3(z(t))^{\kappa_1} - k_2^{\kappa_2} V_3(z(t))^{\kappa_2}\end{aligned}$$

where $k_1 = (2\mu)^{\frac{1+l_1}{2}} 2^{\frac{1+l_1}{2}}$ and $k_2 = (2\mu)^{\frac{1+l_2}{2}} 2^{\frac{1+l_2}{2}}$. Choose $a = \min\{a_1 2^{1-\kappa_1}, k2^{\kappa_1}, k_1^{\kappa_1}\}$ and $b = \min\{a_2, k2^{\kappa_2}, k_2^{\kappa_2}\}$, so that it holds that $\dot{V}(z(t)) \leq -aV(z(t))^{\kappa_1} - bV(z(t))^{\kappa_2}$ for all $t \geq 0$. In this case as well, since the system trajectories evolve in a compact set $\{z \mid V(z) \leq V(z(0))\}$, from (Karimi et al., 2016, Theorem 2), it follows that the function V satisfies the quadratic growth requirement in (Garg et al., 2022, Theorem 3). Thus, all the conditions of (Garg et al., 2022, Theorem 3) are satisfied with $\beta = \frac{1}{2}$, and hence, (25) holds. □

REFERENCES

Baranwal, M., Garg, K., Panagou, D., and Hero, A.O. (2020). Robust distributed fixed-time economic dispatch under time-varying topology. *IEEE Control Systems Letters*, 5(4), 1183–1188.

Bhat, S.P. and Bernstein, D.S. (2000). Finite-time stability of continuous autonomous systems. *SICON*, 38(3), 751–766.

Clarke, F.H., Ledyaev, Y.S., Stern, R.J., and Wolenski, P.R. (2008). *Nonsmooth analysis and control theory*, volume 178. Springer Science & Business Media.

Danca, M.F. (2010). Chaotic behavior of a class of discontinuous dynamical systems of fractional-order. *Nonlinear Dynamics*, 60(4), 525–534.

Feng, Z. and Hu, G. (2017). Finite-time distributed optimization with quadratic objective functions under uncertain information. In *IEEE 56th Annual Conference on Decision and Control*, 208–213. IEEE.

Garg, K., Baranwal, M., Gupta, R., and Benosman, M. (2022). Fixed-time stable proximal dynamical system for solving MVIPs. doi:10.1109/TAC.2022.3214795.

Garg, K., Baranwal, M., and Panagou, D. (2020). A fixed-time convergent distributed algorithm for strongly convex functions in a time-varying network. In *2020 59th IEEE Conference on Decision and Control (CDC)*, 4405–4410. IEEE.

Garg, K. and Panagou, D. (2020). Fixed-time stable gradient flows: Applications to continuous-time optimization. *IEEE Transactions on Automatic Control*, 66(5), 2002–2015.

Hu, Q. and Shao, X. (2016). Smooth finite-time fault-tolerant attitude tracking control for rigid spacecraft. *Aerospace Science and Technology*, 55, 144–157.

Hu, Z. and Yang, J. (2018). Distributed finite-time optimization for second order continuous-time multiple agents systems with time-varying cost function. *Neurocomputing*, 287, 173–184.

Karimi, H., Nutini, J., and Schmidt, M. (2016). Linear convergence of gradient and proximal-gradient methods under the Polyak- łojasiewicz condition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 795–811. Springer.

Koloskova, A., Loizou, N., Boreiri, S., Jaggi, M., and Stich, S. (2020). A unified theory of decentralized sgd with changing topology and local updates. In *International Conference on Machine Learning*, 5381–5393. PMLR.

Lin, P., Ren, W., and Farrell, J.A. (2017). Distributed continuous-time optimization: nonuniform gradient gains, finite-time convergence, and convex constraint set. *IEEE Transactions on Automatic Control*, 62(5), 2239–2253.

Lu, J. and Tang, C.Y. (2012). Zero-gradient-sum algorithms for distributed convex optimization: The continuous-time case. *IEEE Transactions on Automatic Control*, 57(9), 2348–2354.

Mesbahi, M. and Egerstedt, M. (2010). *Graph Theoretic Methods in Multiagent Networks*, volume 33. Princeton University Press.

Nathan, A. and Klabjan, D. (2017). Optimization for large-scale machine learning with distributed features and observations. In *International Conference on Machine Learning and Data Mining in Pattern Recognition*, 132–146. Springer.

Pan, X., Liu, Z., and Chen, Z. (2018). Distributed optimization with finite-time convergence via discontinuous dynamics. In *2018 37th Chinese Control Conference (CCC)*, 6665–6669. IEEE.

Polyakov, A. (2012). Nonlinear feedback design for fixed-time stabilization of linear control systems. *IEEE Transactions on Automatic Control*, 57(8), 2106.

Polyakov, A., Efimov, D., and Brogliato, B. (2019). Consistent discretization of finite-time and fixed-time stable systems. *SIAM Journal on Control and Optimization*, 57(1), 78–103.

Wang, X., Wang, G., and Li, S. (2020). A distributed fixed-time optimization algorithm for multi-agent systems. *Automatica*, 122, 109289.

Zuo, Z. and Tie, L. (2016). Distributed robust finite-time nonlinear consensus protocols for multi-agent systems. *International Journal of Systems Science*, 47(6), 1366–1375.