Input-Feedforward-Passivity-Based Distributed Optimization Over Directed and Switching Topologies

Mengmou Li, Graziano Chesi, and Yiguang Hong

***Abstract*— In this paper, a distributed optimization problem is investigated via input feedforward passivity. First, an input-feedforward-passivity-based continuous-time distributed algorithm is proposed. It is shown that the error system of the proposed algorithm can be interpreted as output feedback interconnections of a group of Input Feedforward Passive (IFP) systems. Second, based on this IFP framework, the distributed algorithm is studied over weight-balanced directed and uniformly jointly strongly connected switching topologies. Specifically, the continuous-time distributed algorithm for uniformly jointly strongly connected digraphs has never been considered before. Sufficient convergence conditions are derived for the design of a suitable coupling gain.**

I. Introduction

Distributed optimization over multi-agent systems has been widely investigated in recent years, due to its broad applications in various aspects including wireless networks, smart grids, and machine learning. In addition to the discrete-time algorithms (e.g., [1], [2]), a variety of continuous-time distributed algorithms have been proposed to solve distributed optimization problems [3]–[6], owing to the benefit of continuous-time stability theory for convergence analysis. However, many of the proposed algorithms are only for undirected networks and not applicable in directed networks [3]–[5]. To deal with this difficulty, the work in [7] tunes some parameters in the original algorithm to stabilize gradient dynamics while the work in [8] proposes a variant algorithm. However, compared with these methods, a more systematic approach is needed for this problem.

It is well known that dissipativity (and its special case, passivity) is a useful tool for stability analysis and control design [9]. Recently, there emerge some continuous-time passivity-based algorithms on distributed optimization under some communication constraints [10]–[12]. However, these passivity-based algorithms can only be applied over undirected graphs, while it is shown that output consensus can be achieved over directed graphs through simple output feedback interconnections of passive systems [13]. Motivated by these works, we aim to study distributed algorithms in directed graphs via dissipativity/passivity techniques. On one hand, we conjecture that it is in general difficult to directly construct a distributed algorithm that can be interpreted as output feedback interconnections of passive systems. On the other hand, works in [14]–[16] point out that output consensus can be achieved over directed graphs even among IFP systems (or passivity-short systems). Therefore, if a distributed algorithm inherits input feedforward passivity, it can be directly applied to directed graphs through output feedback interconnections. As a byproduct of having the IFP properties, the distributed algorithm is robust against the switching topologies, while the effort in constructing complicated candidate Lyapunov functions is greatly reduced in convergence analysis. To the best of our knowledge, though the case of uniformly jointly strongly connected switching topologies has been considered in discrete-time algorithms [17], it has never been considered in continuous-time algorithms before.

The challenges in our work lie in the construction of a group of verifiable nonlinear IFP systems that solve the distributed optimization problem, and the convergence analysis over directed and switching topologies.

The rest of this paper is organized as follows. In Section II, some background knowledge of convex analysis, graph theory, and dissipativity/passivity is reviewed. In Section III, an IFP-based distributed algorithm is proposed. In Section IV, the proposed distributed algorithm is studied over directed and switching topologies. In Section V, a numerical example is presented to demonstrate the effect of the proposed algorithm. Finally, the paper is concluded in Section VI.

II. Preliminaries

A. Notation

Let $\mathbb{R}$ and $\mathbb{Z}$ be the set of real and integer numbers, respectively. The notation $ker(A)$ denotes the kernel of A. The Kronecker product is denoted as $\otimes$. $\|A\|$ denotes the 2-norm of A. Given a symmetric matrix $M \in \mathbb{R}^{m\times m}$, the notation $M > 0$ ($M \geq 0$) means that M is positive definite (positive semi-definite). Denote the eigenvalues of M in ascending order as $s_1(M) \leq s_2(M) \leq \ldots \leq s_m(M)$. I and $\mathbf{0}$ denote the identity matrix and zero matrix (or vector) of proper dimensions, respectively. $\mathbf{1}_m := (1,\ldots,1)^T \in \mathbb{R}^m$. $col(v_1,\ldots,v_m) := (v_1^T,\ldots,v_m^T)^T$ denotes the column vector stacked with vectors $v_1,\ldots,v_m$. The notation $diag\{\alpha_i\}$ denotes a (block) diagonal matrix with its ith diagonal element (block) being α_i. The notation $\mathcal{C}^k$ is used to denote a $k \in \mathbb{Z}_{\geq 1}$ times continuously differentiable function.

B. Convex Analysis

A differential function $f : \mathbb{R}^m \to \mathbb{R}$ is *convex* over a convex set $\mathcal{X} \subset \mathbb{R}^m$ if and only if $[\nabla f(x) - \nabla f(y)]^T (x - y) \geq 0$, $\forall x, y \in \mathcal{X}$ and *strictly convex* if and only if the strict inequality holds. It is μ*-strongly convex* if and only

Mengmou Li and Graziano Chesi are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: mengmou_li@hku.hk; chesi@eee.hku.hk).

Yiguang Hong is with the Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 100190, Beijing, China (e-mail: yghong@iss.ac.cn).

if $\left[\nabla f(x)-\nabla f(y)\right]^T(x-y)\geq \mu\|x-y\|_2^2$, $\forall x,y\in\mathcal{X}$. An equivalent condition for the strong convexity is the following: $f(y)\geq f(x)+\nabla f(x)^T(y-x)+\frac{\mu}{2}\|y-x\|_2^2$, $\forall x,y\in\mathcal{X}$. A function $\mathbf{f}:\mathbb{R}^m\to\mathbb{R}^m$ is *l-Lipschitz continuous* over a set $\mathcal{X}$ if $\|\mathbf{f}(x)-\mathbf{f}(y)\|\leq l\|x-y\|$, $\forall x,y\in\mathcal{X}$.

C. Graph Theory

The information exchanging network is represented by a graph $\mathcal{G}=(\mathcal{N},\mathcal{E})$, where $\mathcal{N}=\{1,\ldots,N\}$ is the node set of all agents and $\mathcal{E}\subset\mathcal{N}\times\mathcal{N}$ is the edge set. The edge $(i,j)\in\mathcal{E}$ denotes that agent i can obtain information from agent j, and $j\in\mathcal{N}_i$ where $\mathcal{N}_i=\{(i,j)\in\mathcal{E}\}$ is agent i's neighbor set. The graph $\mathcal{G}$ is said to be *undirected* if $(i,j)\in\mathcal{E}\Leftrightarrow(j,i)\in\mathcal{E}$ and *directed* otherwise. A sequence of successive edges $\{(i,p),(p,q),\ldots,(v,j)\}$ is a *direct path* from agent i to agent j. An *undirected path* is defined similarly without considering directions of edges. $\mathcal{G}$ is said to be *strongly* (*weakly*) *connected* if there exists a directed (undirected) path between any two agents. A time-varying graph $\mathcal{G}(t)$ is said to be *uniformly jointly strongly connected* if there exists a $T>0$ such that for any t_k, the union $\cup_{t\in[t_k,t_k+T]}\mathcal{G}(t)$ is strongly connected. The adjacency matrix is defined as $A=[a_{ij}]$, where $a_{ii}=0$; $a_{ij}=1$ if $(i,j)\in\mathcal{E}$, and $a_{ij}=0$, otherwise. The in-degree and out-degree of the ith agent are $d_{in}^i=\sum_{j=1}^N a_{ij}$ and $d_{out}^i=\sum_{j=1}^N a_{ji}$, respectively. The graph $\mathcal{G}$ is said to be *weight-balanced* if $d_{in}^i=d_{out}^i$, $\forall i\in\mathcal{N}$. The in-degree matrix is $W_{in}=diag\{d_{in}^i\}$. The Laplacian matrix of $\mathcal{G}$ is defined as $L=W_{in}-A$.

Lemma 1 ([18]): A weight-balanced digraph $\mathcal{G}$ is strongly connected if and only if it is weakly connected.

D. Passivity

Consider a group of agents having the nonlinear dynamics described by

$$\Sigma_i:\begin{cases}\dot{x}_i=f_i\left(x_i,u_i\right)\\ y_i=h_i\left(x_i,u_i\right)\end{cases},\quad\forall i\in\mathcal{N}\tag{1}$$

where $x_i\in\mathcal{X}_i\subset\mathbb{R}^n$, $u_i\in\mathcal{U}_i\subset\mathbb{R}^m$ and $y_i\in\mathcal{Y}_i\subset\mathbb{R}^m$ are the state, input and output, respectively, and $\mathcal{X}_i$, $\mathcal{U}_i$ and $\mathcal{Y}_i$ are the state, input and output spaces, respectively. The functions $f_i\in\mathbb{R}^{n\times n}$, $h_i\in\mathbb{R}^{n\times m}$ are assumed to be sufficiently smooth.

Let us first give the definition of passivity for a nonlinear system Σ_i based on [19], [20].

Definition 1: System Σ_i is said to be *passive* if there exists a continuously differentiable positive semidefinite function $V_i(x_i)$, called the *storage function*, such that

$$\dot{V}_i\leq u_i^Ty_i,\quad\forall(x_i,u_i)\in\mathcal{X}_i\times\mathcal{U}_i.\tag{2}$$

Moreover, it is said to be *Input Feedforward Passive* (IFP) if $\dot{V}_i\leq u_i^Ty_i-\nu_iu_i^Tu_i$, for some $\nu_i\in\mathbb{R}$, denoted as IFP(ν_i). The sign of the IFP index ν_i denotes an excess or shortage of passivity. Particularly, when $\nu_i>0$, the system is said to be *Input Strictly Passive* (ISP). When $\nu_i<0$, the system is said to be *Input Feedforward Passivity-Short* (IFPS).

Throughout this paper, we consider the storage function to be positive definite and radially unbounded.

III. IFP-Based Distributed Algorithm

Consider the convex distributed optimization problem among a group of N agents

$$\min_{\mathrm{x}}\sum_{i\in\mathcal{N}}f_i(\mathrm{x})\tag{3}$$

where $\mathrm{x}\in\mathbb{R}^m$ and each local objective function $f_i:\mathbb{R}^m\to\mathbb{R}$ satisfies the following assumption.

Assumption 1: Each $f_i(\mathrm{x}),i\in\mathcal{N}$ is $\mathcal{C}^2$ and μ_i–strongly convex, with its gradient $\nabla f_i(\mathrm{x})$ being l_i-Lipschitz continuous.

This assumption also implies that $\|\nabla f_i(\mathrm{x})-\nabla f_i(\mathrm{x}')\|\leq l_i\|\mathrm{x}-\mathrm{x}'\|$ and $\mu_iI\leq\nabla^2 f_i(\mathrm{x})\leq l_iI$, $\forall\mathrm{x},\mathrm{x}'$.

Problem (3) is equivalent to

$$\begin{aligned}\min_{x_i,\forall i\in\mathcal{N}}\quad & f(x)=\sum_{i\in\mathcal{N}}f_i(x_i)\\ \text{subject to}\quad & x_i=x_j,\quad\forall i,j\in\mathcal{N}\end{aligned}\tag{4}$$

where $x_i\in\mathbb{R}^m$ is the local decision variable for agent i.

A. IFP-Based Distributed Algorithm

We propose an IFP-based distributed algorithm for agent i, $\forall i\in\mathcal{N}$ as follows.

Algorithm 1 IFP-Based Distributed Algorithm

$$\dot{x}_i=-\alpha\nabla f_i(x_i)-K_i\lambda_i+\beta u_i\tag{5a}$$
$$\dot{\lambda}_i=-\gamma J_iu_i\tag{5b}$$
$$K_iJ_i=C^T\tag{5c}$$
$$u_i=\sigma\sum_{j\in\mathcal{N}_i}a_{ij}(Cx_j-Cx_i)\tag{5d}$$

For the ith agent, $x_i,\lambda_i,\in\mathbb{R}^m$ and $u_i\in\mathbb{R}^m$ are local variables and input, respectively; $J_i,K_i\in\mathbb{R}^{m\times m}$ are invertible matrices such that $K_iJ_i=C^T$ is a common matrix; $\alpha>0$, $\beta\in\mathbb{R}$ and $\gamma>0$ are constant parameters and $\sigma>0$ is the *coupling gain*. To ease the discussion on parameters, we assume that $\alpha,\beta,\gamma,C,K_i,J_i,\forall i\in\mathcal{N}$ are some pre-given arbitrary values while σ is to be designed. Apparently, Algorithm 1 is a distributed algorithm since each agent only exchanges information with neighboring agents.

Denote $x=col(x_1,\ldots,x_N)$ and $\lambda=col(\lambda_1,\ldots,\lambda_N)$, then the compact form of system (5) is

$$\dot{x}=-\alpha\nabla f(x)-\mathbf{K}\lambda-\sigma\beta\mathbf{L}\mathbf{C}x\tag{6a}$$
$$\dot{\lambda}=\sigma\gamma\mathbf{J}\mathbf{L}\mathbf{C}x\tag{6b}$$

where $\mathbf{K}=diag\{K_i\}$, $\mathbf{J}=diag\{J_i\}$, $\mathbf{C}=I_N\otimes C$ are block diagonal matrices, $\mathbf{L}=L\otimes I_m$, and L is the graph Laplacian of the communication graph $\mathcal{G}$.

Lemma 2: Suppose $\mathcal{G}$ is strongly connected and Assumption 1 holds. If there exists an equilibrium point (x^*,λ^*) that satisfies $\sum_{i\in\mathcal{N}}K_i\lambda_i^*=\mathbf{0}$ in system (6), then (x^*,λ^*) is also unique with x_i^* being the optimal solution to problem (3).

Proof: The equilibrium point (x^*, λ^*) satisfies

$$\dot{x}^* = -\alpha \nabla f(x^*) - \mathbf{K}\lambda^* = \mathbf{0} \tag{7a}$$
$$\dot{\lambda}^* = \gamma\sigma \mathbf{JLC} x^* = \mathbf{0}. \tag{7b}$$

$\dot{\lambda}^* = \mathbf{0}$ implies that $Cx_i^* = Cx_j^*,\ \forall i,j \in \mathcal{N}$. Since $K_i J_i = C^T$ and J_i, K_i are invertible, C is also invertible and $x_i^* = x_j^*,\ \forall i,j \in \mathcal{N}$. Moreover, multiplying (7a) by $(\mathbf{1}_N \otimes I_m)^T$ from the left, one has,

$$\begin{aligned}
&-(\mathbf{1}_N \otimes I_m)^T \alpha \nabla f(x^*) - (\mathbf{1}_N \otimes I_m)^T \mathbf{K}\lambda^* \\
=&-\sum_{i\in\mathcal{N}} \alpha \nabla f_i(x_i^*) - \sum_{i\in\mathcal{N}} K_i \lambda_i^* \\
=&-\alpha \sum_{i\in\mathcal{N}} \nabla f_i(x_i^*) = \mathbf{0}
\end{aligned}$$

which implies that x_i^* is the optimal solution to problem (3). By Assumption 1, x^* is unique. Since $\mathbf{K}$ is invertible, λ^* is unique as well. ■

Hereafter, we call (x^*, λ^*) the *optimal point*. The convergence of Algorithm 1 will be addressed in Section IV.

B. Input Feedforward Passivity of the Error System

Denote $\Delta x_i = x_i - x_i^*$, $\Delta\lambda_i = \lambda_i - \lambda_i^*$. Then, the group of error subsystems between (6) and (7), with each one denoted by Σ_i, is

$$\begin{cases}
\Delta\dot{x}_i &= -\alpha\left[\nabla f_i(x_i) - \nabla f_i(x_i^*)\right] - K_i\Delta\lambda_i + \beta u_i \\
\Delta\dot{\lambda}_i &= -\gamma J_i u_i \\
y_i &= C\Delta x_i
\end{cases}, \forall i \in \mathcal{N} \tag{8}$$

where y_i is the output of the ith subsystem. Then the input u_i, $\forall i \in \mathcal{N}$ can be rewritten as output feedback of neighboring agents

$$u_i = \sigma \sum_{j\in\mathcal{N}_i} a_{ij}(y_j - y_i), \quad \forall i \in \mathcal{N}. \tag{9}$$

Assume that, corresponding to the real agents, there exist a group of virtual agents such that the ith virtual agent possesses the subsystem Σ_i, $i \in \mathcal{N}$. Then Algorithm 1 can be seen as output feedback interconnections of these virtual agents. In fact, no information of (x_i^*, λ_i^*) is needed for communication since $y_i - y_j = C\Delta x_i - C\Delta x_j = C(x_i - x_j)$. Then, each agent possesses the same information as its corresponding virtual agent. The communication topology is the same as well.

Before showing the convergence of Algorithm 1, we first show that each error subsystem Σ_i in (8) is IFP(ν_i) with index $\nu_i \leq 0$.

Lemma 3: Under Assumption 1, each error subsystem Σ_i in (8) is IFP(ν_i).

Proof: Under Assumption 1, one has $\nabla f_i(x_i) - \nabla f_i(x_i^*) = B_{x_i}(x_i - x_i^*)$, where $B_{x_i} = \int_0^1 \nabla^2 f_i(x_i^* + \tau(x_i - x_i^*))d\tau$ is a positive definite matrix such that $\mu_i I \leq B_{x_i} \leq l_i I$ ([21, Lemma 1]). Apparently, B_{x_i} is invertible and $B_{x_i}^{-1}$ is also positive definite. Then, the ith subsystem in (8) can be written as

$$\begin{aligned}
\Delta\dot{x}_i &= -\alpha B_{x_i}\Delta x_i - K_i\Delta\lambda_i + \beta u_i \\
\Delta\dot{\lambda}_i &= -\gamma J_i u_i \\
y_i &= C\Delta x_i.
\end{aligned}$$

Denote

$$z_i = \alpha B_{x_i}\Delta x_i + K_i\Delta\lambda_i. \tag{10}$$

Let us consider the storage function

$$\begin{aligned}
V_i = &\frac{\eta}{2} z_i^T z_i - \frac{1}{\gamma}\Delta x_i^T K_i \Delta\lambda_i + \frac{\alpha}{\gamma}\left[f_i(x_i^*) - f_i(x_i)\right] \\
&+ \frac{\alpha}{\gamma}\left[\nabla f_i(x_i^*)^T \Delta x_i\right]
\end{aligned} \tag{11}$$

where η is a positive parameter such that $\eta > \frac{1}{\mu_i\alpha\gamma}$. By the strong convexity of f_i, one has

$$\begin{aligned}
f_i(x_i^*) \geq & f_i(x_i) - \nabla f_i(x_i)^T \Delta x_i + \frac{\mu_i}{2}\Delta x_i^T \Delta x_i \\
=& f_i(x_i) - \nabla f_i(x_i)^T \frac{B_{x_i}^{-1}}{\alpha}\left(\alpha B_{x_i}\Delta x_i\right) \\
&+ \left(\alpha B_{x_i}\Delta x_i\right)^T \frac{\mu_i B_{x_i}^{-2}}{2\alpha^2}\left(\alpha B_{x_i}\Delta x_i\right).
\end{aligned}$$

Then

$$\begin{aligned}
V_i \geq & \frac{\eta}{2} z_i^T z_i - \frac{1}{\gamma}\Delta x_i^T K_i\Delta\lambda_i + \frac{\alpha}{\gamma}\nabla f_i(x_i^*)^T \frac{B_{x_i}^{-1}}{\alpha}\left(\alpha B_{x_i}\Delta x_i\right) \\
&- \frac{\alpha}{\gamma}\nabla f_i(x_i)^T \frac{B_{x_i}^{-1}}{\alpha}\left(\alpha B_{x_i}\Delta x_i\right) \\
&+ \frac{\alpha}{\gamma}\left(\alpha B_{x_i}\Delta x_i\right)^T \frac{\mu_i B_{x_i}^{-2}}{2\alpha^2}\left(\alpha B_{x_i}\Delta x_i\right) \\
=& \frac{\eta}{2} z_i^T z_i - \left(\alpha B_{x_i}\Delta x_i\right)^T \frac{B_{x_i}^{-1}}{\alpha\gamma} K_i\Delta\lambda_i - \left(\alpha B_{x_i}\Delta x_i\right)^T \cdot \\
&\frac{B_{x_i}^{-1}}{\alpha\gamma}\left(\alpha B_{x_i}\Delta x_i\right) + \left(\alpha B_{x_i}\Delta x_i\right)^T \frac{\mu_i B_{x_i}^{-2}}{2\alpha\gamma}\left(\alpha B_{x_i}\Delta x_i\right) \\
=& \begin{bmatrix} \alpha B_{x_i}\Delta x_i \\ K_i\Delta\lambda_i \end{bmatrix}^T R_i \begin{bmatrix} \alpha B_{x_i}\Delta x_i \\ K_i\Delta\lambda_i \end{bmatrix}
\end{aligned}$$

where $R_i = \begin{bmatrix} \frac{\eta}{2}I + \frac{\mu_i B_{x_i}^{-2}}{2\alpha\gamma} - \frac{B_{x_i}^{-1}}{\alpha\gamma} & \frac{\eta}{2}I - \frac{B_{x_i}^{-1}}{2\alpha\gamma} \\ * & \frac{\eta}{2}I \end{bmatrix}$. By the Schur complement [22], $R_i > 0$ if and only if $\frac{\eta}{2} > 0$ and

$$\frac{\eta}{2}I + \frac{\mu_i B_{x_i}^{-2}}{2\alpha\gamma} - \frac{B_{x_i}^{-1}}{\alpha\gamma} - \frac{2}{\eta}\left(\frac{\eta}{2}I - \frac{B_{x_i}^{-1}}{2\alpha\gamma}\right)^T\left(\frac{\eta}{2}I - \frac{B_{x_i}^{-1}}{2\alpha\gamma}\right) > 0.$$

Select η such that $\eta > \frac{1}{\mu_i\alpha\gamma}$, then $R_i > 0$. Hence, $V_i > 0$ and $V_i = 0$ if and only if $(x_i, \lambda_i) = (x_i^*, \lambda_i^*)$.

Recall (10) and $\dot{x}_i^* = \dot{\lambda}_i^* = \mathbf{0}$, then $\Delta\dot{x}_i = -z_i + \beta u_i$ and $\dot{z}_i = \alpha\nabla^2 f_i(x_i)\Delta\dot{x}_i + K_i\Delta\dot{\lambda}_i$. Thus, the derivative of

V_i gives

$$\begin{aligned}
\dot{V}_i =& \eta z_i^T\left[-\alpha\nabla^2 f_i(x_i)(z_i-\beta u_i) - K_i\gamma J_i u_i\right] \\
&- \frac{1}{\gamma}\left[\Delta x_i^T K_i(-\gamma J_i u_i) + (-z_i+\beta u_i)^T K_i \Delta\lambda_i\right] \\
&+ \frac{\alpha}{\gamma}\left\{-\left[\nabla f_i(x_i) - \nabla f_i(x_i^*)\right]^T(-z_i+\beta u_i)\right\} \\
=& -\eta\alpha z_i^T\nabla^2 f_i(x_i)z_i + \eta z_i^T[\alpha\beta\nabla^2 f_i(x_i) - \gamma K_i J_i]u_i \\
&+ \Delta x_i^T K_i J_i u_i + \frac{1}{\gamma}z_i^T K_i\Delta\lambda_i - \frac{\beta}{\gamma}u_i^T K_i \Delta\lambda_i \\
&+ \frac{1}{\gamma}\left(\alpha B_{x_i}\Delta x_i\right)^T z_i - \frac{\beta}{\gamma}\left(\alpha B_{x_i}\Delta x_i\right)^T u_i \\
=& -\eta\alpha z_i^T\nabla^2 f_i(x_i)z_i + \eta z_i^T[\alpha\beta\nabla^2 f_i(x_i) - \gamma K_i J_i]u_i \\
&+ (C\Delta x_i)^T u_i + \frac{1}{\gamma}z_i^T z_i - \frac{\beta}{\gamma}z_i^T u_i \\
\le& -\left(\mu_i\eta\alpha - \frac{1}{\gamma}\right)z_i^T z_i + y_i^T u_i \\
&+ z_i^T\underbrace{\left\{\eta[\alpha\beta\nabla^2 f_i(x_i) - \gamma C^T] - \frac{\beta}{\gamma}I\right\}}_{g_i}u_i \\
\le& -\left(\mu_i\eta\alpha - \frac{1}{\gamma}\right)\|z_i\|^2 + \|z_i\|\|g_i\|\|u_i\| + y_i^T u_i \\
\le& y_i^T u_i - \nu_i u_i^T u_i
\end{aligned} \tag{12}$$

where $\mu_i\eta\alpha - \frac{1}{\gamma} > 0$ follows from $\eta > \frac{1}{\mu_i\alpha\gamma}$, and $\nu_i \le -\frac{\|g_i\|^2}{4\left(\mu_i\eta\alpha - \frac{1}{\gamma}\right)} \le 0$. Since $\nabla^2 f_i(x_i)$ and parameters in g_i are bounded, given finite η, a constant ν_i can be obtained. Thus, Σ_i is IFP(ν_i). ■

When the error system (8) is linear, i.e., each f_i is quadratic, $\forall i \in \mathcal{N}$, by solving an LMI in [23], it can also be proved numerically that Σ_i is IFP(ν_i) with index $\nu_i \le 0$.

As pointed out by [16], it is in general difficult to derive the exact IFP index for a nonlinear system, and only its lower bound can be obtained by specifying the storage function. With the storage function (11), the lower bound of IFP index can be obtained locally by solving the minimax problem

$$\nu_i = -\min_{\eta}\max_{x_i}\frac{\left\|\eta[\alpha\beta\nabla^2 f_i(x_i) - \gamma C^T] - \frac{\beta}{\gamma}I\right\|^2}{4\left(\mu_i\eta\alpha - \frac{1}{\gamma}\right)}. \tag{13}$$

The problem of reducing this gap remains open and leaves to the future work.

Remark 1: Let $J_i,\ K_i = I$, and $\sigma = 1$. When $\gamma = \alpha\beta$, Algorithm 1 reduces to the distributed algorithm in [8]. When $\alpha, \gamma = 1$, and $\beta = 0$, Algorithm 1 reduces to the simplified algorithm in [8]. Compared with algorithms in [8], Algorithm 1 includes more general cases whose convergence cannot be proved by methods in [8], e.g., when β is negative and when γ is independent of α, β. Besides, agents in Algorithm 1 can exchange the information of Cx_i instead of x_i thanks to extra matrices J_i, K_i. Moreover, it is shown later that Algorithm 1 is valid over uniformly jointly strongly connected topologies in addition to directed and strongly connected switching networks [8].

IV. Directed and Switching Topologies

In this section, we show that the IFP framework allows the study of distributed algorithms over directed topologies as well as uniformly jointly strongly connected switching topologies, while the effort in constructing complicated candidate Lyapunov functions in convergence analysis is greatly reduced.

A. Directed Graphs

Assumption 2: The communication graph $\mathcal{G}$ is weight-balanced and strongly connected.

Definition 2: The group of agents (1) is said to achieve output consensus if $\lim_{t\to\infty}\|y_i(t) - y_j(t)\| = 0,\ \forall i, j \in \mathcal{N}$.

Theorem 1: Under Assumption 1 and Assumption 2, Algorithm 1 will converge to the optimal point if $\sum_{i\in\mathcal{N}} K_i\lambda_i(0) = \mathbf{0}$ and the coupling gain satisfies

$$0 < \sigma < \frac{s_2\left(L + L^T\right)}{-2\bar{\nu}s_N\left(L^T L\right)} \tag{14}$$

where $\bar{\nu} < 0$ is the smallest value of IFP index $\nu_i,\ i \in \mathcal{N}$ and $s(\cdot)$ is defined in Section II-A.

Proof: Let $V = \sum_{i\in\mathcal{N}} V_i$, where V_i is defined in (11). Since B_{x_i} and K_i are bounded, $\left\|\frac{\Delta x}{\Delta\lambda}\right\| \to \infty$ implies $V \to \infty$, and then V is radially unbounded. The overall output is $y = col(y_1, \ldots, y_N)$ and the overall input is $u = -\sigma\mathbf{L}y$.

$$\begin{aligned}
\dot{V} \le& \sum_{i\in\mathcal{N}} y_i^T u_i - \nu_i u_i^T u_i \\
\le& \sum_{i\in\mathcal{N}} y_i^T u_i - \bar{\nu} u_i^T u_i \\
=& -\sigma y^T(L\otimes I_m)y - \sigma^2\bar{\nu}y^T(L^T L\otimes I_m)y \\
=& y^T\left\{\underbrace{\left[-\frac{\sigma}{2}(L+L^T) - \sigma^2\bar{\nu}L^T L\right]}_{M}\otimes I_m\right\}y \\
=& y^T(M\otimes I_m)y.
\end{aligned} \tag{15}$$

Observe that L, L^T and $L^T L$ have the same null space $\{c\mathbf{1}_N,\ c\in\mathbb{R}\}$. Then there exists a similarity transformation that reduces the zero eigenvalue of $L^T L$ and $(L + L^T)$ at the same time. Besides, since similarity transformation does not change the eigenvalues, we can check the definiteness of M by comparing the nonzero eigenvalues of $L^T L$ and $(L + L^T)$. Since (14) holds, $M \le 0$ and zero is the simple eigenvalue with the eigenvector $\mathbf{1}_N$. Then, $\dot{V} \le 0$ and the system is globally stable.

By (15) and the first inequality in (12), $\dot{V} = 0$ only if $(M\otimes I_m)y = \mathbf{0}$ and $z_i = \mathbf{0}$, where z_i is defined in (10). Thus, $\mathcal{S} = \{z_i = \mathbf{0}, y_i = y_j, \forall i, j \in \mathcal{N}\}$ is the largest invariant set. Then, by (8), (10), and the LaSalle's Invariance Principle [19], $\Delta\dot{x} \to 0$, $\Delta\dot{\lambda} \to 0$ as $t \to \infty$. The states converge to an equilibrium point.

Since $\lambda - \lambda(0) = \int_0^t \dot{\lambda}(\tau)d\tau$, given the initial condition

$\sum_{i\in\mathcal{N}} K_i\lambda_i(0)=\mathbf{0}$,

$$\begin{aligned}
&(\mathbf{1}_N\otimes I_m)^T\mathbf{K}\lambda\\
=&(\mathbf{1}_N\otimes I_m)^T\mathbf{K}\left(\int_0^t \sigma\gamma\mathbf{J}\mathbf{L}\mathbf{C}x(\tau)d\tau+\lambda(0)\right)\\
=&\sigma\gamma\int_0^t(\mathbf{1}_N\otimes I_m)^T(I_N\otimes C^T)(L\otimes I_m)\mathbf{C}x(\tau)d\tau\\
&+\sum_{i\in\mathcal{N}} K_i\lambda_i(0)\\
=&\sigma\gamma\int_0^t(\mathbf{1}_N^T L\otimes C^T)\mathbf{C}x(\tau)d\tau\\
=&\mathbf{0}
\end{aligned}$$

where the third equality follows from rules of Kronecker product and the fourth follows from $\mathbf{1}_N^T L=\mathbf{0}$. Then Lemma 2 holds, the equilibrium point is the optimal point. Therefore, Algorithm 1 will converge to the optimal point. ■

To obtain $\bar{\nu}$, it does not require the knowledge of the objective functions of other agents but only the strong convexity index ν_i and Lipschitz index l_i to estimate ν_i in (13). Details of the design of σ are given in Section IV-C.

Note that only weight-balanced digraphs are considered here. The consensus under unbalanced graphs can be guaranteed similarly with $V=\sum_{i\in\mathcal{N}}\xi_i V_i$, where $\xi_i>0$ is the ith element of the left eigenvalue of L [14], [16]. However, the sum of local objective functions will have a shift from global optimum [17]. Thus, some modification is needed, which will be discussed in the future.

B. Uniformly Jointly Strongly Connected Digraphs

Consider the distributed algorithm over uniformly jointly strongly connected switching digraphs. To the best of our knowledge, the continuous-time algorithm for uniformly jointly strongly connected networks has never been considered before.

Assumption 3: The agents interact with each other over a uniformly jointly strongly connected digraph $\mathcal{G}(t)$ that is weight-balanced pointwise in time with Laplacian $L(t)\neq\mathbf{0}$, $\forall t\geq 0$.

Here the trivial case of $L(t)=\mathbf{0}$ is omitted without affecting the feasible range of the coupling gain σ.

Theorem 2: Under Assumption 1 and Assumption 3, Algorithm 1 will converge to the optimal point if $\sum_{i\in\mathcal{N}} K_i\lambda_i(0)=\mathbf{0}$ and the coupling gain σ satisfies

$$0<\sigma<\frac{s_+\left(L(t)+L^T(t)\right)}{-2\bar{\nu}s_N\left(L^T(t)L(t)\right)},\ \forall t>0 \qquad (16)$$

where $s_+(\cdot)$ denotes the nonzero smallest eigenvalue.

Proof: Since $\mathcal{G}(t)$ is weight-balanced pointwise in time, by Lemma 1, if agent i and j are weakly connected, there must exist an index set $\mathcal{N}_q$ containing these two agents such that its corresponding graph $\mathcal{G}_{\mathcal{N}_q}$ is strongly connected. Therefore, the graph $\mathcal{G}(t)$ at any time t can be seen as the union of one or several disjoint strongly connected digraphs. Consequently, $L(t)$ and $L^T(t)$ have the same null space, i.e., $ker\left(L(t)\right)=ker\left(L^T(t)\right)$. Then, due to the well known fact that $ker(L(t))=ker\left(L^T(t)L(t)\right)$, there exists a similarity transformation that reduces the zero eigenvalues of $L^T(t)L(t)$ and $(L(t)+L^T(t))$ at the same time. Besides, for weight-balanced digraphs, $\zeta^T\left(L(t)+L^T(t)\right)\zeta=\sum_{i,j\in\mathcal{N}}a_{ij}(t)(\zeta_j-\zeta_i)^2\geq 0$, $\forall\zeta\in\mathbb{R}^N$. Then, $\left(L(t)+L^T(t)\right)$ and $L^T(t)L(t)$ are both nonnegative. The rest of the lines are similar to the proof in Theorem 1. Define $Q(t)=-\frac{1}{2}\left(L(t)+L^T(t)\right)-\sigma\bar{\nu}L^T(t)L(t)$. Since (16) holds, $Q(t)\leq 0$, which leads to $\dot{V}\leq 0$. Then the system is globally stable.

Consider an infinite sequence $V(t_i)$, $i=1,\ldots$, where the time t_i approaches infinity as i approaches infinity. Notice from $Q(t)$ that $\dot{V}(t_i)=0$ only if all the locally connected agents at time t_i reach consensus. There exist t_k and t_l, where $t_l-t_k\geq T$ such that $[t_k,t_l]$ encompasses some time interval across which the agents are uniformly jointly strongly connected. Then, $\lim_{k\to+\infty}\dot{V}(t_k)=\lim_{k\to+\infty}\dot{V}(t_{k+1})=\ldots=\lim_{k\to+\infty}\dot{V}(t_l)=0$, which implies that $y\in\mathcal{S}=\{y_i=y_j,\forall\, i,j\}$, i.e., output consensus is achieved.

Therefore, following the proof in Theorem 1, the system will converge to the equilibrium point that is exactly the unique optimal point. Consequently, Algorithm 1 will converge to the optimal point. ■

C. Design of the Coupling Gain

Theorems 1 and 2 provide sufficient conditions for convergence to the optimal point. In this subsection, we proceed to discuss the design of the coupling gain σ given the values of $\alpha,\beta,\gamma,C,K_i,J_i,\forall i\in\mathcal{N}$.

Note that all agents should have the same σ in order to converge to the optimal point, which means that all agents should have a predetermined protocol to design a proper identical coupling gain. For instance, the coupling gain can be simply chosen as $\sigma=k\sigma_e$, where $k<1$ is a predetermined positive constant and σ_e is the threshold of coupling gains obtained in the above theorems.

To say the least, though some graph information is required in order to obtain the exact threshold σ_e, by the fact that the upper bounds in (14) and (16) are positive, there always exists a small-enough σ such that the trajectories of Algorithm 1 will converge to the optimal point. In fact, for proper parameters, there is usually a wide feasible range for the coupling gain. Let us take for instance the quadratic functions (i.e., linear time-invariant systems in (8)) from the perspective of passivity, with $\alpha,\beta,\gamma=1$, $C=I$. When the strongly convex index $\mu_i>1$, it can be shown by solving an LMI in [23] that the IFP index ν_i is infinitesimal for each agent. Then σ_e can be arbitrarily large based on the above theorems.

We will derive design methods free of graph information in the near future.

V. A Numerical Example

We present a numerical example to demonstrate the effect of Algorithm 1 over switching digraphs in this section.

Consider a network of 4 agents possessing the following local objective functions, respectively,

$$f_i(x) = 0.4(x-i)^2,\ x \in \mathbb{R},\ i = 1, 2, 3, 4.$$

Let $\alpha, \beta, \gamma = 1$, and $J_i = diag\{1/i\}, K_i = diag\{i\}$. Then it can be obtained that each subsystem in (8) is IFP with $\nu_i = -0.3125, \forall i$. Next, we consider two cases of topologies. The graph $\mathcal{G}(t)$ is arbitrarily switching among three modes $①\leftarrow②\leftarrow③\leftarrow①$, $②\leftarrow③\leftarrow④\leftarrow②$, and $①\leftarrow③\leftarrow④\leftarrow①$. The corresponding graph Laplacians are $L_1 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, $L_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 1 \end{bmatrix}$, and $L_3 = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 \end{bmatrix}$, respectively.

The threshold coupling gain is obtained as $\sigma_e = 1.60$ in (16). Algorithm 1 is tested on MATLAB with $x_i(0) \in [0, 1]$, $\lambda(0) = \mathbf{0}$ satisfying the initial condition, and $\sigma = 0.70$ lower than the threshold. The convergence results are shown in Fig. 1. It can be observed that the trajectories of x_i asymptotically converge to the optimal solution $x_i^* = 2.5,\ \forall i$.

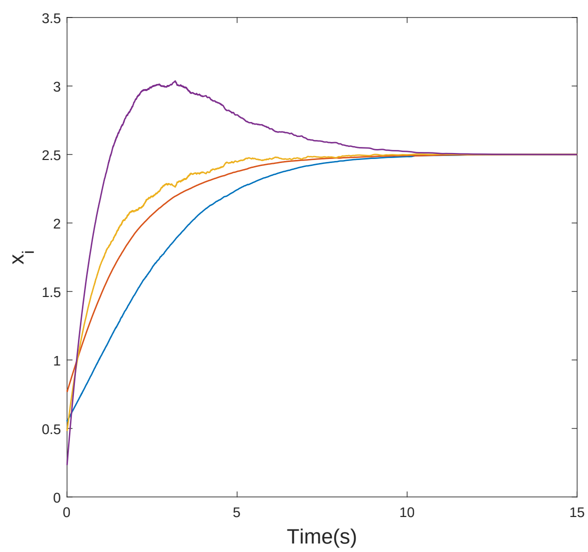

Fig. 1. The trajectories of x_i over a uniformly jointly strongly connected switching digraph.

VI. Conclusion

This paper has addressed a distributed optimization problem via input feedforward passivity. An IFP framework has been adopted to construct a distributed algorithm that is applicable over directed and uniformly jointly connected switching topologies. Sufficient convergence conditions have been derived for the design of a suitable coupling gain. Future work may consider the proposed distributed algorithm in more complicated communication constraints.

References

[1] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[3] J. Wang and N. Elia, "Control approach to distributed optimization," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2010, pp. 557–561.

[4] P. Yi, Y. Hong, and F. Liu, "Distributed gradient algorithm for constrained optimization with application to load sharing in power systems," *Systems & Control Letters*, vol. 83, pp. 45–52, 2015.

[5] M. Li, "Generalized lagrange multiplier method and KKT conditions with an application to distributed optimization," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 2, pp. 252–256, 2019.

[6] S. Liang, X. Zeng, and Y. Hong, "Distributed sub-optimal resource allocation over weight-balanced graph via singular perturbation," *Automatica*, vol. 95, pp. 222–228, 2018.

[7] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 781–786, 2014.

[8] S. S. Kia, J. Cortés, and S. Martínez, "Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication," *Automatica*, vol. 55, pp. 254–264, 2015.

[9] T. Liu, D. J. Hill, and J. Zhao, "Output synchronization of dynamical networks with incrementally-dissipative nodes and switching topology," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 9, pp. 2312–2323, 2015.

[10] Y. Tang, Y. Hong, and P. Yi, "Distributed optimization design based on passivity technique," in *Control and Automation (ICCA), 2016 12th IEEE International Conference on*. IEEE, 2016, pp. 732–737.

[11] Y. Tang and P. Yi, "Distributed coordination for a class of non-linear multi-agent systems with regulation constraints," *IET Control Theory & Applications*, vol. 12, no. 1, pp. 1–9, 2017.

[12] T. Hatanaka, N. Chopra, T. Ishizaki, and N. Li, "Passivity-based distributed optimization with communication delays using PI consensus algorithm," *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4421–4428, 2018.

[13] N. Chopra and M. W. Spong, "Passivity-based control of multi-agent systems," in *Advances in robot control*. Springer, 2006, pp. 107–134.

[14] Z. Qu and M. A. Simaan, "Modularized design for cooperative control and plug-and-play operation of networked heterogeneous systems," *Automatica*, vol. 50, no. 9, pp. 2405–2414, 2014.

[15] A. V. Proskurnikov and M. Mazo Jr, "Simple synchronization protocols for heterogeneous networks: beyond passivity," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 9426–9431, 2017.

[16] M. Li, L. Su, and G. Chesi, "Consensus of heterogeneous multi-agent systems with diffusive couplings via passivity indices," *IEEE Control Systems Letters*, vol. 3, no. 2, pp. 434–439, 2019.

[17] P. Xie, K. You, R. Tempo, S. Song, and C. Wu, "Distributed convex optimization with inequality constraints over time-varying unbalanced digraphs," *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4331–4337, 2018.

[18] C. Godsil and G. F. Royle, *Algebraic graph theory*. Springer Science & Business Media, 2013, vol. 207.

[19] H. K. Khalil, "Nonlinear systems," *Prentice-Hall, New Jersey*, 1996.

[20] R. Sepulchre, M. Jankovic, and P. V. Kokotovic, *Constructive nonlinear control*. Springer Science & Business Media, 2012.

[21] G. Qu and N. Li, "On the exponential stability of primal-dual gradient dynamics," *IEEE Control Systems Letters*, vol. 3, no. 1, pp. 43–48, 2019.

[22] J. Gallier, "The schur complement and symmetric positive semidefinite (and definite) matrices," *Penn Engineering*, 2010.

[23] N. Kottenstette, M. J. McCourt, M. Xia, V. Gupta, and P. J. Antsaklis, "On relationships among passivity, positive realness, and dissipativity in linear systems," *Automatica*, vol. 50, no. 4, pp. 1003–1016, 2014.