

Finite-Time Convergent Algorithms for Time-Varying Distributed Optimization

Xinli Shi, Guanghui Wen, and Xinghuo Yu

Abstract—This paper focuses on finite-time (FT) convergent distributed algorithms for solving time-varying (TV) distributed optimization (TVDO). The objective is to minimize the sum of local TV cost functions subject to the possible TV constraints by the coordination of multiple agents in finite time. Specifically, two classes of TVDO are investigated including unconstrained distributed consensus optimization and distributed optimal resource allocation problems (DORAP) with both TV cost functions and coupled equation constraints. For the previous one, based on non-smooth analysis, a continuous-time distributed discontinuous dynamics with FT convergence is proposed based on an extended zero-gradient-sum method with a local auxiliary subsystem. Then, an FT convergent distributed dynamics is further obtained for TV-DORAP by dual transformation. Particularly, the inversion of the cost functions' Hessians is not required in the dual variables' dynamics, while another local optimization needs to be solved to obtain the primal variable at each time instant. Finally, two numerical examples are conducted to verify the proposed algorithms.

Index Terms—Finite-time convergence, time-varying distributed optimization, resource allocation, discontinuous dynamics.

I. INTRODUCTION

Recently, time-varying distributed optimization (TVDO) has received increasing attention as it is more practical than time-invariant distributed optimization (TIDO) in a dynamic environment where the objective function or constraints can change over the time. TVDO has found applications in many areas, such as power system, machine learning and robotics, see [8] and references therein. In TVDO, the optimal solution can be TV and hence the traditional algorithms designed for TIDO to approach a static optimizer can not be applied directly. To solve TVDO, many discrete-time algorithms (DTAs) have been provided for solving TVDO [2]–[6]. For example, prediction-correction methods are used in [3], [4] for solving TVDO with sample period, and the tracking error bound is related to the size of sample period. More DTAs can be found in the survey [7] and literature therein. However, in general, it is difficult for CTAs to track the TV optimal trajectory asymptotically due to the sample period, step size or errors in the local optimization.

To track the optimal solution trajectory with vanishing errors, several continuous-time algorithms (CTAs) have been proposed for solving TV (distributed) optimization [8]–[12]. Compared with DTAs, typical CTAs can achieve fast convergence performance, such as finite-time (FT) convergence to the optima, and further inspire the discrete-time counterpart design. Besides, CTAs are also preferred for optimizing the swarm tracking behavior of a multi-robotic system with physical dynamics [8] [9].

X. Shi is with the School of Cyber Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: xinli_shi@seu.edu.cn).

G. Wen is with the the Jiangsu Provincial Key Laboratory of Networked Collective Intelligence, School of Mathematics, Southeast University, Nanjing 210096, China (e-mail: wenguanghui@gmail.com).

X. Yu is with the School of Engineering, RMIT University, Melbourne, VIC 3001, Australia (e-mail: x.yu@rmit.edu.au).

Moreover, CTAs can track the TV optimal solution with vanishing errors. Since CTAs exhibit such good tracking performance for TV problems, we design a CTA, with the hope that even when discretized it will show improved performance. Thus, in this work, we mainly focus on CTAs for solving TVDO, including TV distributed consensus optimization (TVDCO) and optimal resource allocation problems (TV-DORAP). For TVDCO, in [9], both single- and double-integrator dynamics are investigated for distributed consensus optimization with TV cost functions, which can achieve asymptotic convergence to the optimal trajectory under some strict assumptions such as identical Hessians. Furthermore, TVDCO is considered for a class of nonlinear multi-agent system in [10]. In [11], an edge-based distributed protocol is provided for solving TVDCO subject to identical Hessians. In [12], gradient-based searching methods are used to track the TV solution of TVDCO with quadratic objective functions.

Different from the distributed consensus optimization aiming to achieve multi-agent consensus on the optimal solution, in DORAP with coupled constraints, the optimal resource allocation of agents can be heterogeneous but should satisfy the total demand while minimizing the sum of local cost functions. For time-invariant DORAP, extensive CTAs have been proposed to approach the static optimal solution with agents' coordination [13], [14]. As variations always exist in the generations of renewable resources and the noncontrollable load demand in a dynamic power system [8], TV cost functions and constraints are more practical in the real-time DORAP. In [15], a robust distributed algorithm is designed for economic dispatch and then the TV loads are discussed, under which the power mismatch is only shown to be ultimately bounded. In [16], the prediction–correction method combined with the discontinuous consensus protocol is used for DORAP with TV quadratic cost functions of identical Hessians, and then distributed average tracking (DAT) estimator-based methods are used to deal with the case with nonidentical constant Hessians. Moreover, Wang et al. provide two dynamics for handling DORAP with quadratic cost functions subject to identical and nonidentical Hessians in [17] and [18], respectively, as well as TV resources.

Although some CTAs have been proposed for solving TVDO including TVDCO and DORAP, there exist several issues required to be further addressed. One is that most CTAs can only achieve an asymptotic convergence to the optimal trajectory, which means that tracking errors tend to zero as the time goes to infinity. The second is that almost all the CTAs with simple structures require that the Hessians of all local cost functions are identical, and for DORAP, the cost functions are of the quadratic form, which limit the applications of the existing methods. To deal with nonidentical Hessians, a non-smooth estimator based on DAT methods can be introduced to track the global information, but it suffers from high gains and more

computation/communication costs. To tackle these two issues, the task of this note is to design FT convergent CTAs for solving TVDO without identical Hessians. That is to say, the provided dynamics will track the optimal trajectory in finite time without mismatch errors. In literature, the existing distributed FT algorithms are designed for time-invariant unconstrained optimization [20]. For example, an FT convergent algorithm is proposed in [20] based on the continuous-time zero-gradient-sum (ZGS) method [21], where the initial states should be the minimizers of the local cost functions. Recently, an FT convergent primal-dual method has been proposed in [22] and further applied to constrained distributed optimization including DORAP. In [23], a DAT estimator for tracking global information is used to design FT dynamics for TVDO with TV cost functions.

As pointed out in [8], the best algorithm for the invariant case may be the worst one in the dynamic case. Similarly, the existing FT CTAs designed for TIDO might not be applied to TVDO directly, and there are few results on the FT convergent algorithms with simple structures for TVDO. In this note, we will provide two FT distributed algorithms for solving TVDO such as TVDCO and DORAP, respectively. The contributions are listed as follows.

- 1) For a class of TVDCO with strongly convex and smooth cost functions, a discontinuous distributed dynamics with FT convergence is proposed, based on an extended ZGS approach. Specifically, the introduced auxiliary dynamics can drive all the local states to an invariant set in finite time, on which the sum of local gradients is zero. After that, TVDCO will be solved once the multi-agent system reaches consensus in finite time by using non-smooth consensus protocol. Different from existing ZGS method based algorithms (e.g., [20], [21]) for TIDO, the agents' initial states are not required to minimize the local functions (see [20], [21]). Moreover, compared with exiting methods [9]–[11], [23], the provided algorithm has a simpler structure without estimating the global information and can be used for TVDO with nonidentical Hessians. However, similar to [9]–[11], [23], the inversion of Hessian is required in the proposed method.
- 2) An FT convergent distributed dynamics is further obtained for DORAP with TV cost functions and coupled constraints by dual transformation. Compared with existing works [16]–[18], the cost functions can be of non-quadratic form with nonidentical Hessians. Moreover, the inversion of the cost functions' Hessians is not required in the dual variables' dynamics. However, another local optimization needs to be solved to obtain the primal variable at each time instant.
- 3) In the proposed CTAs, only binary information is required and the neighboring agents only need to know whether their relative position is positive or negative, which benefits the online implementation in a coarse sensing scenario [24]–[26]. Despite such a coarse information, the FT convergence to the exact TV optimal trajectory is guaranteed.

The rest of this paper is organized as follows. Section II introduces some preliminary notations and concepts. In Section III, the problem statements and main methodologies are provided. Finally, two numerical examples on TVDCO and TV DORAP are presented in Section IV to verify the proposed algorithms. Conclusions are drawn in Section V.

II. PRELIMINARIES

A. Notation and Network Representation

\mathbb{R}_+^n denotes the set of n -dimensional vectors with non-negative entries and $I_n \in \mathbb{R}^{n \times n}$ is an identity matrix. Let $\langle n \rangle = \{1, 2, \dots, n\}$. For $x = [x_1, \dots, x_n]^T \in \mathbb{R}^n$, we write the p -norm of x as $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$ and $\text{sgn}^\alpha(x) = [\text{sign}(x_1)|x_1|^\alpha, \dots, \text{sign}(x_n)|x_n|^\alpha]^T$ for $\alpha \in \mathbb{R}_+$. By default, we denote $\|x\|$ as the Euclidean norm of x . For the matrix $M \in \mathbb{R}^{n \times n}$, $\mathcal{N}(M)$ denotes the null space of M and $\Pi_{\mathcal{N}(M)}(x)$ denotes the projection of x onto $\mathcal{N}(M)$. When M is positive semidefinite, $\lambda_2(M)$ denotes its smallest positive eigenvalue.

A continuously differentiable function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ is called μ -strongly convex if for any $x, y \in \mathbb{R}^n$, $f(y) - f(x) - \nabla f(x)^T(y - x) \geq \frac{\mu}{2}\|y - x\|^2$. Moreover, f is said to be θ -smooth if for any $x, y \in \mathbb{R}^n$, $\|\nabla f(y) - \nabla f(x)\| \leq \theta\|y - x\|$. For a convex function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$, its Legendre–Fenchel conjugate f^* is defined by

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \{\langle y, x \rangle - f(x)\}, \quad \forall y \in \mathbb{R}^n. \quad (1)$$

The following result can be found in [27, Proposition 12.60].

Lemma 1: For a proper, lower semicontinuous and convex function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$, f^* is $\frac{1}{\sigma}$ -strongly convex iff f is differentiable and σ -smooth.

A weighted undirected graph is represented by $\mathcal{G}(\mathcal{V}, \mathcal{E}, A)$, where $\mathcal{V} = \langle N \rangle$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denote the node set and edge set, respectively, and the matrix $A = [a_{ij}]_{N \times N}$ is the weighted adjacency matrix with $a_{ij} > 0$ iff $(j, i) \in \mathcal{E}$ and $a_{ii} = 0, \forall i \in \langle N \rangle$. Let \mathcal{N}_i denote the set of neighbors of agent i .

B. Finite-Time Stability

Consider the following differential autonomous system

$$\dot{x}(t) = f(x(t)), x(t_0) = x_0, \quad (2)$$

with $x \in \mathbb{R}^n$. When the right-hand side function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is discontinuous at some points, its Filippov solution will be investigated, which is an absolutely continuous map $x : I \rightarrow \mathbb{R}^n$ defined on an interval $I \subset \mathbb{R}$ satisfying the differential inclusion:

$$\dot{x}(t) \in \mathcal{F}[f](x(t)) \quad (3)$$

with $\mathcal{F}[f](\cdot)$ being the Filippov set-valued map [28]. As known, the Filippov solution to (2) always exists if f is measurable and locally essentially bounded. The *set-valued Lie derivative* of a locally Lipschitz continuous map $V : \mathbb{R}^n \rightarrow \mathbb{R}$ associated with $\mathcal{F}[f]$ at x is defined as

$$\tilde{\mathcal{L}}_{\mathcal{F}[f]}V(x) \triangleq \{a \in \mathbb{R} \mid \exists \zeta \in \mathcal{F}[f](x) \Rightarrow \zeta^T \nu = a, \forall \nu \in \partial V(x)\},$$

in which $\partial V(x)$ represents the generalized gradient of V [28].

The FT/FxT stability of the system (2) is given in Definition 1. Lemma 2 is helpful for the non-smooth analysis when dealing with Filippov solutions to (2).

Definition 1: The system (2) is *FT stable* at the origin if it is Lyapunov stable and there exists a *finite settling time* $T(x_0)$ such that

$$\lim_{t \rightarrow T(x_0)} x(t; x_0) = 0 \text{ and } x(t) = 0, \forall t \geq T(x_0)$$

for any solution $x(t; x_0)$ starting from $x_0 \in \mathcal{X}_0$. When $\mathcal{X}_0 = \mathbb{R}^n$, the system (2) is called globally FT stable.

Lemma 2: [28] Suppose that $V : \mathbb{R}^n \rightarrow \mathbb{R}$ is locally Lipschitz continuous and regular, and $x(t) : I \subset \mathbb{R} \rightarrow \mathbb{R}^n$ is a Filippov solution satisfying (3). Then, $V(x(t))$ is absolutely continuous with

$$\dot{V}(x(t)) \stackrel{a.e.}{\in} \tilde{\mathcal{L}}_{\mathcal{F}[f]} V(x(t)), \quad t \in I. \quad (4)$$

III. PROBLEM STATEMENT AND METHODOLOGY

In this section, we first provide a unified framework for designing FT convergent algorithms to solve TV centralized optimization. Then, based on an extended ZGS method, two FT convergent distributed algorithms will be designed for TVDCO and DORAP over a networked system, respectively. Meanwhile, the comparisons with existing works are discussed in remarks.

A. FT Convergent TV Optimization

First, we focus on the following TV convex optimization

$$\min_{x \in \mathbb{R}^n} f_0(x, t), \quad (5)$$

where the objective function $f_0 : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}$ is convex and further satisfies Assumption 1.

Assumption 1: The function $f_0(x, t)$ is twice continuously differentiable with invertible Hessian $H_0(x, t)$ and there exists a continuous trajectory $x^*(t)$ that solves (5).

Then, we aim to provide a continuous-time dynamics $\dot{x} = u(x, t)$ such that $x(t)$ tracks $x^*(t)$ in FT $T < +\infty$, i.e.,

$$\lim_{t \rightarrow T} \|x(t) - x^*(t)\| = 0. \quad (6)$$

Then, to achieve (6), the proposed dynamics is designed as

$$\begin{cases} \dot{x} = -H_0^{-1}(x, t)(\varphi(z) + \frac{\partial \nabla f(x, t)}{\partial t}), \\ \dot{z} = -\varphi(z), \end{cases} \quad (7a)$$

$$(7b)$$

with $z(0) = \nabla f_0(x(0), 0)$, where the function $\varphi(z)$ is chosen such that the origin is FT/FxT stable for the subsystem (7b). Several typical functions can be chosen, e.g.,

$$\varphi(z) = a \cdot z / \|z\|_r^p + b \cdot z \|z\|_r^{q-1} \text{ or } a \cdot \text{sgn}^{1-p}(z) + b \cdot \text{sgn}^q(z)$$

with $p \in (0, 1]$, $r \geq 1$, $q > 1$, for which the FT/FxT stability of (7b) can be shown with the Lyapunov function $V(z) = \frac{1}{2}z^T z$ based on [30, Lemma 1]. Note that the subsystem (7b) is only related to $z(t)$ and thus it can be freely designed with prescribed performance. In this work, we consider a class of FT stable subsystem (7b). Then, we have the following result.

Theorem 1: Suppose that Assumption 1 holds and (7b) is FT stable at origin with settling time T_{\max} . Then, with (7), $x(t)$ will track $x^*(t)$ in the same FT T_{\max} .

Proof. See Appendix A. \square

Remark 1: In fact, from the proof of Theorem 1, (7) is equivalent to the following dynamics

$$\dot{x} = -H_0^{-1}(x, t)(\varphi(\nabla f(x, t)) + \frac{\partial \nabla f(x, t)}{\partial t}). \quad (8)$$

The introduced variable $z(t)$ is used to estimate $\nabla f(x, t)$ when the initialization coincides. The auxiliary system (7b) will be helpful for designing FT distributed algorithms to solve TVDO in the next part.

B. FT Algorithm for Solving TVDCO

Consider an interactive network $\mathcal{G}(\mathcal{V}, \mathcal{E}, A)$ consisting of N agents, each of which is equipped with a local TV cost function. Then, the multi-agent system aims to solve the following TVDO by coordinating with neighbors.

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^N f_i(x, t), \quad (9)$$

where $f_i : \mathbb{R}^n \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is the local TV cost function only known by the agent i . Moreover, f_i is supposed to satisfy Assumptions 2 and 3, where $\nabla f_i(x, t)$ is the gradient of f_i w.r.t. x . Under Assumption 2, it can be deduced that there exists a unique continuous trajectory $x^*(t)$ that solves TVDO (9). Besides, Assumption 3 covers a group of functions such as $f_i(x, t) = f_i^0(x) + g(t)x$ when $\dot{g}(t)$ is globally bounded. For example, in smart grid, $f_i(x, t)$ can be the cost of a generator with x and $-g(t)$ representing the generation power and the TV electricity price, respectively. The objective of this work is to design a distributed CTA with only local information that tracks $x^*(t)$ in finite time.

Assumption 2: For each local cost function $f_i, i \in \langle N \rangle$, f_i is twice continuously differentiable w.r.t. x , θ_i -strongly convex and θ_i -smooth for some positive scalars $\theta_i, \bar{\theta}_i$, i.e., $\theta_i \mathbf{I}_n \leq H_i(x, t) \leq \bar{\theta}_i \mathbf{I}_n$, where $H_i(x, t)$ is the Hessian matrix of f_i .

Assumption 3: For each $i \in \langle N \rangle$, $\frac{\partial \nabla f_i(x, t)}{\partial t}$ is measurable w.r.t. t and bounded by $\|\frac{\partial \nabla f_i(x, t)}{\partial t}\| \leq \kappa$ for some $\kappa > 0$.

Assumption 4: The interactive graph \mathcal{G} is undirected and connected.

Let x_i be the local copy of the system decision x at agent i . Then, under Assumption 4, (9) can be reformulated into the following equivalent distributed optimization over the network $\mathcal{G}(\mathcal{V}, \mathcal{E})$.

$$\min_{x_i \in \mathbb{R}^n} \sum_{i=1}^N f_i(x_i, t), \text{ s.t. } x_i = x_j, \forall (i, j) \in \mathcal{E}. \quad (10)$$

Then, the purpose of this subsection is to design distributed dynamics which guarantees that there exists a finite time $T < \infty$ such that

$$\lim_{t \rightarrow T} \|x_i(t) - x^*(t)\| = 0, \forall i \in \langle N \rangle. \quad (11)$$

In order to track the global solution $x^*(t)$ in finite time, we propose the following distributed discontinuous dynamics as an extension of (7)

$$\begin{cases} \dot{x}_i = -H_i^{-1}(x_i, t)(\varphi_i(z_i) + \frac{\partial \nabla f_i(x_i, t)}{\partial t} + \alpha \phi_i(x_i, x_{\mathcal{N}_i})), \\ \dot{z}_i = -\varphi_i(z_i), \quad i \in \langle N \rangle \end{cases} \quad (12a)$$

$$(12b)$$

with $z_i(0) = \nabla f_i(x_i(0), 0)$, where $\phi_i(x_i, x_{\mathcal{N}_i}) = \sum_{j=1}^N a_{ij} \text{sgn}(x_i - x_j)$ with $x_{\mathcal{N}_i} \triangleq [x_j]_{j \in \mathcal{N}_i}$. Note that under Assumptions 2, 3 and 5, the right-hand side of (12) is bounded over the time and thus its Filippov solution always exists. We call (12) an extended ZGS method as the initial state $x_i(0)$ is not required to be the local minimizer of local cost function to be distinct from [21]. However, with the introduced auxiliary subsystem (12b), the local states $x_i(t)$ will be driven to an invariant set where the sum of local TV gradients is zero. Particularly, when the solution to (12b) can be explicitly given as $z_i(t) = \Phi_i(t, z_0)$, $\varphi_i(z_i)$ in (12a) can be simply replaced by $\varphi_i(\Phi_i(t, z_0))$ and then (12b) can be removed.

Remark 2: For (12), it might be restrictive that each agent knows the closed form of $\partial \nabla f_i(x_i, t)/\partial t$. However, in some applications, the local time-varying cost function is available for each agent, such as a motion control with an optimization objective, or in smart grid where smart devices should coordinate with each other to maximize the overall utility function with a TV electricity price whose rate is known beforehand [9]. The term $\partial \nabla f_i(x_i, t)/\partial t$ here is used for prediction to enhance the exact convergence to $x^*(t)$, which has been used in [4], [9], [10], [16], [18]. As pointed in [8], without the predictor, a tracking error exists, possibly depending on the variation of the gradient with the time and the control gain. Besides, the discontinuous term ϕ_i is used for achieving consensus among agents as well as constraining $\partial \nabla f_i(x_i, t)/\partial t$. Sometimes, the noise in the channels can affect the performance of the algorithms [24], [25]. For ϕ_i , as only the binary information that represents whether the neighboring agents' relative position is positive or negative is used, the presence of small noise in a coarse sensing channel may not affect the performance of algorithms when $\text{sign}(x_i - x_j)$ (i.e., $\{\pm 1, 0\}$) remains the same. Detailed properties of the binary protocols can be found in [26].

Assumption 5: The subsystem (12b) is FT stable at the origin, $\forall i \in \langle N \rangle$.

In the next, we will show that $\sum_{i=1}^N z_i(t)$ as an estimation of $\sum_{i=1}^N \nabla f_i(x_i, t)$ will converge to zero in finite time and then all $x_i(t)$ will reach consensus on $x^*(t)$ in finite time due to ϕ_i in (12a).

Given a weighted adjacency matrix $A = [a_{ij}]_{N \times N}$ of the undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with $m = |\mathcal{E}|$, we define a weighted incidence matrix $B_0 = [b_{ik}]_{N \times m}$: $b_{ik} = -b_{jk} = a_{ij}$ if $i < j$ for any edge $e_k = \{i, j\} \in \mathcal{E}$. Let $\mathbf{x} = [x_i]_{i \in \langle N \rangle}$, $\delta = [\delta_{ij}]_{(i,j) \in \mathcal{E}}$, $\mathbf{z} = [z_i]_{i \in \langle N \rangle}$, $B = B_0 \otimes I_n$, $f(\mathbf{x}, t) = \sum_{i=1}^N f_i(x_i, t)$, $H = \text{blkdiag}(H_1(x_1, t), \dots, H_N(x_N, t))$ and $\varphi(\mathbf{z}) = [\varphi_i(z_i)]_{i \in \langle N \rangle}$. Then, (12) can be rewritten in a compact form:

$$\begin{cases} \dot{\mathbf{x}} = -H^{-1}(\mathbf{x}, t)(\varphi(\mathbf{z}) + \frac{\partial \nabla f(\mathbf{x}, t)}{\partial t} + \alpha B \text{sgn}(B^T \mathbf{x})), & (13a) \\ \dot{\mathbf{z}} = -\varphi(\mathbf{z}), & (13b) \end{cases}$$

with $\mathbf{z}(0) = \nabla f(\mathbf{x}(0), 0) = [\nabla f_i(x_i(0), 0)]_{i \in \langle N \rangle}$. Then, based on the non-smooth analysis, the FT convergence of (13) is established in the following result.

Theorem 2: Let Assumptions 2-5 hold, $\underline{\theta} = \min_{i \in \langle N \rangle} \{\theta_i\}$ and $\bar{\theta} = \max_{i \in \langle N \rangle} \{\bar{\theta}_i\}$. Under the distributed protocol (12) with $\alpha > \kappa \sqrt{\frac{N\bar{\theta}}{\underline{\theta} \lambda_2(B^T B)}}$, $z_i(t)$ will converge to 0 in finite time and then $x_i(t)$ will track the optimal trajectory $x^*(t)$ to TVDO (10) in finite time.

Proof. See Appendix B. \square

Actually, by the Eq. (26), the strong convexity assumption and Assumption 3 can be further relaxed as Assumption 6, which is more general than [9, Ass. 3.8], [10, Ass. 4], [11, Ass. 2] and [4]. It can be easily verified that when f_i is strongly convex and $\|\partial \nabla f_i(x_i, t)/\partial t\|$ is bounded, Assumption 6 holds directly. With the alternative Assumption 6, we have the following result.

Assumption 6: For any $i, j \in \langle N \rangle$, there exists $\varpi > 0$ such that $\|H_i^{-1}(x_i, t) \frac{\partial \nabla f_i(x_i, t)}{\partial t} - H_j^{-1}(x_j, t) \frac{\partial \nabla f_j(x_j, t)}{\partial t}\|_1 \leq \varpi$.

Proposition 1: Let Assumptions 4-6 hold. Suppose that $f_i(x_i, t)$ has invertible Hessian and is θ_i -smooth. Let $\bar{\theta} = \max_{i \in \langle N \rangle} \{\bar{\theta}_i\}$ and $\bar{a} = \max_{(i,j) \in \mathcal{E}} \{a_{ij}\}$. Under the distributed protocol (12) with $\alpha > \frac{m\varpi\bar{a}\bar{\theta}}{\lambda_2(B^T B)}$, $z_i(t)$ will converge to 0 in finite time and then $x_i(t)$ will track the optimal trajectory $x^*(t)$ to TVDO (10) in finite time.

Proof. See Appendix C. \square

Remark 3: In [4], a dual prediction–correction DTA is developed for solving equation constrained TV optimization and then its distributed implementation is given by using constraint matrix (i.e., the incidence/Laplacian matrix) to ensure the local consensus. In [9], an asymptotically convergent CTAs is provided for solving (10) subject to TV cost functions with identical Hessians, based on an FT DAT dynamics to track the global information. The work [9] first makes all the states achieve consensus in finite time and then the TV optimization is solved by a centralized Newton–Raphson (NR) dynamics, i.e., $\dot{x} = -(\nabla^2 f(x))^{-1} \nabla f(x)$, which we call “consensus+central NR” dynamics. Differently, non-identical Hessians are allowable in the proposed ZGS-based algorithm (12) and the state consensus term is inside the local NR dynamics, which we call “consensus within local NR” dynamics. The advantage of the ZGS-based algorithm lies in that the TVDCO is solved once the consensus of local states is reached as the ZGS manifold \mathcal{M}_0 is invariant. By the convergence analysis, one can only obtain an upper bound of the finite settling time applicable for all considered TVDCO in worst case as most results for FT algorithms do. For specific problems, such as DAT system, the derived control gain condition in Theorem 2 is less conservative than that (i.e., $\alpha > \kappa \sqrt{2N/\lambda_2(B^T B)}$) obtained in [16, Lemma 1], for which $\bar{\theta} = \underline{\theta} = 1$. To improve the convergence rate, one can design the auxiliary system (12b) to achieve the expected finite settling time when the local states reach the \mathcal{M}_0 , and then enlarge the control gain α to further reduce the finite consensus time. However, large gain might result in high chattering magnitude with discretization.

Remark 4: Assumption 3 used in Theorem 2 is more general than existing works [9]–[11], and holds for an important class of functions such as $f_i(x_i, t) = (a_i x_i + b_i(t))^2/2$, where a_i is heterogeneous among agents and only the boundedness of $\|\dot{b}_i(t)\|$ is required, e.g. $b_i(t) = t$ and $\sin(t)$. However, in [9], for the distributed method with neighbors' position, with identical Hessians, i.e., $a_i = a_j$, the boundedness of $\|b_i(t) - b_j(t)\|$, $\|\dot{b}_i(t) - \dot{b}_j(t)\|$ and $\|\ddot{b}_i(t) - \ddot{b}_j(t)\|$ is required to satisfy [9, Assumption 3.8]. Similarly, identical Hessians are used in [10], [11]. For DAT-based method provided in [9], to deal with non-identical Hessians, the boundedness of $\|\frac{d\nabla f_i(x_i, t)}{dt}\|$, $\|\frac{dH_i(x_i, t)}{dt}\|$ and $\|\frac{d(\partial f_i(x_i, t)/\partial t)}{dt}\|$ is required and known for the FT convergence of the DAT estimators. In fact, these items not only depend on $\dot{b}_i(t)$ and $\ddot{b}_i(t)$, but also rely on the evolution of local states x_i , and thus are difficult to estimate. For a general class of quadratic cost functions, e.g., $f_i(x_i, t) = \frac{1}{2} x_i^T H_i(t) x_i + b_i^T(t) x_i$, as $\nabla f_i(x_i, t) = H_i(t) x_i + b_i(t)$ also depends on the value of x_i , the value κ in Assumption 3 is not easy to be known generally. However, if $x_i(t)$ is known to be upper bounded, with the boundedness of $\|\dot{H}_i(t)\|$ and $\dot{d}_i(t)$, Assumption 3 holds directly.

C. FT Algorithm for Solving TV-DORAP

In this subsection, the TV-DORAP will be studied over a multi-agent system, where each agent is associated with a TV local cost function. All the involved agents will coordinate with each other to minimized the whole cost function subject to a coupled equation constraint. The DORAP can be formulated as

follows:

$$\min \sum_{i=1}^N f_i(x_i, t) \quad \text{s.t.} \quad \sum_{i=1}^N x_i = d(t), \quad (14)$$

where $x_i \in \mathbb{R}^n$ is the local state of agent i , $d(t) \in \mathbb{R}^n$ is the TV total resource that will be allocated among all agents, and $f_i : \mathbb{R}^n \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is the TV local cost function of agent i . We further suppose that $d(t) = \sum_{i=1}^N d_i(t)$, where $d_i(t) \in \mathbb{R}^n$ can be accessed only by the i th agent. Note that (14) is different from (10) where all the local states are required to achieve consensus on the optimal solution.

For the TV-DORAP (14), by introducing the Lagrangian multiplier associated with the equality constraint, the TV Lagrangian function can be written as

$$\mathcal{L}(\mathbf{x}, \lambda, t) = \sum_{i=1}^N f_i(x_i, t) + \lambda \sum_{i=1}^N (d_i(t) - x_i), \quad (15)$$

where $\mathbf{x} = [x_i]_{i \in \langle N \rangle} \in \mathbb{R}^{Nn}$. Then, by [29, Sec. 5], one can derive the Lagrangian dual function $g(\lambda, t)$

$$g(\lambda, t) = \inf_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda, t) = \sum_{i=1}^N (\lambda d_i(t) - f_i^*(\lambda, t)) = \sum_{i=1}^N g_i(\lambda, t),$$

where $g_i(\lambda, t) = \lambda d_i(t) - f_i^*(\lambda, t)$ can be regarded as the local distributed dual function. Under Lemma 1 and Assumption 2, it can be concluded that g_i is $\frac{1}{\theta_i}$ -strongly concave and $\frac{1}{\theta_i}$ -smooth. Moreover, as ∇f_i^* is an inverse map of ∇f_i and $\nabla^2 f_i(x_i(\lambda), t)$ is positive definite, it holds that $\nabla^2 f_i^*(\lambda, t) = (\nabla^2 f_i(x_i(\lambda, t), t))^{-1}$, where $x_i(\lambda, t) = \arg \sup_{x_i} (\lambda x_i - f_i(x_i, t))$. Besides, $\nabla g_i(\lambda, t) = d_i(t) - x_i(\lambda, t)$. Then, the dual problem of (14) is given by

$$\max_{\lambda \in \mathbb{R}^n} \sum_{i=1}^N g_i(\lambda, t), \quad (16)$$

which can be transferred to the TVDCO (10) as

$$\min_{\lambda \in \mathbb{R}^n} \sum_{i=1}^N -g_i(\lambda_i, t), \quad \text{s.t.} \quad \lambda_i = \lambda_j, \forall (i, j) \in \mathcal{E} \quad (17)$$

with $\boldsymbol{\lambda} = [\lambda_i]_{i \in \langle N \rangle} \in \mathbb{R}^{Nn}$. Then, the distributed algorithm (12) can be used to solve (17) as follows.

$$\begin{cases} \dot{\lambda}_i = -H_i(x_i, t)(\varphi_i(z_i) - \frac{\partial \nabla g_i(\lambda_i, t)}{\partial t} + \alpha \phi_i(\lambda_i, \lambda_{\mathcal{N}_i})), & (18a) \\ \dot{z}_i = -\varphi_i(z_i), & (18b) \\ x_i(\lambda_i, t) = \arg \sup_{x_i \in \mathbb{R}^n} (\lambda_i x_i - f_i(x_i, t)), \quad \forall i \in \langle N \rangle & (18c) \end{cases}$$

with $z_i(0) = -\nabla g_i(\lambda_i(0), 0) = x_i(\lambda_i, 0) - d_i(0)$ and $\frac{\partial \nabla g_i(\lambda_i, t)}{\partial t} = \dot{d}_i(t) - \frac{\partial x_i(\lambda_i, t)}{\partial t}$. Take $f_i(x_i, t) = \frac{1}{2} a_i x_i^2 + b_i(t) x_i + c_i(t)$ as an example with $x_i \in \mathbb{R}$. One can derive that $x_i(\lambda_i, t) = \frac{\lambda_i - b_i(t)}{a_i}$ and $\frac{\partial \nabla g_i(\lambda_i, t)}{\partial t} = \dot{d}_i(t) + \frac{\dot{b}_i(t)}{a_i}$. To ensure the boundedness of $\frac{\partial \nabla g_i(\lambda_i, t)}{\partial t}$, Assumption 7 is imposed on $d_i(t)$. Then, one can show the boundedness on $\frac{\partial \nabla g_i(\lambda_i, t)}{\partial t}$ in Lemma 3 under Assumptions 2, 3 and 7.

Assumption 7: There exists $\delta > 0$ such that $\|\dot{d}_i(t)\| \leq \delta, \forall i \in \langle N \rangle$.

Lemma 3: Let Assumptions 2, 3 and 7 hold. It satisfies that

$$\left\| \frac{\partial \nabla g_i(\lambda_i, t)}{\partial t} \right\| \leq \frac{\kappa}{\theta_i} + \delta, \quad \forall i \in \langle N \rangle. \quad (19)$$

Proof. See Appendix D. \square

By applying Theorem 2 and Lemma 3, one can derive the following corollary.

Corollary 1: Suppose that Assumptions 2-7 hold. Let $\underline{\theta} = \min_{i \in \langle N \rangle} \{\theta_i\}$ and $\bar{\theta} = \max_{i \in \langle N \rangle} \{\theta_i\}$. Consider the distributed protocol (18) with $\alpha > (\frac{\kappa}{\bar{\theta}} + \delta) \sqrt{\frac{N \bar{\theta}}{\underline{\theta} \lambda_2(B^T B)}}$, $z_i(t)$ will converge to 0 in finite (resp. fixed) time and then $x_i(t)$ will track the optimal trajectory $x_i^*(t)$ to the TV DORAP (14) in finite time.

Remark 5: In [16] and [17], several asymptotically convergent algorithms are provided for solving DORAP with TV quadratic cost functions. Moreover, TV demand $d_i(t)$ is considered in [18], where both $\|d_i(t)\|_\infty$ and $\|\dot{d}_i(t)\|_\infty$ are required to be bounded. In [16] and [18], an estimator based on DAT method is first provided for tracking the global information in finite time, and then a distributed CTA is designed to track the TV optimal solution asymptotically. Differently, the proposed algorithm (18) in this note has a simple form and can track the optimal trajectory in finite time. Moreover, compared with existing works [16]–[18], the cost functions can be of non-quadratic form with non-identical Hessians and the inverses of Hessian of f_i is not required in (18) from a dual perspective, which extends the existing results. However, another local optimization needs to be solved to obtain the primal variable at each time instant.

IV. NUMERICAL EXAMPLES

In this section, two case studies will be conducted to demonstrate the effectiveness of the proposed algorithms (12) and (18), respectively. We consider a multi-agent system consisting of $N = 12$ agents over the communication network $\mathcal{G}(\langle N \rangle, \mathcal{E})$ as shown in Fig. 1, for which the edge weights are all ones. Then, one can calculate that $\lambda_2(B^T B) = 1.239$ with B being the incidence matrix of \mathcal{G} . All the experiments are conducted using Matlab R2021b on a 2.9 GHz Intel Core i7.

For the first case study, we consider the following TV binary logistic regression model with l_2 -regularization:

$$\min_x f(x) = \sum_{i=1}^N (\log(1 + \exp(-l_i y_i^T(t) x)) + \frac{\beta_i}{2} \|x\|_2^2), \quad (20)$$

where $l_i \in \{-1, 1\}$, $y_i(t) \in \mathbb{R}^2$ and β_i is the regularization parameter, an integer randomly chosen in [1 8]. In the simulation, the TV sample data $y_i(t) = (1 + \sin(\omega t)) y_i(0)$ with $\omega = \pi/10$ and $y_i(0)$ chosen randomly in $[0 1]^2$. The proposed ‘‘consensus within local NR’’ dynamics (12) with $\alpha = 4$ and $\varphi(z_i) = 10 \text{sgn}^{0.5}(z_i)$ is compared with the ‘‘consensus+central NR’’ dynamics [9, Alg. (9)] and ‘‘prediction–correction’’ algorithm [4] with Laplacian constraint matrix. The performance of three algorithms is testified online in a real scenario. The former two CTAs are implemented by Euler discretization with step $h = 0.4$ ms as the computation/communication of each update consumes less than 0.4 ms. The ‘‘prediction–correction’’ algorithm is executed with $P = 3$ prediction steps and $C = 3$ correction steps, for which the sampling period is set as $h = 0.03$ s considering the computation/communication time for 3 prediction and correction operations using the `fmincon` function in MATLAB. Note that two rounds of communication are required for each prediction/correction operation. The primal states $x_i(0), i \in \mathcal{V}$ are randomly chosen in $[0 1]^2$. Let $E_x(t) = \log_{10}(\frac{1}{N} \sum_{i=1}^N \|x_i(t) - x^*(t)\|)$ denote the tracking error between real-time states and the TV optimal solution $x^*(t)$.

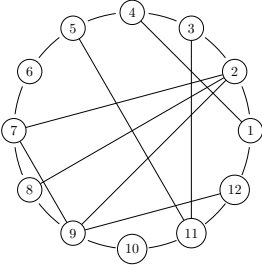
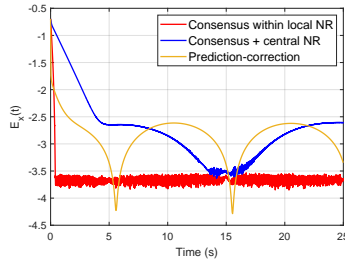
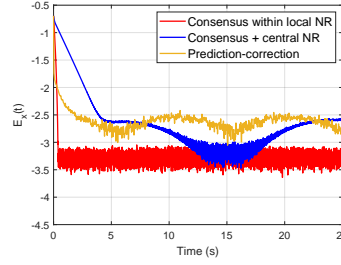
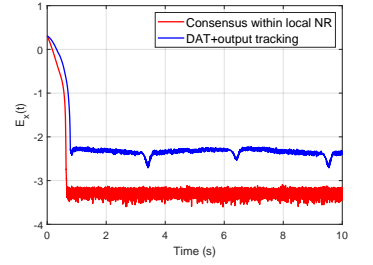


Fig. 1: The network with 12 agents.


 Fig. 2: Case 1: Values of $E_x(t)$.

 Fig. 3: Case 1: $E_x(t)$ with noise.

 Fig. 4: Case 2: Values of $E_x(t)$.

The simulation results are shown in Fig. 2. It can be seen that the proposed algorithm converges to a small neighborhood of $x^*(t)$ in finite time less than 1s and then stay therein, which is faster and more robust than the other two algorithms. However, the final chattering exists due to the discontinuous controller and Euler discretization. For comparison, the states generated by [9, Alg. (9)] asymptotically converge to a larger neighborhood of $x^*(t)$ compared with (12) partially due to the fact that the local Hessians are not identical [9]. For the “prediction–correction” algorithm, the local solutions also fluctuate around $x^*(t)$ possibly due to the sample period and the TV optimal solution. To testify the effects of noisy and inexact information on the performance of algorithms, the Gaussian noise $\epsilon \sim \mathcal{N}(0, 10^{-4})$ is added to each link and $\partial \nabla f_i(x_i, t)/\partial t$. The simulation results shown in Fig. 3 indicate that the accuracy of all algorithms is reduced because of the noise. However, due to the binary protocol, the convergence behavior of the former two algorithms is less affected than that of the last one.

For the second case study, the DORAP (14) is considered. For comparison, as the existing algorithms mostly consider DORAP with quadratic cost functions, here the cost functions f_i are given by $f_i(x_i, t) = 0.5(2 + 0.1i)x_i^2 + \sin(0.1it)x_i + \sin(0.6it)$. The local TV demand is set as $d_i(t) = i + \sin(t + i\pi/N)$, $\forall i \in \langle N \rangle$. The proposed “consensus within local NR” dynamics (18) with $\alpha = 5$ and $\varphi(z_i) = 10\text{sgn}^{0.5}(z_i)$ is compared with [18, Alg. (6)] (with all gains equal to 3), named by “DAT+output tracking” method as two DAT dynamics are firstly performed to track the global information and then an output tracking dynamics is given for ensuring the KKT condition. The algorithms are implemented by Euler discretization with step $h = 0.2$ ms by considering the computation/communication time in each update. Let $E_x(t) = \log_{10}(\frac{1}{N} \sum_{i=1}^N \|x_i(t) - x_i^*(t)\|)$ where $x_i^*(t)$ is the optimal TV solution. The simulation results shown in Fig. 4 indicate that the states driven by the proposed algorithm converge to a small neighborhood of $x^*(t)$ in finite time less than 1s, with a higher accuracy than [18, Alg. (6)]. Significantly, in this case (only quadratic functions are considered in [18]), both the computation/communication and storage cost of the proposed algorithm at each instant are less than those of [18, Alg. (6)] as more auxiliary variables are used and communicated in [18, Alg. (7)].

V. CONCLUSION

In this paper, both TV centralized and distributed optimization have been investigated. For TV centralized optimization, a unified approach is given for designing FT convergent algorithms. Then, a distributed discontinuous dynamics with FT convergence is proposed for solving TVDCO based on an extended

ZGS method with the local auxiliary dynamics. Furthermore, the proposed distributed dynamics is used for solving TV-DORAP with additional TV coupled equation constraints by dual transformation. Significantly, the cost functions can be of non-quadratic form with non-identical Hessians and the inversion of the cost functions’ Hessians is not required in the dual variables’ dynamics. However, another local optimization needs to be solved to obtain the primal variable at each time instant. Further work will focus on TVDO with general TV constraints over switching networks, as well as the convergence analysis on the discretized version of the proposed algorithm.

APPENDIX

A. Proof of Theorem 1

Considering the dynamics (7), one can derive that

$$d\nabla f(x, t)/dt = H_0(x, t)\dot{x} + \partial \nabla f(x, t)/\partial t = -\varphi(z) = \dot{z}. \quad (21)$$

Since $z(0) = \nabla f(x(0), 0)$, $z(t)$ and $\nabla f(x, t)$ coincide, i.e., $z(t) \equiv \nabla f(x, t)$, $\forall t \geq 0$. As (7b) is FT stable at origin with settling time T_{\max} , it indicates that $\nabla f(x(t), t)$ will converge to zero in finite time T_{\max} , i.e., (6) holds with $T = T_{\max}$.

B. Proof of Theorem 2

By Assumption 5, there exists a finite time T_1 such that $z(t) = 0$, $\forall t \geq T_1$. Based on (12), it can be derived that

$$\begin{aligned} \frac{d}{dt} \sum_{i=1}^N \nabla f_i(x_i, t) &= \sum_{i=1}^N (H_i(x_i, t)\dot{x}_i + \frac{\partial \nabla f_i(x_i, t)}{\partial t}) \\ &= -\sum_{i=1}^N \varphi_i(z_i) = \sum_{i=1}^N \dot{z}_i. \end{aligned}$$

Since $\sum_{i=1}^N z_i(0) = \sum_{i=1}^N \nabla f_i(x_i(0), 0)$, then it holds that

$$\sum_{i=1}^N z_i(t) = \sum_{i=1}^N \nabla f_i(x_i(t), t), \quad \forall t \geq 0. \quad (22)$$

From the previous analysis, $z(t) = 0$ for any $t \geq T_1$, which indicates that $\mathbf{x}(t) \in \mathcal{M}_0 \triangleq \{\mathbf{x}(t) : \sum_{i=1}^N \nabla f_i(x_i(t), t) = 0\}$. Then, to show that $x_i(t)$ tracks $x_i^*(t)$ in finite time, it is sufficient to prove that all the local states x_i achieve consensus in finite time. By (13b), for $t \geq T_1$, one can restrain that $\varphi(z) = 0$ since $z = 0$. Then, (13a) reduces to

$$\dot{\mathbf{x}} = -H^{-1}(\mathbf{x}, t) \left(\frac{\partial \nabla f(\mathbf{x}, t)}{\partial t} + \alpha B \text{sgn}(B^T \mathbf{x}) \right).$$

Denote $\hat{\mathbf{x}} = B^T \mathbf{x}$, $g(t) = \frac{\partial \nabla f(\mathbf{x}, t)}{\partial t}$ and $H(t) = H(\mathbf{x}, t)$. It holds that

$$\dot{\hat{\mathbf{x}}} = -B^T H^{-1}(t)(g(t) + \alpha B \text{sgn}(\hat{\mathbf{x}})). \quad (23)$$

Denote the right-hand side of (23) as X_r . By the sum rule, we have

$$\mathcal{F}[X_r](\mathbf{x}, t) \subseteq -B^T H^{-1}(t)(\mathcal{F}[g](t) + \alpha B \mathcal{F}[\text{sgn}](\hat{\mathbf{x}})). \quad (24)$$

Consider the Lyapunov function $V_2(t) = \|\hat{\mathbf{x}}\|_1$. By Lemma 2, it holds that $\dot{V}_2 \stackrel{a.e.}{\in} \tilde{\mathcal{L}}_{\mathcal{F}[X_r]} V_2$, i.e., there exist $\gamma \in \mathcal{F}[g](t)$ and $\zeta \in \mathcal{F}[\text{sgn}](\hat{\mathbf{x}})$ such that

$$\dot{V}_1 \stackrel{a.e.}{\in} -\nu^T B^T H^{-1}(t)(\gamma + \alpha B \zeta), \quad \forall \nu \in \partial \|\hat{\mathbf{x}}\|_1. \quad (25)$$

Note that $\partial \|\hat{\mathbf{x}}\|_1 = \mathcal{F}[\text{sgn}](\hat{\mathbf{x}}(t))$. Then, by choosing $\nu = \zeta$, it gives

$$\begin{aligned} \dot{V}_1 &\stackrel{a.e.}{\in} -\zeta^T B^T H^{-1}(t)(\gamma + \alpha B \zeta) \\ &\stackrel{a.e.}{\in} -\zeta^T B^T H^{-1}(t)\gamma - \alpha \zeta^T B^T H^{-1}(t)B\zeta \\ &\leq -(\alpha - \frac{\varepsilon}{2})\zeta^T B^T H^{-1}(t)B\zeta + \frac{1}{2\varepsilon}\gamma^T H^{-1}(t)\gamma \\ &\leq -(\alpha - \frac{\varepsilon}{2})\zeta^T B^T H^{-1}(t)B\zeta + \frac{N\kappa^2}{2\varepsilon\theta}, \end{aligned} \quad (26)$$

where the first inequality is obtained by applying $a^T b \leq \frac{1}{2}(\varepsilon a^T a + \frac{1}{\varepsilon} b^T b)$ with $a = -H^{-\frac{1}{2}}(t)B\zeta$ and $b = H^{-\frac{1}{2}}(t)g(t)$, and the second inequality is due to Assumptions 2 and 3 by setting $\theta = \min_{i \in \langle N \rangle} \{\theta_i\}$. Denote $M(t) = B^T H^{-1}(t)B$. It holds that

$$\zeta^T M(t)\zeta \geq \lambda_2(H^{-1}(t))\zeta^T B^T B\zeta \geq \frac{1}{\theta}\zeta^T B^T B\zeta \quad (27)$$

with $\bar{\theta} = \max_{i \in \langle N \rangle} \{\theta_i\}$. As $B^T B$ is positive semidefinite, then it holds that

$$\zeta^T B^T B\zeta \geq \lambda_2(B^T B)\|\zeta - \Pi_{\mathcal{N}(B)}(\zeta)\|^2 \quad (28)$$

since $\mathcal{N}(B^T B) = \mathcal{N}(B)$. Let $\eta = \Pi_{\mathcal{N}(B)}(\zeta)$. As $\hat{\mathbf{x}} = B^T \mathbf{x}$, then $\hat{\mathbf{x}}^T \eta = 0$. Due to $\zeta = \text{sign}(\hat{\mathbf{x}})$, when $\hat{\mathbf{x}} \neq 0$, there exists at least one entry k such that $\zeta_k = \text{sign}(\hat{x}_k) \neq 0$ and $\zeta_k \neq \text{sign}(\eta_k)$, which implies that $\|\zeta - \eta\| \geq \|\zeta_k - \eta_k\| \geq 1$ and thus $\zeta^T B^T B\zeta \geq \lambda_2(B^T B)$. Then, we have

$$\dot{V}_1 \leq -(\alpha - \frac{\varepsilon}{2})\frac{1}{\theta}\lambda_2(B^T B) + \frac{N\kappa^2}{2\varepsilon\theta} \triangleq -\varrho,$$

for which the right side is minimized at $\varepsilon^* = \kappa\sqrt{\frac{N}{\theta\lambda_2(B^T B)}}$.

By choosing $\alpha > \kappa\sqrt{\frac{N\bar{\theta}}{\theta\lambda_2(B^T B)}}$, one gets $\varrho > 0$ and $\dot{V}_1 \leq -\varrho$, implying that $V_1(t)$ will converge to zero in finite time T_2 . In other words, all the local states x_i reach consensus in T_2 . As $\mathbf{x}(t) \in \mathcal{M}_0$ for $t \geq T_1$, one can conclude that $x_i(t) = x^*(t)$ for any $t \geq T_2$.

C. Proof of Proposition 1

Following the proof of Theorem 2 and denoting $\bar{a} = \max_{(i,j) \in \mathcal{E}} \{a_{ij}\}$, Eq. (26) can be further relaxed as

$$\begin{aligned} \dot{V}_1 &= -\zeta^T B^T H^{-1}(t)\gamma - \alpha \zeta^T B^T H^{-1}(t)B\zeta \\ &\leq -\alpha \zeta^T B^T H^{-1}(t)B\zeta + m\varpi\bar{a} \\ &\leq -\frac{\alpha}{\theta}\lambda_2(B^T B) + m\varpi\bar{a}. \end{aligned}$$

By choosing $\alpha > \frac{m\varpi\bar{a}\bar{\theta}}{\lambda_2(B^T B)}$, the conclusion can be shown similar to the proof of Theorem (2).

D. Proof of Lemma 3

By the definition of $x_i(\lambda_i, t)$ in (18c), it holds that $\lambda_i - \nabla f_i(x_i(\lambda_i, t), t) = 0$, from which it can be derived that

$$H_i(x_i, t)\frac{\partial x_i(\lambda_i, t)}{\partial t} + \frac{\partial \nabla f_i(x_i, t)}{\partial t} = 0. \quad (29)$$

Then, we have $\frac{\partial x_i(\lambda_i, t)}{\partial t} = -H_i^{-1}(x_i, t)\frac{\partial \nabla f_i(x_i, t)}{\partial t}$ and hence by applying Assumptions 2, 3 and 7, for any $i \in \langle N \rangle$, it holds that

$$\begin{aligned} \left\| \frac{\partial \nabla g_i(\lambda_i, t)}{\partial t} \right\| &= \left\| \frac{\partial x_i(\lambda_i, t)}{\partial t} - \dot{d}_i(t) \right\| \\ &\leq \|H_i^{-1}(x_i, t)\frac{\partial \nabla f_i(x_i, t)}{\partial t}\| + \|\dot{d}_i(t)\| \leq \frac{\kappa}{\theta_i} + \delta. \end{aligned}$$

REFERENCES

- [1] A. Simonetto, E. Dall'Anese, S. Paternain, G. Leus, and G. B. Giannakis, "Time-varying convex optimization: Time-structured algorithms and applications," *Proc. of the IEEE*, vol. 108, no. 11, pp. 2032–2048, Nov. 2020.
- [2] S. Hosseini, A. Chapman, and M. Mesbahi, "Online distributed convex optimization on dynamic networks," *IEEE Trans. Autom. Control*, vol. 61, no. 11, pp. 3545–3550, Nov. 2016.
- [3] A. Simonetto, A. Koppel, A. Mokhtari, G. Leus, and A. Ribeiro, "Decentralized prediction-correction methods for networked time-varying convex optimization," *IEEE Trans. Autom. Control*, vol. 62, no. 11, pp. 5724–5738, Nov. 2017.
- [4] A. Simonetto, "Dual prediction–correction methods for linearly constrained time-varying convex programs," *IEEE Trans. Autom. Control*, vol. 64, no. 8, pp. 3355–3361, 2018.
- [5] R. Dixit, A. S. Bedi, and K. Rajawat, "Online learning over dynamic graphs via distributed proximal gradient algorithm," *IEEE Trans. Autom. Control*, vol. 66, no. 11, pp. 5065–5079, 2020.
- [6] K. Yuan, W. Xu, and Q. Ling, "Can primal methods outperform primal-dual methods in decentralized dynamic optimization," *IEEE Trans. Signal Process.*, vol. 68, pp. 4466–4480, 2020.
- [7] X. Li, L. Xie, and N. Li, A survey of decentralized online learning. arXiv preprint arXiv:2205.00473.
- [8] E. Dall'Anese, A. Simonetto, S. Becker, and L. Madden, "Optimization and learning with information streams: Time-varying algorithms and applications," *IEEE Sig. Process. Mag.*, vol. 37, no. 3, pp. 71–83, 2020.
- [9] S. Rahili and W. Ren, "Distributed continuous-time convex optimization with time-varying cost functions," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1590–1605, Apr. 2017.
- [10] B. Huang, Y. Zou, Z. Meng, and W. Ren, "Distributed time-varying convex optimization for a class of nonlinear multiagent systems," *IEEE Trans. Autom. Control*, vol. 65, no. 2, pp. 801–808, Feb. 2020.
- [11] B. Ning, Q.-L. Han, and Z. Zuo, "Distributed optimization for multiagent systems: An edge-based fixed-time consensus approach," *IEEE Trans. Cybern.*, vol. 49, no. 1, pp. 122–132, Jan. 2019.
- [12] C. Sun, M. Ye, and G. Hu, "Distributed time-varying quadratic optimization for multiple agents under undirected graphs," *IEEE Trans. Autom. Control*, vol. 62, no. 7, pp. 3687–3694, Jul. 2017.
- [13] Y. Zhu, W. Ren, W. Yu, and G. Wen, "Distributed resource allocation over directed graphs via continuous-time algorithms," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 2, pp. 1097–1106, Feb. 2021.
- [14] W. Jia, N. Liu, and S. Qin, "An adaptive continuous-time algorithm for nonsmooth convex resource allocation optimization," *IEEE Trans. Autom. Control*, vol. 67, no. 11, pp. 6038–6044, Nov. 2022.
- [15] A. Cherukuri and J. Cortés, "Initialization-free distributed coordination for economic dispatch under varying loads and generator commitment," *Automatica*, vol. 74, pp. 183–193, 2016.
- [16] B. Wang, S. Sun, and W. Ren, "Distributed continuous-time algorithms for optimal resource allocation with time-varying quadratic cost functions," *IEEE Trans. Control of Netw. Syst.*, vol. 7, no. 4, pp. 1974–1984, Dec. 2020.
- [17] B. Wang, Q. Fei, and Q. Wu, "Distributed time-varying resource allocation optimization based on finite-time consensus approach," *IEEE Contr. Syst. Lett.*, vol. 5, no. 2, pp. 599–604, Apr. 2021.
- [18] B. Wang, S. Sun, and W. Ren, "Distributed time-varying quadratic optimal resource allocation subject to nonidentical time-varying Hessians with application to multiquadrotor hose transportation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 10, pp. 6109–6119, Oct. 2022.
- [19] L. Bai, C. Sun, Z. Feng, and G. Hu, "Distributed continuous-time resource allocation with time-varying resources under quadratic cost functions," in *Proc. IEEE Conf. Decis. Control*, Miami Beach, FL, USA, Dec. 2018, pp. 823–828.

- [20] Y. Song and W. Chen, "Finite-time convergent distributed consensus optimisation over networks," *IET Control Theory Appl.*, vol. 10, no. 11, pp. 1314-1318, 2016.
- [21] J. Lu and C. Y. Tang, "Zero-gradient-sum algorithms for distributed convex optimization: The continuous-time case," *IEEE Trans. Autom. Control*, vol. 57, no. 9, pp. 2348-2354, Sept. 2012.
- [22] X. Shi, X. Xu, X. Yu, and J. Cao, "Finite-time convergent primal-dual gradient dynamics with applications to distributed optimization," *IEEE Trans. Cybern.*, vol. 53, no. 5, pp. 3240-3252, 2023.
- [23] Z. Hu and J. Yang, "Distributed finite-time optimization for second order continuous-time multiple agents systems with time-varying cost function," *Neurocomputing*, vol. 287, pp. 173-184, 2018.
- [24] Z. Jiang, K. Mukherjee and S. Sarkar, "Convergence and noise effect analysis for generalized gossip-based distributed optimization," 2017 American Control Conference (ACC), Seattle, WA, USA, 2017, pp. 4353-4358.
- [25] V. P. Chellapandi, A. Upadhyay, A. Hashemi, and S. H. Zak, "On the convergence of decentralized federated learning under imperfect information sharing," *IEEE Contr. Syst. Lett.*, vol. 7, pp. 2982-2987, 2023.
- [26] M. Jafarian and C. De Persis, "Formation control using binary information," *Automatica*, vol. 53, pp. 125-135, 2015.
- [27] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, 2009.
- [28] X. Shi, G. Wen, J. Cao, and X. Yu, "Finite-time distributed average tracking for multi-agent optimization with bounded inputs," *IEEE Trans. Autom. Control*, vol. 68, no. 8, 2023.
- [29] S. P. Boyd and L. Vandenberghe. Convex optimization. Cambridge university press, 2004.
- [30] A. Polyakov, "Nonlinear feedback design for fixed-time stabilization of linear control systems," *IEEE Trans. Autom. Control*, vol. 57, no. 8, pp. 2106-2110, Aug. 2011.